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

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

[1] State Key Laboratory of Systematic and Evolutionary Botany, Institute of Botany, Chinese Academy of Sciences, 20 Nanxincun, Haidian District, Beijing 100093, China

[2] Peking-Tsinghua Center for Life Sciences, Academy for Advanced Interdisciplinary Studies, Peking University, 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, 390 Beibinhe West Road, Anning District, Lanzhou, Gansu 730070, China

[5] Beijing Botanical Garden, Institute of Botany, Chinese Academy of Sciences, 20 Nanxincun, Haidian District, Beijing 100093, China

[6] Institute of Desertification Studies, Chinese Academy of Forestry, 10, Huai-shu-ju Road, Haidian District, Beijing 100091, China

Corresponding Authors: Qi Lu, Zhili Suo
Email address: Luqi@caf.ac.cn, zlsuo@ibcas.ac.cn

The *Haloxylon* genus belongs to the Amaranthaceae (formerly Chenopodiaceae) family. The small trees or shrubs in this genus are referred to as the King of psammophytic plants, and perform important functions in environmental protection, including wind control and sand fixation in deserts. To better understand these beneficial plants, we sequenced the chloroplast (cp) genomes of *Haloxylon ammodendron* (HA) and *Haloxylon persicum* (HP) and conducted comparative genomic analyses on these and two other representative Amaranthaceae species. Similar to other higher plants, we found that the *Haloxylon* cp genome is a quadripartite, double-stranded, circular DNA molecule of 151,570 bp in HA and 151,586 bp in HP. It contains a pair of inverted repeats (24,171 bp in HA and 24,177 bp in HP) that separate the genome into a large single copy region of 84,214 bp in HA and 84,217 bp in HP, and a small single copy region of 19,014 bp in HA and 19,015 bp in HP. Each *Haloxylon* cp genome contains 112 genes, including 78 coding, 30 tRNA, and four ribosomal RNA genes. We detected 59 different simple sequence repeat loci, including 44 mono-nucleotide, three di-nucleotide, one tri-nucleotide, and 11 tetra-nucleotide repeats. Comparative analysis revealed only 67 mutations between the two species, including 44 substitutions, 23 insertions/deletions, and two micro-inversions. The two inversions, with lengths of 14 and 3 bp, occur in the *petA-psbJ* intergenic region and *rpl16* intron, respectively, and are predicted to form hairpin structures with repeat sequences of 27 and 19 bp, respectively, at the two ends. The ratio of transitions to transversions was 0.76. These results are valuable for future studies on *Haloxylon* genetic diversity and will enhance our understanding of the phylogenetic evolution of Amaranthaceae.

1 **Comparative analysis of the complete chloroplast genome sequences in**

2 **psammophytic *Haloxylon* species (Amaranthaceae)**

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

9 [1]State Key Laboratory of Systematic and Evolutionary Botany, Institute of Botany, Chinese

10 Academy of Sciences, 20 Nanxincun, Haidian District, Beijing 100093, China

11 E-mail: wpdong@ibcas.ac.cn (W.P.D.); xuchao@ibcas.ac.cn (X.C.)

12 [2]Peking-Tsinghua Center for Life Sciences, Academy for Advanced Interdisciplinary Studies,

13 Peking University, 5 Yiheyuan Road Haidian District, Beijing 100871, China

14 [3]University of Chinese Academy of Sciences, 19 A Yuquan Road, Shijingshan District, Beijing

15 100049, China

16 [4]Gansu Desert Control Research Institute, 390 Beibinhe West Road, Anning District, Lanzhou,

17 Gansu 730070, China

18 E-mail: lidlu2008@163.com

19 [5]Beijing Botanical Garden, Institute of Botany, Chinese Academy of Sciences, 20 Nanxincun,

20 Haidian District, Beijing 100093, China

21 E-mail: jinxiaobai@ibcas.ac.cn

22 E-mail: lirl@ibcas.ac.cn, 20778033@qq.com

23 [6]Institute of Desertification Studies, Chinese Academy of Forestry, 10, Huai-shu-ju Road,

24 Haidian District, Beijing 100091, China

26 ***Correspondence:** zlsuo@ibcas.ac.cn; Tel.: 86 10 13520435137

27 Luqi@caf.ac.cn; Tel.: 86 10 13910830860

31 **ABSTRACT**

32 The *Haloxylon* genus belongs to the Amaranthaceae (formerly Chenopodiaceae) family. The small trees or

33 shrubs in this genus are referred to as the King of psammophytic plants, and perform important functions in

34 environmental protection, including wind control and sand fixation in deserts. To better understand these

35 beneficial plants, we sequenced the chloroplast (cp) genomes of *Haloxylon ammodendron* (HA) and *Haloxylon*

36 *persicum* (HP) and conducted comparative genomic analyses on these and two other representative

37 Amaranthaceae species. Similar to other higher plants, we found that the *Haloxylon* cp genome is a

38 quadripartite, double-stranded, circular DNA molecule of 151,570 bp in HA and 151,586 bp in HP. It contains

39 a pair of inverted repeats (24,171 bp in HA and 24,177 bp in HP) that separate the genome into a large single

40 copy region of 84,214 bp in HA and 84,217 bp in HP, and a small single copy region of 19,014 bp in HA and

41 19,015 bp in HP. Each *Haloxylon* cp genome contains 112 genes, including 78 coding, 30 tRNA, and four

42 ribosomal RNA genes. We detected 59 different simple sequence repeat loci, including 44 mono-nucleotide,

43 three di-nucleotide, one tri-nucleotide, and 11 tetra-nucleotide repeats. Comparative analysis revealed only 67

44 mutations between the two species, including 44 substitutions, 23 insertions/deletions, and two micro-

45 inversions. The two inversions, with lengths of 14 and 3 bp, occur in the *petA-psbJ* intergenic region and *rpl16*

46 intron, respectively, and are predicted to form hairpin structures with repeat sequences of 27 and 19 bp,

47 respectively, at the two ends. The ratio of transitions to transversions was 0.76. These results are valuable for

48 future studies on *Haloxylon* genetic diversity and will enhance our understanding of the phylogenetic evolution

49 of Amaranthaceae.

50

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

52

53 **INTRODUCTION**

54 The eudicot clade comprises approximately 75% of all flowering land plant species, including major subclades:

55   rosids, asterids, Saxifragales, Santalales, and Caryophyllales (*APG III, 2009*). *Haloxylon* species, which

56   include psammophytic small trees or shrubs, are positioned phylogenetically in the Amaranthaceae Juss of the

57   Caryophyllales Perleb among core eudicots (*APG III, 2009*; *Pyankov et al., 2001*; *Akhani et al., 2007*). The

58   *Haloxylon* genus has about 11 species, with a distribution from the Mediterranean through Central Asia and

59   into China (*Zhu et al., 2004*). Two *Haloxylon* species, which are known as the King of psammophytic plants,

60   are found in the deserts of northwest China and, play important roles in environmental protection, including

61   wind control and sand fixation (*Zhu et al., 2004*; *Jia & Lu, 2004*). These precious psammophytic woody plants

62   can adapt to harsh environmental conditions, such as drought, desert, high temperature, and sand storms.

63   However, populations of *Haloxylon* plants have been threatened in China in past decades as a result of

64   decreased underground water, overgrazing, and over exploitation of agriculture.

65   Because of the environmental significance of these plants and their declining numbers, genetic research on

66   *Haloxylon* germplasm resources has garnered significant interest (*Song & Jia, 2000*; *Sheng et al., 2004, 2005*;

67   *Zhang et al., 2006a, 2006b*). However, *Haloxylon* plants possess only fine green assimilating shoots, without

68   leaves, making the evaluation of their phenotypic diversity difficult. Further, the detection of genetic diversity

69   within *Haloxylon* germplasm resources has been slowed by a lack of morphological markers (*Sheng et al.,*

70   *2004, 2005*; *Zhang et al., 2006a, 2006b*; *Wang et al., 2009*; *Suo et al., 2012a*). A recent study by Long *et al.*

71   (2014) used RNA-seq data to elucidate the *Haloxylon* transcriptome, providing a valuable sequence resource

72   for further genetic and genomic studies; however, genetic information for members of the *Haloxylon* genus,

73   and how they might differ from one another, is limited.

74   Each leaf cell of plants contains 1,000 to 10,000 chloroplasts (cp), which are key organelles for

75   photosynthesis and other biochemical pathways such as the biosynthesis of starch, fatty acids, pigments, and

76   amino acids (*Dong et al., 2013b*; *Raman and Park, 2016*). Since the first cp genome of *Nicotiana tabacum* was

77   sequenced in 1986, around 800 complete cp genome sequences have been made available in the National

78   Center for Biotechnology Information organelle genome database. These data are valuable sources of genetic

79   markers for phylogenetic analyses, genetic diversity evaluation, and plant molecular identification (*Dong et al.,*

80   *2012, 2013a, 2013b, 2014*; *Ni et al., 2016*; *Suo et al., 2012b*).

81   There are two published complete cp genome sequences (*Spinacia oleracea* and *Beta vulgaris* subsp.

82   *vulgaris*) from members of the Amaranthaceae family (*Li et al., 2014*; *Schmitz-Linneweber et al., 2001*).

83 However, the determination of the cp genome from *Haloxylon* plants is of further significance for potentially

84 enhancing our understanding of their adaptability to severe desert environmental conditions, and their genomic

85 evolution within the Amaranthaceae. Here, we report the complete cp genomes from two *Haloxylon* species, *H.*

86 *ammodendron* and *H. persicum*, including patterns of nucleotide substitutions, microstructural mutation, and

87 simple sequence repeats (SSRs). We further performed genomic comparative analyses on these and two other

88 representative Amaranthaceae species, to better understand the evolutionary relationships within this family.

89

MATERIALS & METHODS

Sampling and DNA extraction

92 Fresh young shoots of *H. ammodendron* (HA) and *H. persicum* (HP) were collected in May 2011 from Minqin

93 Eremophytes Botanical Garden (N 38°34′ , E 102°59′, Altitude 1378 m), Gansu Province, China (under the

94 leadership of Gansu Desert Control Research Institute, 390 Beibinhe West Road, Anning District, Lanzhou,

95 Gansu 730070, China). These HA and HP plants were originally introduced from the Turpan Desert Botanical

96 Garden of Chinese Academy of Sciences, Xinjiang Uygur Autonomous Region. The shoots from each

97 accession were immediately dried using silica gel for future DNA extraction. Total genomic DNA (gDNA)

98 was extracted from each using the Plant Genomic DNA Kit (DP305) from Tiangen Biotech (Beijing) Co., Ltd.,

99 China. The approval numbers are 2012BAD16B0101 and 80117B1001 for field permit of the research.

100

Chloroplast genome sequencing

102 The HA and HP cp genomes were sequenced using the short-range PCR method reported by Dong *et al*. (2012,

103 2013). The PCR protocol was as follows: preheating at 94°C for 4 min, 34 cycles at 94°C for 45 s, annealing at

104 55°C for 40 s, and elongation at 72°C for 1.5 min, followed by a final extension at 72°C for 10 min. PCR

105 amplification was performed in an Applied Biosystems VeritiTM 96-Well Thermal Cycler (Model#: 9902,

106 made in Singapore). The amplicons were sent to Shanghai Majorbio Bio-Pharm Technology Co., Ltd (Beijing)

107 for Sanger sequencing in both the forward and reverse directions using a 3730xl DNA analyzer (Applied

108 Biosystems, Foster City, CA, USA). DNA regions containing poly structures or that were difficult to amplify

109 were further sequenced using newly designed primer pairs for confirming reliable and high quality sequencing

110 results.

111

Chloroplast genome assembling and annotation

113 The cp DNA sequences were manually confirmed and assembled using Sequencher (v4.6) software, and cp

114 genome annotation was performed using the Dual Organellar Genome Annotator (DOGMA) (*Wyman et al.,*

115 *2004*). BLASTX and BLASTN searches were utilized to accurately annotate the protein-encoding genes and to

116 identify the locations of the transfer RNAs (tRNAs) and ribosomal RNAs (rRNAs). Gene annotation

117 information from other closely related plant species was also used for confirmation when the boundaries of the

118 introns or exons could not be precisely determined because of the limited power of BLAST in cp genome

119 annotation (e.g., for some short exons of 6–9 nt in length, such as in the case of *rps16*, *petB,* and *petD)*.

120 Promoter, intron, and exon boundaries, as well as the location of stop codons for all protein-encoding genes,

121 have been identified accurately. The cp genome map was drawn using Genome Vx software (*Conant & Wolfe*,

122 *2008*) (http://wolfe.ucd.ie/GenomeVx/), and the cp genome sequences have been deposited to GenBank with

123 the following accession numbers: KF534478 for HA and KF534479 for HP

124 (https://www.ncbi.nlm.nih.gov/nuccore/?term=Haloxylon+chloroplast+genome).

125

Repeat structure analysis

127 Gramene **Simple Sequence Repeat Identification Tool** software (

128 http://www.gramene.org/db/markers/ssrtool)(*Benson, 1999*) was utilized to search for simple sequence repeat

129 loci in the cp genome sequences, with the threshold value of repeat number as ≥10 for mono-nucleotide repeats,

130 ≥5 for di-nucleotide repeats, ≥4 for tri-nucleotide repeats, and ≥3 for tetra-nucleotide, penta-nucleotide, or

131 hexa-nucleotide repeats.

132

Gene content analysis and comparative genomics

134 The mVISTA program was employed in Shuffle-LAGAN mode (*Frazer et al.*, *2004*) to compare the complete

135 HA and HP cp genomes. These were aligned using MUSCLE software (*Thompson et al.*, *1997*) and were

136 manually adjusted using Se-Al 2.0 (*Rambaut, 1996*). Variable sites in the cp genome were calculated using

137 DnaSP (DNA Sequences Polymorphism version 5.10.01) software (*Librado & Rozas*, *2009*), and the genetic

138 distance (p-distance) was computed using MEGA 6.0 software (*Tamura et al.*, *2011*). Based on the aligned

139 sequence matrix, the micro-structure events were checked manually and were further divided into three

140 categories: (i) microsatellite-related insertions/deletions (indels), (ii) non-microsatellite-related indels, (iii) and

141 inverted sequences. Using the HA cp genome sequence as the standard reference, the size, location, and

142 evolutionary direction of the microstructure events were counted. The proposed secondary structures of the

143 inverted regions in the cp genomes of HA and HP were analyzed using mfold software (*Zuker*, *2003*). The

144 complete cp genome sequences of *S. oleracea* (GenBank accession number AJ400848.1, *Spinacia* L.)

145 (*Schmitz-Linneweber et al.*, *2001*) and *B. vulgaris* subsp. *vulgaris* (GenBank accession number KJ081864.1,

146 *Beta vulgaris* subsp. *vulgaris*) (*Li et al.*, *2014*), two closely related species in the Amaranthaceae family, were

147 downloaded from GenBank databases (www.ncbi.nlm.nih.gov). These were used for comparison with the

148 complete cp genomes of HA and HP.

149

RESULTS & DISCUSSION

Genome features

152 Similar to the typical cp genome structure in other higher plants, the *Haloxylon* cp genome is a double-

153 stranded, circular DNA molecule of 151,570 bp in length in HA and 151,586 bp in length in HP. It also

154 includes a large single copy region (LSC) of 84,214 bp in HA and 84,217 bp in HP and a small single copy

155 region (SSC) of 19,014 bp in HA and 19,015 bp in HP; these are separated by a pair of inverted repeats (IR)

156 (24,171 bp in HA and 24,177 bp in HP) (Fig. 1). The GC content in this IR region is 43.0% in HA and 42.7%

157 in HP, and the GC content in the LSC and SSC regions is 34.4% (LSC) and 29.7% (SSC) in HA and 34.5%

158 (LSC) and 29.7% (SSC) in HP (Table 1).

159 Among the four Amaranthaceae species included in our analyses, which represent three genera, the longest

160 cp genomes (151,570 bp for HA and 151,586 bp for HP) are 1935 bp to 1951 bp larger than the shortest one

161 (149,635 bp for *B. vulgaris* subsp. *vulgaris*) (*Li et al.*, *2014*). The size of the *S. oleracea* cp genome (150,725

162 bp) (*Schmitz-Linneweber et al.*, *2001*) is intermediate (Table 1). Notably, the cp genomes of HP and HA are

163 quite similar in size; the HP cp is only 16 bp longer than that of HA, with minor differences between them.

164 There are a total of 112 genes in the *Haloxylon* cp genome, including 78 coding genes, 18 of which are

165 duplicated genes in the IR region, 30 tRNA genes, and four ribosomal RNA genes (16S, 23S, 5S, 4.5S) (Fig. 1,

166 Table S1). Based on their predicted functions, these genes can be divided into three categories, 1) genes related

167 to transcription and translation; 2) genes related to photosynthesis; 3) genes related to the biosynthesis of

168 amino acids, fatty acids, etc., and some functionally unknown genes (Table S1). The *S. oleracea* cp also

169 contains the same 78 protein-coding genes, whereas the cp in *B. vulgaris* has 79. This species contains an

170 additional gene (*rpl*23), which is a pseudogene in the other species (Fig.1, Table S1). There are 17 genes

171 harboring introns in the cp genomes of the four Amaranthaceae species analyzed (one class I intron, *trn*L^{UUA},

172 and 16 class II introns), and two of these genes, *ycf*3 and *clp*P, contain two introns each (Table 2).

173   Several angiosperm lineages have lost introns from the *rpl*2 gene independently (*Downie et al., 1991*),

174 which could also be regarded as a characteristic feature of the core members of the Caryophyllales (*Logacheva*

175 *et al., 2008*). In each of the four Amaranthaceae cp genomes in our analysis, the *rpl*2 gene has lost its intron.

176 Some authors have proposed that intron loss is not always a dependable marker of phylogenetic relationships

177 (*Millen et al., 2001*; *Dong et al., 2013b*; *Raman & Park, 2016*), and further study, including the sampling of

178 more taxa, is needed to clarify this issue.

179

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

181 To analyze these Amaranthaceae species at the genome-level, the sequences of all the four cp genomes were

182 plotted using the VISTA program (*Frazer et al., 2004*), using the annotation of HA as a reference (Fig. 2).

183 Similar to other angiosperms, we observed that the IR region is more conserved in these species than the LSC

184 and SSC regions.

185   The expansion and contraction of the border regions between the two IR regions and the single copy region

186 have contributed to genome size variations among plant lineages (*Dong et al., 2013b*; *Goremykin et al., 2003*;

187 *Ni et al., 2016*). Therefore, we next compared the exact IR border positions and their adjacent genes among the

188 four Amaranthaceae cp genomes (Fig. 3). From these data, we see that the IRa/LSC border is generally located

189 upstream of the *trn*H^{GUG} gene. The distance between the IRa/LSC border and the *trn*H^{GUG} gene is 1 bp in the

190 *Haloxylon* cp genomes and 2 bp in *Beta* genus, with no separation in *Spinacia* (Fig. 3). The IR region is

191 expanded by 763 bp and enters the 5' end of the *ycf*1 gene in *Haloxylon* species, whereas it is expanded by

192 1427 bp and 1492 bp, respectively, in *Spinacia* and *Beta*. Except for the expansion of the *ycf*1 gene, the IR

193 region extends to the *rps*19 gene in all of four Amaranthaceae cp genomes. The *rps*19 pseudogene was not

194 observed in this study. Although there are expansions or contractions of IR regions observed among the

195   investigated species of the Amaranthaceae, they contribute little to the overall size differences in the cp

196   genomes. The exon at the 5' end of the *rps*12 gene is located in the LSC region, and the intron and 3'-end exon

197   of the gene are situated in the IR region in all four Amaranthaceae species.

198

199   **Indels and SNPs**

200   Indel and single nucleotide polymorphism (SNP) sites are important molecular features valuable for

201   development of DNA markers that are useful for plant identification and genetic analysis of population

202   structure. (*Dong et al.*, *2012, 2013a, 2013b, 2014*; *Suo et al. 2012b*, *2015*, *2016*). We detected 23 indels in the

203   cp genome sequence alignment of HA and HP, including 16 indels caused by microsatellite repeat variations

204   and seven non-microsatellite-related indels (Table 3). Most of the indel events occurred in non-coding regions

205   (21/23). A large portion of the indels related to microsatellite repeat variations are characterized by a single

206   base mutation; six insertions of this type were observed in the HA cp genome. The non-microsatellite-related

207   indels were found to contain mostly five to six variable base sites, and two insertions of this type were detected

208   in the HA cp genome.

209   Forty-four SNPs were detected in the HA and HP cp genomes (Table 4), which is considerably less than

210   what was found between the cp genomes of other closely related plant species, including *Oryza sativa* and

211   *Oryza nivara* (159 SNPs, Masood *et al.*, 2004), *Machilus yunnanensis* and *Machilus balansae* (231 SNPs,

212   *Song et al.*, *2015*), *Citrus sinensis* and *Citrus aurantiifolia* (330 SNPs, *Su et al.*, *2014*), *Panax ginseng* and

213   *Palax notoginseng* (464 SNPs, *Dong et al.*, *2014*), and *Solanum tuberosum* and *Solanum bulbocastanum* (591

214   SNPs, *Chung et al.*, *2006*). Of note, the indel and SNP mutation events in the *Haloxylon* cp genomes were not

215   randomly distributed, but rather, clustered as "hotspots" (*Shaw et al.*, *2007*; *Worberg et al.*, *2007*). It is likely

216   that such mutational dynamics created the highly variable regions in the genome (*Suo et al.*, *2012b*; *Song et al.*,

217   *2015*).

218

219   **Patterns of nucleotide substitutions**

220   Overall, the differences between the HA and HP cp genomes are minor, with a genetic distance of 0.00029

221   between them (Table 4). In total, 44 variable nucleotide sites were detected, 23 of which were found in

222   intergenic regions, six in introns, and 15 in protein-encoding regions.

223   We also found that the probability of occurrence for the various nucleotide substitutions is different,

224 depending on the mutation, as shown in Fig. 4. The most frequently occurring mutations are from A to C and

225 from T to G (12 times each); mutations from A to T and from T to A exhibited the lowest frequency (only one

226 occurrence of each). The ratio of transitions (Ts) and transversions (Tv) was 0.76 in the cp genome of

227 *Haloxylon* species.

228   In the gene-encoding regions of the HA and HP cp genomes, a total of 15 variable base sites were detected

229 in 11 protein-encoding genes. Specifically, we found one mutation in each of the following genes: *atp*A, *atp*I,

230 *mat*K, *ndh*F, *ndh*I, *psb*C, *rpo*B, *rps*15, and *rps*3. Two genes, *rpo*C2 and *ycf*1, each contained three mutation

231 sites (Table 5). These mutations included six Ts and nine Tv. Ten nonsynonymous substitutions occurred

232 simultaneously in seven genes (Table 5).

233

234 **Repeat structure feature**

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

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

237 one is a tri-nucleotide repeat, and 11 are tetra-nucleotide repeats; penta-nucleotide repeats or those containing a

238 higher number of nucleotide repeats were not detected. Among the SSR loci detected, the most frequently

239 observed repeats were A/T and AT/TA, accounting for 77.97% of the total number of SSR loci (Table 6). By

240 comparison, in the cp genomes of *M. yunnanensis* and *M. balansae*, 36 SSR loci were identified (*Song et al.*,

241 *2015*).

242

243 **Inversions**

244 Inversions are important events in the evolution of plant cp genomes. Smaller inversions are less frequent in

245 these genomes, and they are generally associated with hairpins (Fig. 5). Most inversions are found in spacers

246 and introns, and in most cases, the presence/absence of inversions is highly homoplastic during cp genome

247 evolution (*Kim & Lee, 2005*; *Catalano et al., 2009*), even at the population level (*Quandt & Stech, 2004*). A

248 sequence alignment of the *Haloxylon* cp genomes revealed that an inversion event of 14 bp and one of 3 bp

249 occur in the *petA-psbJ* intergenic region and in the *rpl16* intron, respectively. The two inverted sequences are

250 predicted to form secondary hairpin structures, with repeat sequences of 27 bp and 19 bp at the two ends,

251   respectively (Fig. 5).

252

253   **Pseudogenes**

254   Pseudogenes have been defined as nonfunctional regions of genomic DNA that originally derived from

255   functional genes (*Balakirev & Ayala, 2003*). These are evolutionary relics of functional components in the

256   genome that provide important information regarding the history of the gene and genome evolution (*Balakirev*

257   *& Ayala, 2003*; *Zou et al., 2009*; *Choi & Park, 2015*). The *rpl*22 and *rps*18 genes are putative pseudogenes in

258   the Paeoniaceae (*Dong et al., 2013b*), whereas the *atp*B gene is a pseudogene in *Aster spathulifolius*.

259   Conversely, the *rpl*22, *rps*18, and *atp*B genes are predicted to be normal and functional in the *Haloxylon*

260   species, whereas *rpl*23 is a present as a pseudogene in the *Haloxylon* cp genomes (Fig. 1 and Table S1).

261

262   ## CONCLUSIONS

263   Two *Haloxylon* cp genomes were sequenced and characterized for the first time, and we found that they share

264   the same overall organization and gene content found in most angiosperm cp genomes, including that of the

265   closely related *Spinacia* and *Beta* species. The location and distribution of repeat sequences and differing

266   nucleotide mutation sites between the two cp genomes were identified. The LSC/IRB/SSC/IRA boundary

267   regions of the Amaranthaceae cp genomes were compared, and lightly intense variations were identified within

268   the genus *Haloxylon*. The complete *Haloxylon* cp genome sequences reported here enhance the genomic

269   information available for the Amaranthaceae family and further contribute to the study of germplasm diversity.

270   These data represent a valuable source of markers for future research on *Haloxylon* population genetics.

271

272   ## ACKNOWLEDGEMENTS

273   The authors thank Prof. Borong Pan for advice and helpful discussion.

274

275   ## REFERENCES

276   **Akhani H., Edwards G., Roalson E.H. 2007.** Diversification of the old world Salsoleae *s.l.*

277   (Chenopodiaceae): molecular phylogenetic analysis of nuclear and chloroplast data sets and a revised

278   classification. *International J Plant Sci* 168: 931–956

279   **APG III. 2009.** An update of the Angiosperm Phylogeny Group classification for the orders and families of

280   flowering plants: APG III. *Botanical J Linnean Society* 161: 105–121

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

282   123–151

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

284   573–580

285   **Catalano S.A., Saidman B.O., Vilardi J.C. 2009.** Evolution of small inversions in chloroplast genome: a case

286   study from a recurrent inversion in angiosperms. *Cladistics* 25, 93–104

287   **Choi K.S., Park S.J. 2015.** The complete chloroplast genome sequence of *Aster spathulifolius*

288   (Asteraceae): genomic features and relationship with Asteraceae. *Gene* 572: 214–221

289   **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

290   complete chloroplast genome sequences of *Solanum tuberosum* and comparative analysis with *Solanaceae*

291   species identified the presence of a 241-bp deletion in cultivated potato chloroplast DNA sequence. *Plant*

292   *Cell Rep* 25: 1369–1379 DOI:10.1007/s00299-006-0196-4.

293   **Conant G.C., Wolfe K.H. 2008.** GenomeVx: simple web-based creation of editable circular chromosome

294   maps. *Bioinformatics* 24(6): 861–862.

295   **Cosner M.E., Jansen R.K., Palmer J.D., Downie S.R. 1997.** The highly rearranged chloroplast genome of

296   *Trachelium caeruleum* (Campanulaceae): multiple inversions, inverted repeat expansion and   contraction,

297   transposition, insertions/deletions, and several repeat families. *Curr Genet* 31: 419–429   DOI:

298   10.1007/s002940050225

299   **Dong W.P., Liu H., Xu C., Zuo Y.J., Chen Z.J., Zhou S.L. 2014.** A chloroplast genomic strategy for

300   designing taxon specific DNA mini-barcodes: a case study on ginsengs. *BMC Genetics* 15:138

301   **Dong W.P., Xu C., Cheng T., Lin K., Zhou S.L. 2013a.** Sequencing angiosperm plastid genomes made easy:

302   a complete set of universal primers and a case study on the phylogeny of Saxifragales. *Genome Biol Evol*

303   5(5): 989–997

304   **Dong W.P., Xu C., Cheng T., Zhou S.L. 2013b.** Complete chloroplast genome of *Sedum sarmentosum* and

305   chloroplast genome evolution in Saxifragales. *PLoS ONE* 8(10): e77965 DOI:10.1371/journal.

306   pone.0077965.

307   **Dong W.P., Liu J., Yu J., Wang L., Zhou S.L. 2012.** Highly variable chloroplast markers for evaluating

308   plant phylogeny at low taxonomic levels and for DNA barcoding. *PLoS ONE* 7(4): e35071 DOI:10.1371/

309   journal.pone.0035071.

310   **Doyle J.J., Doyle J.L., Palmer J.D. 1995.** Multiple independent losses of two genes and one intron from

311   legume chloroplast genomes. *Syst Bot* 20: 272–294 DOI:10.2307/2419496.

312   **Downie S.R., Olmstead R.G., Zurawski G., Soltis D.E., Soltis P.S., Watson J.C., Palme J.D. 1991.** Six

313   independent losses of the chloroplast DNA *rpl*2 intron in dicotyledons: molecular and phylogenetic

314   implications. *Evolution* 45: 1245–1259 DOI:10. 2307/2409731.

315   **Downie S.R., Palmer J.D. 1992.** Use of chloroplast DNA rearrangements in reconstructing plant phylogeny.

316   In Soltis P.S., D.E. Soltis, J.J. Doyle (eds) Molecular Systematics of Plants. Chapman and Hall, New York,

317   London, pp14–35

318   **Goremykin V.V., Hirsch-Ernst K.I., Wolfl S., Hellwig F.H. 2003.** Analysis of the *Amborella trichopoda*

319   chloroplast genome sequence suggests that *Amborella* is not a basal angiosperm. *Mol Biol Evol* 20: 1499–

320   1505 DOI: 10.1093/molbev/msg159.

321   **Frazer K.A., Pachter L., Poliakov A., Rubin E.M., Dubchak I. 2004.** VISTA: computational tools for

322   comparative genomics. *Nucleic Acids Res* 32: W273-W279 DOI:10.1093/nar/gkh053. PubMed: 15215394.

323   **Jansen R.K., Kaittanis C., Saski C., Lee S.B., Tomkins J., Alverson A.J., Daniell H. 2006.** Phylogenetic

324   analyses of *Vitis* (Vitaceae) based on complete chloroplast genome sequences: effects of taxon sampling and

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

326   2148-6-32.

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

328   **Kim K.J., Lee H.L. 2005.** Wide spread occurrence of small inversions in the chloroplast genomes of land

329   plants. *Molecules and Cells* 19:104–113

330   **Li H., Cao H., Cai Y.F., Wang J.H., Qu S.P., Huang X.Q. 2014.** The complete chloroplast genome sequence

331   of sugar beet (*Beta vulgaris* ssp. *vulgaris*). *Mitochondrial DNA* 25: 209–211

332   **Librado P., Rozas J. 2009.** DnaSP v5: a software for comprehensive analysis of DNA polymorphism data.

333   *Bioinformatics* 25: 1451–1452

334 **Logacheva M.D., Samigullin T.H., Dhingra A., Penin A.A. 2008.** Comparative chloroplast genomics and

335   phylogenetics of *Fagopyrum esculentum* ssp. *ancestrale* –a wild ancestor of cultivated buckwheat. *BMC*

336   *Plant Biology* 8: 59 DOI:10.1186/1471-2229-8-59 PMID: 18492277.

337 **Long Y., Zhang J., Tian X., Wu S.S., Zhang Q., Zhang J.P., Dang Z.H., Pei X.W. 2014.** De novo assembly

338   of the desert tree *Haloxylon ammodendron* (C. A. Mey.) based on RNA-Seq data provides insight into

339   drought response, gene discovery and marker identification. *BMC Genomics* 15: 1111. DOI:10.1186/1471-

340   2164-15-1111.

341 **Masood M.S., Nishikawa T., Fukuoka S., Njenga P.K., Tsudzuki T., Kadowaki K. 2004.** The complete

342   nucleotide sequence of wild rice (*Oryza nivara*) chloroplast genome: first genome wide comparative

343   sequence analysis of wild and cultivated rice. *Gene* 340: 133–139 DOI:10.1016/j.gene. 2004.06.008.

344 **Millen R.S., Olmstead R.G., Adams K.L., Palmer J.D., Lao N.T., Heggie L., Kavanaghd T.A., Hibberde**

345   **J.M., Graye J.C., Mordenf C.W., Calieg P.J., Jermiinh L.S., Wolfe K.H. 2001.** Many parallel losses of

346   *inf*A from chloroplast DNA during angiosperm evolution with multiple independent transfers to the

347   nucleus. *Plant Cell* 13: 645–658 DOI:10.1105/tpc.13.3.645.

348 **Ni L.H., Zhao Z.L., Xua H.X., Chen S.L., Dorje G. 2016.** The complete chloroplast genome of *Gentiana*

349   *straminea* (Gentianaceae), an endemic species to the Sino-Himalayan subregion. *Gene* 577: 281–288

350 **Pyankov V.I., Artyusheva E.G., Edwards G.E., Black C.C. JR, Soltis P.S. 2001.** Phylogenetic analysis of

351   tribe Salsoleae (Chenopodiaceae) based on ribosomal ITS sequences: implications for the evolution of

352   photosynthesis types. *Am J Bot* 88(7): 1189–1198

353 **Quandt D., Stech M. 2004.** Molecular evolution and phylogenetic utility of the chloroplast *trn*T-*trn*F region

354   in bryophytes. *Plant Biol* 6: 545–554

355 **Raman G., Park S. 2016.** The complete chloroplast genome sequence of Ampelopsis: gene organization,

356   comparative analysis, and phylogenetic relationships to other angiosperms. *Front Plant Sci* 7: 341 DOI:

357   10.3389/fpls.2016.00341.

358 **Rambaut A. 1996.** Se-Al: sequence alignment editor. version 2.0. Oxford: University of Oxford, Department

359   of Zoology

360 **Schmitz-Linneweber C., Maier R.M., Alcaraz J.P., Cottet A., Herrmann R.G., Mache R. 2001.** The

361   plastid chromosome of spinach (*Spinacia oleracea*): complete nucleotide sequence and gene organization.

362   *Plant Mol Biol* 45: 307–315 DOI:10.1023/A:1006478403810. PubMed: 11292076.

363   **Shaw J., Lickey E.B., Schilling E.E., Small R.L. 2007.** Comparison of whole chloroplast genome sequences

364   to choose non-coding regions for phylogenetic studies in angiosperms: the tortoise and the hare III. *Am J Bot*

365   94: 275–288 DOI:10.3732/ajb.94.3.275.

366   **Sheng Y., Zheng W.H., Quan P.K., Ma K.P. 2004.** Population genetic structure of a dominant desert tree,

367   *Haloxylon ammodendron* (Chenopodiaceae), in the southeast Gurbantunggut desert detected by RAPD

368   and ISSR markers. *Acta Botanica Sinica* 46: 675–681 (in Chinese with English Abstract)

369   **Sheng Y., Zheng W.H., Quan P.K., Ma K.P. 2005**. Genetic variation within and among populations of a

370   dominant desert tree *Haloxylon ammodendron* (Amaranthaceae) in China. *Ann Bot London* 96:

371   245–252.

372   **Song C.S., Jia K.F. 2000.** Scientific survey of Wulate *Haloxylon ammodendron* forest nature reserve (The

373   series of nature reserve). China Forestry Publishing House, Beijing, China (in Chinese with English

374   Overview)

375   **Song Y., Dong W., Liu B., Xu C., Yao X., Gao J., Corlett R.T. 2015.** Comparative analysis of complete

376   chloroplast genome sequences of two tropical trees *Machilus yunnanensis* and *Machilus balansae* in the

377   family Lauraceae. *Front. Plant Sci.* 6: 662 DOI:10.3389/fpls.2015.00662.

378   **Su H.J., Hogenhout S.A., Al-Sadi A.M., Kuo C.H. 2014.** Complete chloroplast genome sequence of omani

379   lime (*Citrus aurantiifolia*) and comparative analysis within the Rosids. *PLoS ONE* 9: e113049 DOI:

380   10.1371/journal.pone.0113049.

381   **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.**

382   Distinguishing *Haloxylon persicum* and *H. ammodendron* (*Haloxylon* Bunge, Amaranthaceae) using DNA

383   Marker. AASRI Procedia 1: 305– 310

384   **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

385   of tree peonies at micro-evolution level with hyper-variable chloroplast markers and floral traits. *Plant Cell*

386   *Reports* 31: 2199–2213

387   **Suo Z.L., Chen L.N., Pei D., Jin X.B., Zhang H.J. 2015.** A new nuclear DNA marker from ubiquitin ligase

388   gene region for genetic diversity detection of walnut germplasm resources. *Biotechnology Reports* 5: 40–45

389   **Suo Z.L., Li W.Y., Jin X.B., Zhang H.J. 2016.** A new nuclear DNA marker revealing both microsatellite

390 variations and single nucleotide polymorphic loci: a case study on classification of cultivars in

391 *Lagerstroemia indica* L. *J Microb Biochem Technol* 8: 266–271 DOI:10.4172/1948–5948.1000296.

392 **Tamura K., Stecher G., Peterson D., Filipski A., Kumar S. 2013.** MEGA6: molecular evolutionary genetics

393 analysis version 6.0. *Mol Biol Evol* 30: 2725–2729. DOI:http://dx.doi.org/10.1093/molbev/mst197.

394 **Thompson J.D., Gibson T.J., Plewniak F., Jeanmougin F., Higgins D.G. 1997.** The CLUSTAL_X windows

395 interface: flexible strategies for multiple sequence alignment aided by quality analysis tools. *Nucleic Acids*

396 *Res* 25: 4876–4882 DOI:10.1093/nar/25.24.4876. PubMed: 9396791.

397 **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

398 repeats analysis of *Haloxylon ammodendron* from seeds carried back by "Shenzhou No.4" spaceship. J

399 Northwest University 39: 259–263 (in Chinese with English Abstract)

400 **Worberg A., Quandt D., Barniske A.M., Lohne C., Hilu K.W., BorschT. 2007.** Phylogeny of basal

401 eudicots: insights from non-coding and rapidly evolving DNA. *Organ Diver Evol* 7: 55–77 DOI:10.1016/

402 j.ode.2006.08.001.

403 **Wyman S.K., Jansen R.K., Boore J.L. 2004.** Automatic annotation of organellar genomes with DOGMA.

404 *Bioinformatics* 20: 3252–3255 DOI: 10.1093/bioinformatics/bth352. PubMed: 15180927.

405 **Zhang P., Dong Y.Z., Wei Y., Hu C.Z. 2006a.** ISSR analysis of genetic diversity of *Haloxylon ammodendron*

406 (C. A. Mey.) Bunge in Xinjiang. *Acta Botanica Boreali-Occidentalia Sinica* 26: 1337–1341 (in Chinese

407 with English Abstract)

408 **Zhang P., Dong Y.Z., Wei Y., Hu C.Z. 2006b.** Analysis of genetic diversity of *Haloxylon persicum*

409 (Chenopodiaceae) in Xinjiang by ISSR. *Acta Bot Yunnaica* 28: 359–362 (in Chinese with English

410 Abstract)

411 **Zhu G.L., Mosyakin S.L., Clemants S.E. 2004.** *Haloxylon* Bunge (Chenopodiaceae). In Flora of China

412 Editorial Committee (eds) Flora of China. Sci. Press, Beijing/Missouri Botanic Garden Press, St. Louis. 5:

413 395–396

414 **Zou S.H., Lehti-Shiu M.D., Thibaud-Nissen F., Prakash T., Buell C.R., Shiu S.H. 2009**. Evolutionary and

415 expression signatures of pseudogenes in Arabidopsis and rice. *Plant Physiol* 151: 3–15

416 **Zuker M. 2003.** Mfold web server for nucleic acid folding and hybridization prediction. *Nucleic Acids*

417 *Research* 31: 3406–3415

418

Table and Figure Legends

420 **Figure 1 Representative map of the two *Haloxylon* chloroplast genomes.** Genome annotation was

421 performed using DOGMA. Genes drawn outside of the circle are transcribed clockwise, whereas those

422 represented inside the circle are transcribed counterclockwise. Small single copy (SSC), large single copy

423 (LSC), and inverted repeat (IRa, IRb) regions are indicated.

424 **Figure 2 Identity plot comparing the chloroplast genomes of four Amaranthaceae species using**

425 ***Haloxylon ammodendron* as a reference sequence.** The vertical scale indicates the percent identity, ranging

426 from 50%-100%. The horizontal axis indicates the coordinates within the chloroplast genome. Genomic

427 regions are color coded as protein-coding, rRNA, tRNA, intron, and conserved non-coding sequences (CNS).

428 Abbreviations HP: *H. persicum*; SO: *Spinacia oleracea*; BV: *Beta vulgaris* subsp. *vulgaris*.

429 **Figure 3 Comparison of the junction positions between the single copy and IR regions among four**

430   **Amaranthaceae genomes.**

431 **Figure 4 The nucleotide substitution patterns in the two *Haloxylon* chloroplast genomes.** The patterns

432   were divided into six types, as indicated by the six non-strand-specific base-substitution types (i.e.,

433   numbers of G to A and C to T sites for each respective set of associated mutation types). The *H.*

434   *ammodendron* chloroplast genome was used as a standard.

435 **Figure 5 The hairpin loops predicted to be formed by inversions in the *Haloxylon* chloroplast genomes.**

436

437 **Table 1 Summary of complete chloroplast genome features in *Haloxylon*.**

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

439 **Table 3 Indel mutation events in the chloroplast genomes of *Haloxylon ammodendron* and *H. persicum*.**

440 **Table 4 The nucleotide substitution patterns present in the two *Haloxylon* chloroplast genomes.**

441 **Table 5 Comparison of the mutational changes, number of transitions (Ts) and transversions (Tv), and**

442 **synonymous (S) and nonsynonymous (N) substitutions per protein-coding chloroplast gene in *Haloxylon***

443 ***ammodendron* and *H. persicum*.**

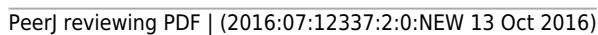
444 **Table 6 Location of repeats in the *Haloxylon ammodendron* chloroplast genome.**

445

446 **Supplemental Information**

447 Table S1 **Genes found in the *Haloxylon* chloroplast genomes.**

Representative map of the two *Haloxylon* chloroplast genomes.

Genome annotation was performed using DOGMA. Genes drawn outside of the circle are transcribed clockwise, whereas those represented inside the circle are transcribed counterclockwise. Small single copy (SSC), large single copy (LSC), and inverted repeat (IRa, IRb) regions are indicated.

Haloxylon persicum

151586bp

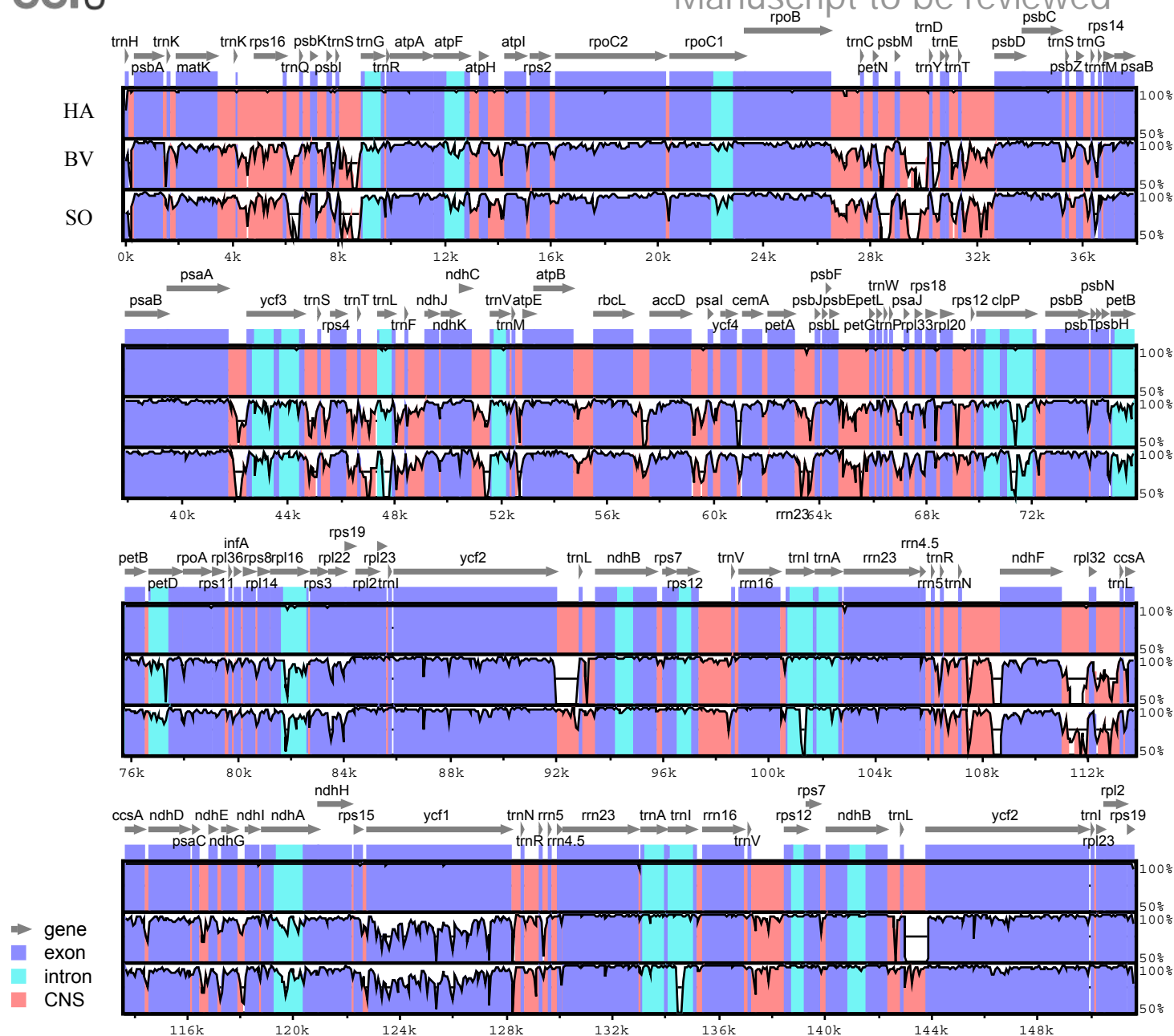
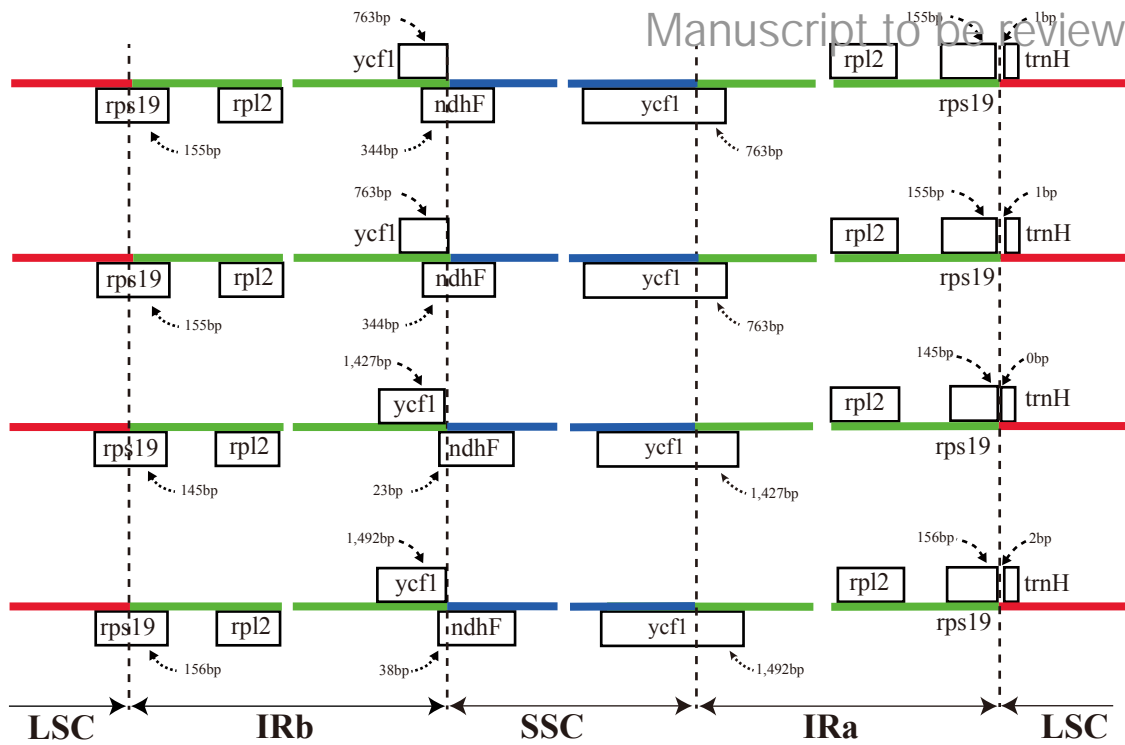
Haloxylon ammodendron

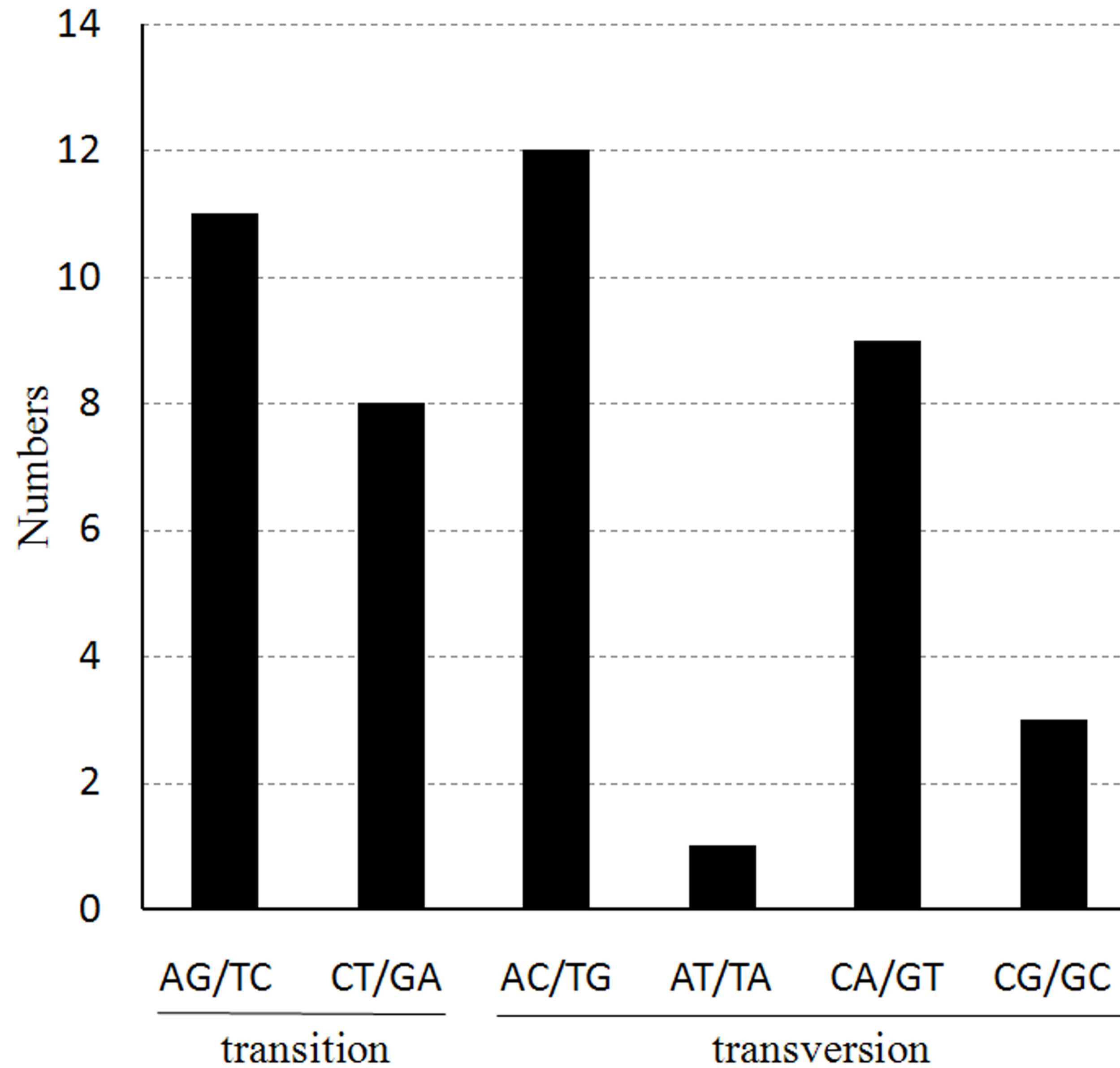
151570bp

Rubisco subnuit
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

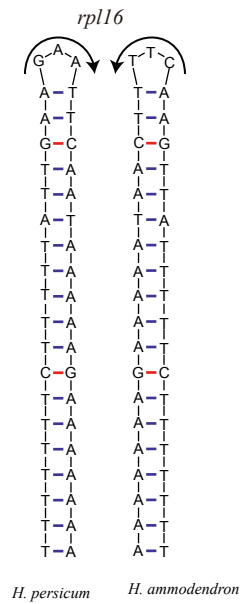
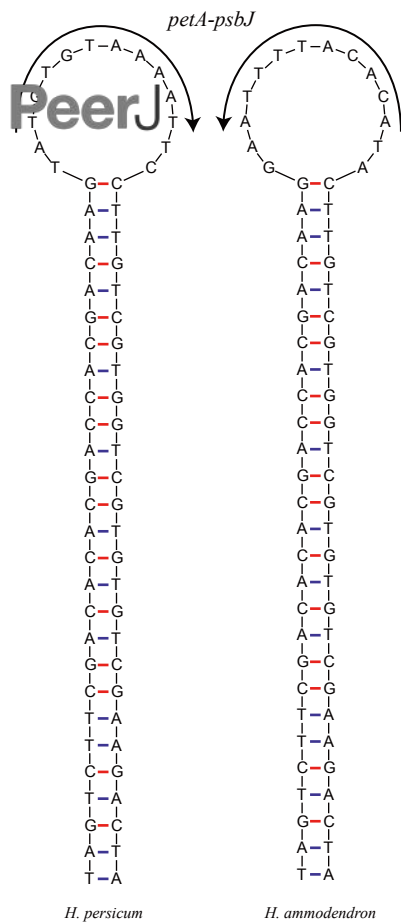
Identity plot comparing the chloroplast genomes of four Amaranthaceae species using *Haloxylon ammodendron* as a reference sequence.

The vertical scale indicates the percent identity, ranging from 50%-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*.

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

The nucleotide substitution patterns in the two *Haloxylon* chloroplast genomes.

The patterns were divided into six 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 *H. ammodendron* chloroplast genome was used as a standard.

The hairpin loops predicted to be formed by inversions in the *Haloxylon* chloroplast genomes.

petA-psbJ

rpl16

H. persicum   *H. ammodendron*

H. persicum   *H. ammodendron*

Summary of complete chloroplast genome features in *Haloxylon.*

2 **Table 1 Summary of complete chloroplast genome features in *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

Genes with introns in *Haloxylon ammodendron* and *H. persicum* and length of exons and introns.

2   **Table 2** **Genes with introns in *Haloxylon ammodendron* and *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.

Indel mutation events in the chloroplast genomes of *Haloxylon ammodendron* and *H.persicum*.

Table 3 Indel mutation events in the chloroplast genomes of *Haloxylon ammodendron* and *H. persicum*.

Region	Location	Types	HA	HP	Length (bp)	Direction [a]
*acc*D-*psa*I	Intergenic	homopolymeric indel	AA	-	2	insertion
*atp*A-*atp*F	Intergenic	homopolymeric indel	T	-	1	insertion
*atp*F	intron	homopolymeric indel	-	T	1	deletion
*ndh*I-*ndh*A	Intergenic	homopolymeric indel	-	A	1	deletion
*ndh*J-*ndh*K	Intergenic	homopolymeric indel	-	T	1	deletion
*psb*I-*trn*S	Intergenic	homopolymeric indel	-	T	1	deletion
*psb*I-*trn*S	Intergenic	homopolymeric indel	-	A	1	deletion
*rbc*L-*acc*D	Intergenic	homopolymeric indel	-	A	1	deletion
*rps*18-*rpl*20	Intergenic	homopolymeric indel	T	-	1	insertion
*trn*E-*trn*T	Intergenic	homopolymeric indel	-	A	1	deletion
*trn*K-*rps*16	Intergenic	homopolymeric indel	A	-	1	insertion
*trn*K-*rps*16	Intergenic	homopolymeric indel	A	-	1	insertion
*trn*L	intron	homopolymeric indel	-	A	1	deletion
*trn*L	intron	homopolymeric indel	A	-	1	insertion
*trn*L	intron	homopolymeric indel	-	T	1	deletion
*trn*R-*apt*A	Intergenic	homopolymeric indel	-	T	1	deletion
*atp*H-*atp*I	Intergenic	Indel	TTATT	-	5	insertion
*clp*P-*psb*B	Intergenic	Indel	-	GTCTT	5	deletion
*pet*L-*pet*G	Intergenic	Indel	-	G	1	deletion
*rpo*B-*trn*C	Intergenic	Indel	-	TGTAT	5	deletion
*rpo*B-*trn*C	Intergenic	Indel	TACAA	-	5	insertion
*rrn*23	coding	Indel	-	AATTAA	6	deletion
*rrn*23	coding	Indel	-	TTAATT	6	deletion

[a] The chloroplast genome of *H. ammodendron* was used as a standard. HA= *H. ammodendron*, HP= *H. persicum*.

The nucleotide substitution patterns present in the two *Haloxylon* chloroplast genomes.

1   **Table 4 The nucleotide substitution patterns present in the two *Haloxylon* chloroplast genomes.**

Region	Location	*H. ammodendron*	*H. persicum*
*atp*A	coding	G	A
*atp*I	coding	T	C
*mat*K	coding	C	A
*ndh*F	coding	C	T
*ndh*I	coding	G	T
*psb*C	coding	A	C
*rpo*B	coding	C	T
*rpo*C2	coding	C	A
*rpo*C2	coding	C	G
*rpo*C2	coding	G	T
*rps*15	coding	A	G
*rps*3	coding	T	G
*ycf*1	coding	A	G
*ycf*1	coding	G	C
*ycf*1	coding	G	T
*atp*B-*rbc*L	Intergenic	A	C
*atp*F-*atp*H	Intergenic	G	C
*atp*H-*atp*I	Intergenic	G	A
*ndh*F-*rpl*32	Intergenic	G	T
*psa*J-*rpl*33	Intergenic	C	T
*psa*J-*rpl*33	Intergenic	T	A
*psb*E-*pet*L	Intergenic	C	A
*psb*M-*trn*D	Intergenic	A	G
*rpl*14-*rpl*16	Intergenic	T	G
*rpl*20-*rps*12	Intergenic	G	T
*rpl*33-*rps*18	Intergenic	T	C
*rpo*A-*rps*11	Intergenic	A	G
*rpo*A-*rps*11	Intergenic	T	C
*rpo*B-*trn*C	Intergenic	G	T
*rpo*B-*trn*C	Intergenic	T	G
*rps*18-*rpl*20	Intergenic	T	G
*rps*8-*rpl*14	Intergenic	G	A
*trn*G-*trn*R	Intergenic	A	C
*trn*H-*psb*A	Intergenic	T	G
*trn*K-*mat*K	Intergenic	A	C
*trn*K-*rps*16	Intergenic	A	C
*trn*P-*psa*J	Intergenic	C	T
*trn*P-*psa*J	Intergenic	C	T
*clp*P	intron	T	G
*ndh*A	intron	T	C
*rpl*16	intron	T	C

*rps*16	intron	**T**	**G**
*trn*V	intron	**T**	**C**
*ycf*3	intron	**T**	**C**

Comparison of the mutational changes, number of transitions (Ts) and transversions (Tv), and synonymous (S) and nonsynonymous (N) substitutions per protein-coding chloroplast gene in *Haloxylon ammodendron* and *H. persicum.*

2 **Table 5** **Comparison of the mutational changes, number of transitions (Ts) and**

3 **transversions (Tv), and synonymous (S) and nonsynonymous (N) substitutions per protein-coding**

4 **chloroplast gene in *Haloxylon ammodendron* and *H. persicum*.**

Gene	Ts	Tv	S	N
*atp*A	1	0	1	0
*atp*I	1	0	1	0
*mat*K	0	1	0	1
*ndh*F	1	0	0	1
*ndh*I	0	1	0	1
*psb*C	0	1	1	0
*rpo*B	1	0	1	0
*rpo*C2	0	3	0	3
*rps*15	1	0	0	1
*rps*3	0	1	0	1
*ycf*1	1	2	1	2
Total	6	9	5	10

Location of repeats in the *Haloxylon ammodendron* chloroplast genome.

1 **Table 6 Location of repeats in the *Haloxylon ammodendron* chloroplast genome.**

No.	Location	Motif	No. of Repeats	SSR start	SSR end
1	*trn*K-*mat*K	A	11	1658	1668
2	*trn*K-*rps*16	A	12	4210	4221
3	*rps*16-*trn*Q	A	10	6461	6470
4	*trn*Q-*psb*K	A	10	6957	6966
5	*psb*K-*psb*I	A	10	7578	7587
6	*psb*I-*trn*S	A	12	7854	7865
7	*atp*F intron	A	10	12476	12485
8	*rpo*C1 intron	A	10	22386	22395
9	*trn*E-*trn*T	A	10	31169	31178
10	*trn*L-intron	A	12	47464	47475
11	*trn*F-*ndh*J	A	10	48982	48991
12	*rbc*L-*acc*D	A	12	57323	57334
13	*acc*D-*psa*I	A	10	59584	59593
14	*psb*F	A	10	64309	64318
15	*clp*P intron	A	10	71717	71726
16	*pet*B intron	A	18	75505	75522
17	*ndh*I-*ndh*A	A	10	118705	118714
18	*psa*A	C	10	40165	40174
19	*trn*K-*rps*16	T	10	4464	4473
20	*psb*I-*trn*S	T	10	7745	7754
21	*trn*R-*atp*A	T	11	9948	9958
22	*atp*A-*atp*F	T	10	11532	11541
23	*atp*F intron	T	11	12457	12467
24	*rps*2-*rpo*C2	T	11	15957	15967
25	*rps*2-*rpo*C2	T	11	18156	18166
26	*rpo*B	T	10	25865	25874
27	*trn*D-*trn*Y	T	10	30323	30332
28	*trn*L-*trn*F	T	10	48029	48038
29	*ndh*J-*ndh*K	T	10	49646	49655
30	*trn*V intron	T	15	52214	52228
31	*trn*M-*atp*E	T	10	52658	52667
32	*rbc*L-*acc*D	T	14	57377	57390
33	*pet*L-*pet*G	T	10	66141	66150
34	*psa*J-*rpl*33	T	12	67499	67510
35	*rps*18-*rpl*20	T	10	68447	68456
36	*rpo*A	T	10	78219	78228
37	*rps*11-*rpl*36	T	12	79577	79588
38	*rpl*32-*trn*L	T	11	112371	112381
39	*ndh*A intron	T	12	119581	119592
40	*ndh*A intron	T	10	119793	119802

5
6
7

No.	Location	Motif	No. of Repeats	SSR start	SSR end
41	*ycf*1	T	12	125285	125296
42	*ycf*1	T	10	125890	125899
43	*ycf*1	T	14	126895	126908
44	*ycf*1	T	10	127195	127204
45	*rps*16-*trn*Q	AT	5	6277	6286
46	*trn*S-*trn*G	AT	5	8177	8186
47	*trn*S-*trn*G	AT	5	8300	8309
48	*trn*N-*ndh*F	TAA	4	109380	109391
49	*psb*A-*trn*K	TTGT	3	1522	1533
50	*mat*K-*trn*K	TTCT	3	3873	3884
51	*atp*I-*rps*2	ATTA	3	15121	15132
52	*trn*E-*trn*Y	ATTA	3	31084	31095
53	*acc*D-*psa*I	TAAT	4	59721	59736
54	*rps*18-*rpl*20	TTTA	3	68474	68485
55	*clp*P intron	TTTC	3	71598	71609
56	*rrn*23	AGGT	3	104481	104492
57	*trn*L-*ccs*A	AACC	3	113312	113323
58	*ycf*1	TAAT	3	124297	124308
59	*rrn*23	CTAC	3	131310	131321