



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

Wenpan Dong^{1,2}, Chao Xu^{1,3}, Delu Li⁴, Xiaobai Jin⁵, Ruili Li⁵, Qi Lu⁶ and Zhili Suo¹

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 University of Chinese Academy of Sciences, Beijing, China

⁴ Gansu Desert Control Research Institute, Gansu, China

⁵ Beijing Botanical Garden, Institute of Botany, Chinese Academy of Sciences, Beijing, China

⁶ Institute of Desertification Studies, Chinese Academy of Forestry, Beijing, China

ABSTRACT

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.

Submitted 2 August 2016
Accepted 18 October 2016
Published 10 November 2016

Corresponding authors
Qi Lu, luqi@caf.ac.cn
Zhili Suo, zlsuo@ibcas.ac.cn

Academic editor
Thiago Venancio

Additional Information and
Declarations can be found on
page 16

DOI 10.7717/peerj.2699

© Copyright
2016 Dong et al.

Distributed under
Creative Commons CC-BY 4.0

OPEN ACCESS

Subjects Biodiversity, Biotechnology, Genetics, Genomics, Plant Science

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

INTRODUCTION

The eudicot clade comprises approximately 75% of all flowering land plant species, including major subclades: rosids, asterids, Saxifragales, Santalales, and Caryophyllales (APG III, 2009). *Haloxylon* species, which include psammophytic small trees or shrubs, are positioned phylogenetically in the Amaranthaceae Juss of the Caryophyllales Perleb among core eudicots (APG III, 2009; Pyankov et al., 2001; Akhani, Edwards & Roalson, 2007). The *Haloxylon* genus has about 11 species, with a distribution from the Mediterranean through Central Asia and into China (Zhu, Mosyakin & Clemants, 2004). Two *Haloxylon* species, which are known as the King of psammophytic plants, are found in the deserts of northwest China and, play important roles in environmental protection, including wind control and sand fixation (Zhu, Mosyakin & Clemants, 2004; Jia & Lu, 2004). These precious psammophytic woody plants can adapt to harsh environmental conditions, such as drought, desert, high temperature, and sand storms. However, populations of *Haloxylon* plants have been threatened in China in past decades as a result of decreased underground water, overgrazing, and over exploitation of agriculture.

Because of the environmental significance of these plants and their declining numbers, genetic research on *Haloxylon* germplasm resources has garnered significant interest (Song & Jia, 2000; Sheng et al., 2004; Sheng et al., 2005; Zhang et al., 2006a; Zhang et al., 2006b). However, *Haloxylon* plants possess only fine green assimilating shoots, without leaves, making the evaluation of their phenotypic diversity difficult. Further, the detection of genetic diversity within *Haloxylon* germplasm resources has been slowed by a lack of morphological markers (Sheng et al., 2004; Sheng et al., 2005; Zhang et al., 2006a; Zhang et al., 2006b; Wang et al., 2009; Suo et al., 2012a). A recent study by Long et al. (2014) used RNA-seq data to elucidate the *Haloxylon* transcriptome, providing a valuable sequence resource for further genetic and genomic studies; however, genetic information for members of the *Haloxylon* genus, and how they might differ from one another, is limited.

Each leaf cell of plants contains 1,000–10,000 chloroplasts (cp), which are key organelles for photosynthesis and other biochemical pathways such as the biosynthesis of starch, fatty acids, pigments, and amino acids (Dong et al., 2013b; Raman & 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 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; Dong et al., 2013a; Dong et al., 2013b; Dong et al., 2014; Ni et al., 2016; Suo et al., 2012b).

There are two published complete cp genome sequences (*Spinacia oleracea* and *Beta vulgaris* subsp. *vulgaris*) from members of the Amaranthaceae family (Li et al., 2014; Schmitz-Linneweber et al., 2001). However, the determination of the cp genome from *Haloxylon* plants is of further significance for potentially enhancing our understanding of their adaptability to severe desert environmental conditions, and their genomic evolution within the Amaranthaceae. Here, we report the complete cp genomes from two *Haloxylon* species, *H. ammodendron* and *H. persicum*, including patterns of nucleotide substitutions,

microstructural mutation, and simple sequence repeats (SSRs). We further performed genomic comparative analyses on these and two other representative Amaranthaceae species, to better understand the evolutionary relationships within this family.

MATERIALS & METHODS

Sampling and DNA extraction

Fresh young shoots of *H. ammodendron* (HA) and *H. persicum* (HP) were collected in May 2011 from Minqin Eremophytes Botanical Garden (N38°34', E102°59', Altitude 1,378 m), Gansu Province, China (under the leadership of Gansu Desert Control Research Institute, Lanzhou, Gansu, China). These HA and HP plants were originally introduced from the Turpan Desert Botanical Garden of Chinese Academy of Sciences, Xinjiang Uygur Autonomous Region. The shoots from each accession were immediately dried using silica gel for future DNA extraction. Total genomic DNA was extracted from each using the Plant Genomic DNA Kit (DP305) from Tiangen Biotech (Beijing) Co., Ltd., China. The approval numbers are 2012BAD16B0101 and 80117B1001 for field permit of the research.

Chloroplast genome sequencing

The HA and HP cp genomes were sequenced using the short-range PCR method reported by [Dong et al. \(2012\)](#), [Dong et al. \(2013a\)](#) and [Dong et al. \(2013b\)](#). The PCR protocol was as follows: preheating at 94 °C for 4 min, 34 cycles at 94 °C for 45 s, annealing at 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 amplification was performed in an Applied Biosystems Veriti™ 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 in both the forward and reverse directions using a 3730xl DNA analyzer (Applied Biosystems, Foster City, CA, USA). DNA regions containing poly structures or that were difficult to amplify were further sequenced using newly designed primer pairs for confirming reliable and high quality sequencing results.

Chloroplast genome assembling and annotation

The cp DNA sequences were manually confirmed and assembled using Sequencher (v4.6) software, and cp genome annotation was performed using the Dual Organellar Genome Annotator (DOGMA) ([Wyman, Jansen & Boore, 2004](#)). BLASTX and BLASTN searches were utilized to accurately annotate the protein-encoding genes and to identify the locations of the transfer RNAs (tRNAs) and ribosomal RNAs (rRNAs). Gene annotation information from other closely related plant species was also used for confirmation when the boundaries of the introns or exons could not be precisely determined because of the limited power of BLAST in cp genome annotation (e.g., for some short exons of 6–9 nt in length, such as in the case of *rps16*, *petB*, and *petD*). Promoter, intron, and exon boundaries, as well as the location of stop codons for all protein-encoding genes, have been identified accurately. The cp genome map was drawn using Genome Vx software ([Conant & Wolfe, 2008](#)) (<http://wolfe.ucd.ie/GenomeVx/>), and the cp genome sequences have been deposited to GenBank with the following accession numbers: [KF534478](#) for HA and [KF534479](#) for HP (<https://www.ncbi.nlm.nih.gov/nuccore/?term=Haloxylon+chloroplast+genome>).

Repeat structure analysis

Gramene **Simple Sequence Repeat Identification Tool** software (<http://www.gramene.org/db/markers/ssrtool>) (Benson, 1999) was utilized to search for simple sequence repeat loci in the cp genome sequences, with the threshold value of repeat number as ≥ 10 for mono-nucleotide repeats, ≥ 5 for di-nucleotide repeats, ≥ 4 for tri-nucleotide repeats, and ≥ 3 for tetra-nucleotide, penta-nucleotide, or hexa-nucleotide repeats.

Gene content analysis and comparative genomics

The mVISTA program was employed in Shuffle-LAGAN mode (Frazer *et al.*, 2004) to compare the complete HA and HP cp genomes. These were aligned using MUSCLE software (Thompson *et al.*, 1997) and were manually adjusted 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), and the genetic distance (*p*-distance) was computed using MEGA 6.0 software (Tamura *et al.*, 2013). Based on the aligned sequence matrix, the micro-structure events were checked manually and were further divided into three categories: (i) microsatellite-related insertions/deletions (indels), (ii) non-microsatellite-related indels, (iii) and inverted sequences. Using the HA cp genome sequence as the standard reference, the size, location, and evolutionary direction of the microstructure events were counted. The proposed secondary structures 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 *S. oleracea* (GenBank accession number [AJ400848.1](#), *Spinacia* L.) (Schmitz-Linneweber *et al.*, 2001) and *B. vulgaris* subsp. *vulgaris* (GenBank accession number [KJ081864.1](#), *Beta vulgaris* subsp. *vulgaris*) (Li *et al.*, 2014), two closely related species in the Amaranthaceae family, were downloaded from GenBank databases (www.ncbi.nlm.nih.gov). These were used for comparison with the complete cp genomes of HA and HP.

RESULTS & DISCUSSION

Genome features

Similar to the 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 in HA and 151,586 bp in length in HP. It also includes a large single copy region (LSC) of 84,214 bp in HA and 84,217 bp in HP and a small single copy region (SSC) of 19,014 bp in HA and 19,015 bp in HP; these are separated by a pair of inverted repeats (IR) (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% 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% (LSC) and 29.7% (SSC) in HP (Table 1).

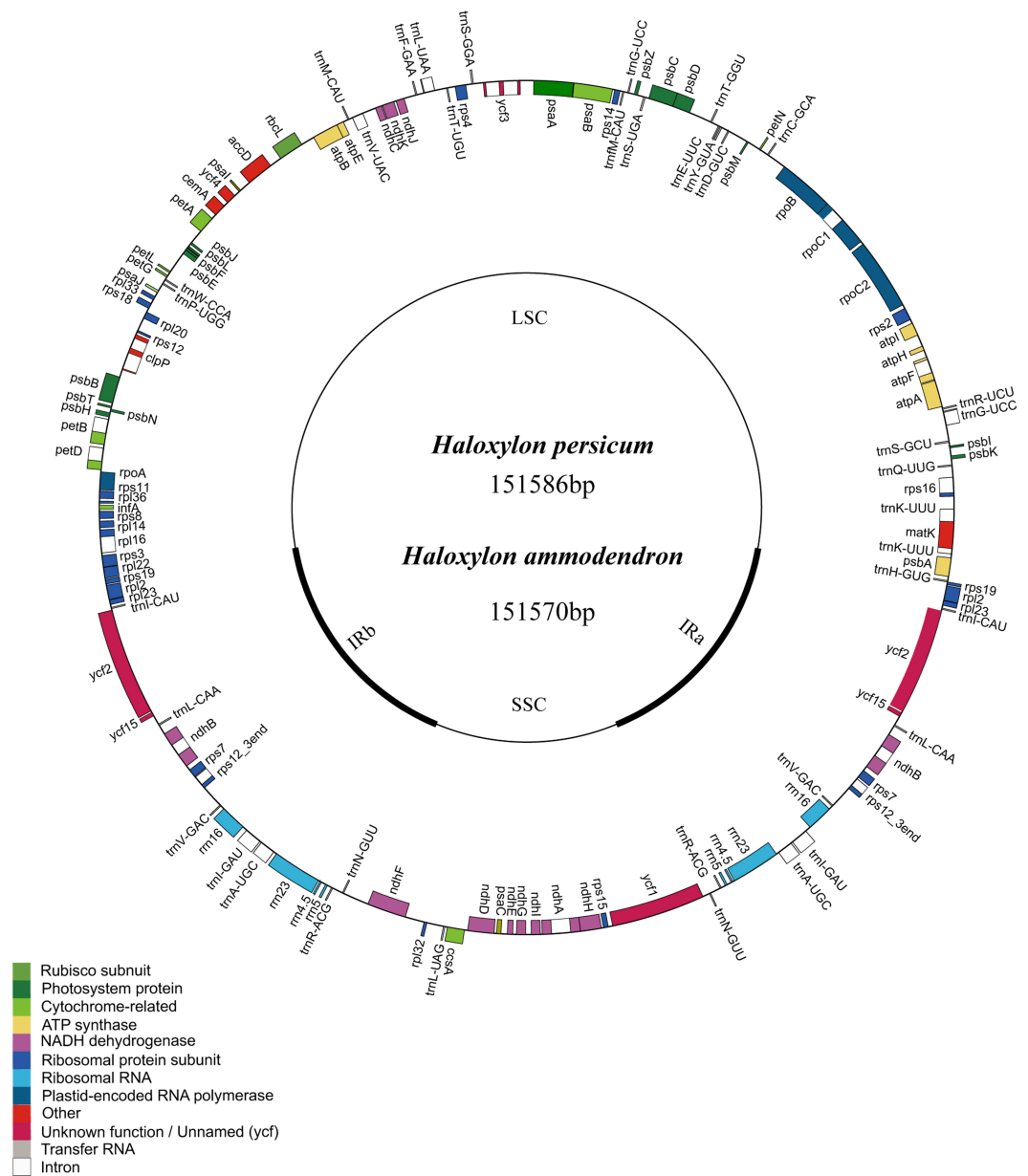


Figure 1 Representative map of the two *Haloxylon* chloroplast genomes. Genome annotation was performed using DOGMA. Genes drawn outside of the circle are transcribed clockwise, whereas those represented inside the circle are transcribed counter-clockwise. Small single copy (SSC), large single copy (LSC), and inverted repeat (IRA, IRb) regions are indicated.

Among the four Amaranthaceae species included in our analyses, which represent three genera, the longest cp genomes (151,570 bp for HA and 151,586 bp for HP) are 1,935 bp to 1,951 bp larger than the shortest one (149,635 bp for *B. vulgaris* subsp. *vulgaris*) ([Li et al., 2014](#)). The size of the *S. oleracea* cp genome (150,725 bp) ([Schmitz-Linneweber et al., 2001](#)) is intermediate ([Table 1](#)). Notably, the cp genomes of HP and HA are quite similar in size; the HP cp is only 16 bp longer than that of HA, with minor differences between them.

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

	<i>H. ammodendron</i>	<i>H. persicum</i>	<i>Spinacia oleracea</i>	<i>Beta vulgaris</i>
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

There are a total of 112 genes in the *Haloxylon* cp genome, including 78 coding genes, 18 of which are duplicated genes in the IR region, 30 tRNA genes, and four ribosomal RNA genes (16S, 23S, 5S, 4.5S) (Fig. 1 and Table S1). Based on their predicted functions, these genes can be divided into three categories, (1) genes related to transcription and translation; (2) genes related to photosynthesis; (3) genes related to the biosynthesis of amino acids, fatty acids, etc., and some functionally unknown genes (Table S1). The *S. oleracea* cp also contains the same 78 protein-coding genes, whereas the cp in *B. vulgaris* has 79. This species contains an additional gene (*rpl23*), which is a pseudogene in the other species (Fig. 1 and Table S1). There are 17 genes harboring introns in the cp genomes of the four Amaranthaceae species analyzed (one class I intron, *trnL-UUA*, and 16 class II introns), and two of these genes, *yef3* and *clpP*, contain two introns each (Table 2).

Several angiosperm lineages have lost introns from the *rpl2* gene independently (Downie et al., 1991), which could also be regarded as a characteristic feature of the core members of the Caryophyllales (Logacheva et al., 2008). In each of the four Amaranthaceae cp genomes in our analysis, the *rpl2* gene has lost its intron. Some authors have proposed that intron loss is not always a dependable marker of phylogenetic relationships (Millen et al., 2001; Dong et al., 2013b; Raman & Park, 2016), and further study, including the sampling of more taxa, is needed to clarify this issue.

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 cp genomes were plotted using the VISTA program (Frazer et al., 2004), using the annotation of HA as a reference (Fig. 2). Similar to other angiosperms, we observed that 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 next compared the exact

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

Notes.

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

IR border positions and their adjacent genes among the four Amaranthaceae cp genomes (Fig. 3). From these data, we see that the IRa/LSC border is generally located upstream of the *trnH*-GUG gene. The distance between the IRa/LSC border and the *trnH*-GUG gene is 1 bp in the *Haloxylon* cp genomes and 2 bp in *Beta* genus, with no separation in *Spinacia* (Fig. 3). The IR region is expanded by 763 bp and enters the 5' end of the *ycf1* gene in *Haloxylon* species, whereas it is expanded by 1,427 bp and 1,492 bp, respectively, in *Spinacia* and *Beta*. Except for the expansion of the *ycf1* gene, the IR region extends to the *rps19* gene in all of four Amaranthaceae cp genomes. The *rps19* 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 exon at the 5' end of the *rps12* gene is located in the LSC region, and the intron and 3'-end exon of the gene are situated in the IR region in all four Amaranthaceae species.

Indels and SNPs

Indel and single nucleotide polymorphism (SNP) sites are important molecular features valuable for development of DNA markers that are useful for plant identification and genetic analysis of population structure (Dong et al., 2012; Dong et al., 2013a; Dong et al., 2013b; Dong et al., 2014; Suo et al., 2012b; Suo et al., 2015; Suo et al., 2016). We detected 23 indels in the cp genome sequence alignment of HA and HP, including 16 indels caused by microsatellite repeat variations and seven non-microsatellite-related indels (Table 3). Most

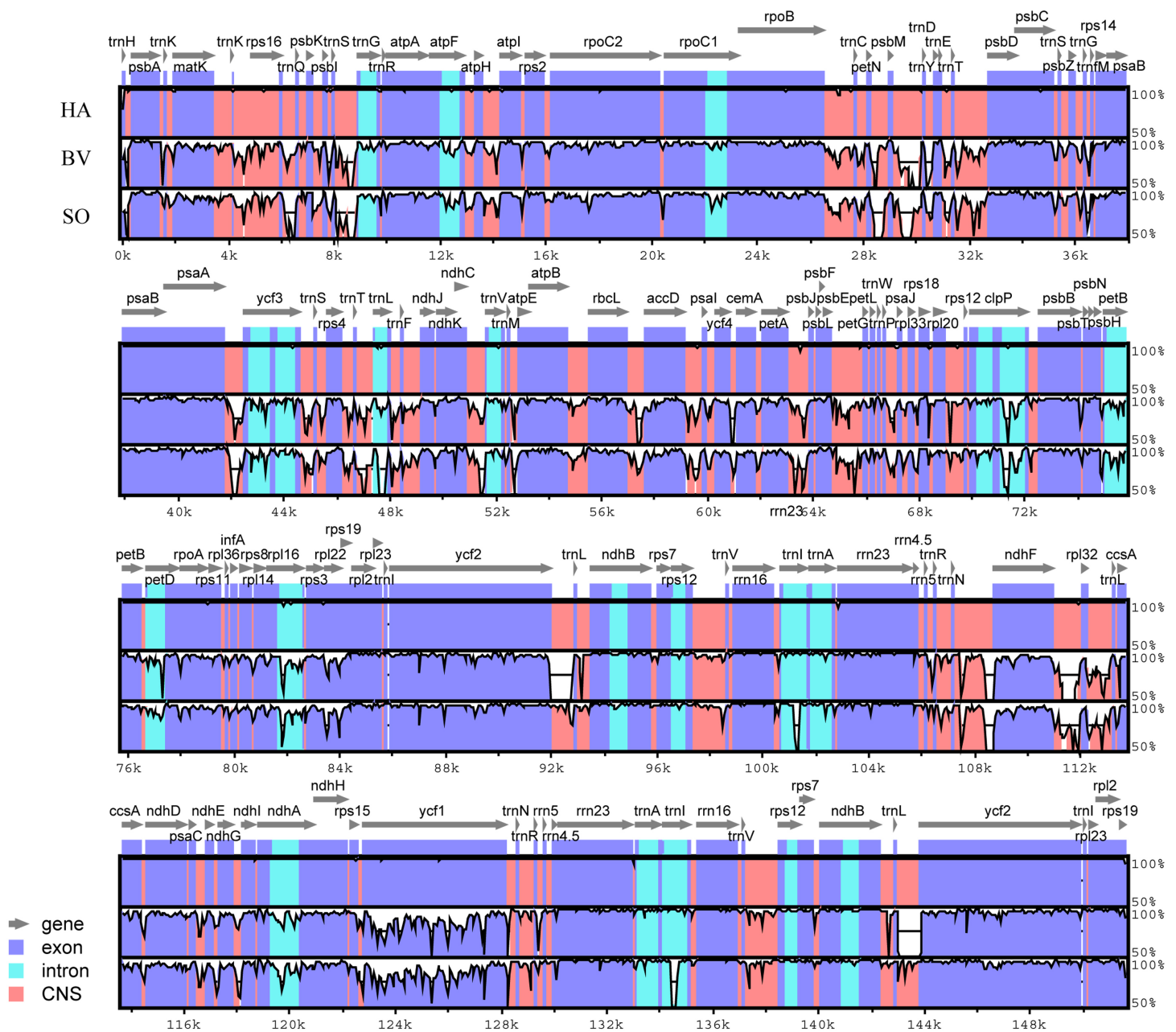


Figure 2 Identity plot comparing the chloroplast genomes of four Amaranthaceae species using *Haloxylon ammodendron* as a reference sequence. The vertical scale indicates the percent identity, ranging from 50% to 100%. The horizontal axis indicates the coordinates within the chloroplast genome. Genomic 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*.

of the indel events occurred in non-coding regions (21/23). A large portion of the indels related to microsatellite repeat variations are characterized by a single base mutation; six insertions of this type were observed in the HA cp genome. The non-microsatellite-related indels were found to contain mostly five to six variable base sites, and two insertions of this type were detected in the HA cp genome.

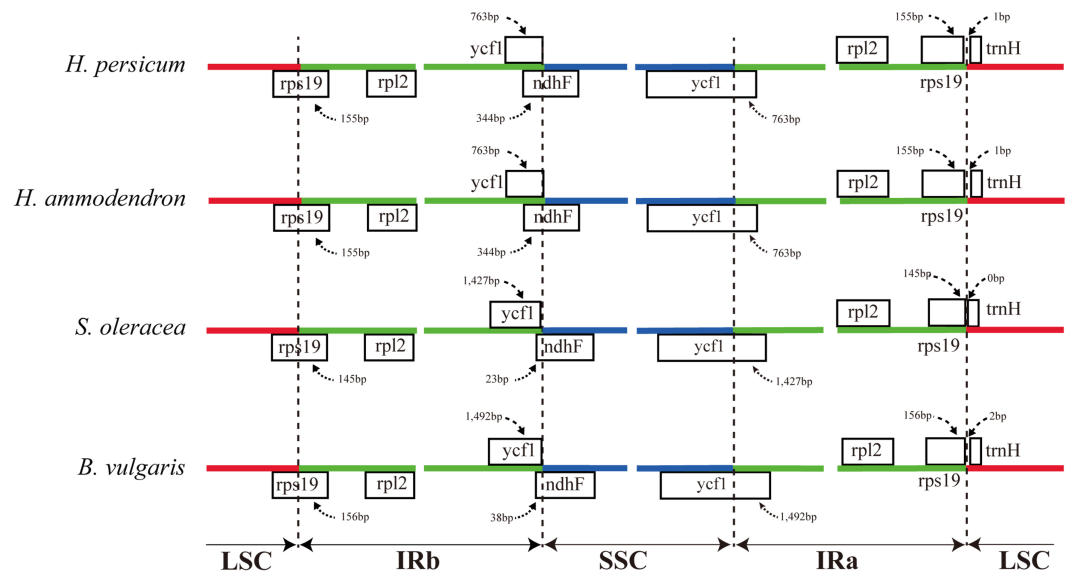


Figure 3 Comparison of the junction positions between the single copy and IR regions among four Amaranthaceae genomes.

Forty-four SNPs were detected in the HA and HP cp genomes (Table 4), which is considerably less than what was found between the cp genomes of other closely related plant species, including *Oryza sativa* and *Oryza nivara* (159 SNPs, Masood et al., 2004), *Machilus yunnanensis* and *Machilus balansae* (231 SNPs, Song et al., 2015), *Citrus sinensis* and *Citrus aurantiifolia* (330 SNPs, Su et al., 2014), *Panax ginseng* and *Palax notoginseng* (464 SNPs, Dong et al., 2014), and *Solanum tuberosum* and *Solanum bulbocastanum* (591 SNPs, Chung et al., 2006). Of note, the indel and SNP mutation events in the *Haloxylon* cp genomes were not randomly distributed, but rather, clustered as “hotspots” (Shaw et al., 2007; Worberg et al., 2007). It is likely that such mutational dynamics created the highly variable regions in the genome (Suo et al., 2012b; Song et al., 2015).

Patterns of nucleotide substitutions

Overall, the differences between the HA and HP cp genomes are minor, with a genetic distance of 0.00029 between them (Table 4). In total, 44 variable nucleotide sites were detected, 23 of which were found in intergenic regions, six in introns, and 15 in protein-encoding regions.

We also found that the probability of occurrence for the various nucleotide substitutions is different, depending on the mutation, as shown in Fig. 4. The most frequently occurring mutations are from A to C and from T to G (12 times each); mutations from A to T and from T to A exhibited the lowest frequency (only one occurrence of each). The ratio of transitions (Ts) and transversions (Tv) was 0.76 in the cp genome of *Haloxylon* species.

In the gene-encoding regions of the HA and HP cp genomes, a total of 15 variable base sites were detected in 11 protein-encoding genes. Specifically, we found one mutation in each of the following genes: *atpA*, *atpI*, *matK*, *ndhF*, *ndhI*, *psbC*, *rpoB*, *rps15*, and *rps3*. Two genes, *rpoC2* and *ycf1*, each contained three mutation sites (Table 5). These mutations

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

Region	Location	Types	HA	HP	Length (bp)	Direction ^a
<i>accD-psaI</i>	Intergenic	Homopolymeric indel	AA	–	2	Insertion
<i>atpA-atpF</i>	Intergenic	Homopolymeric indel	T	–	1	Insertion
<i>atpF</i>	Intron	Homopolymeric indel	–	T	1	Deletion
<i>ndhI-ndhA</i>	Intergenic	Homopolymeric indel	–	A	1	Deletion
<i>ndhJ-ndhK</i>	Intergenic	Homopolymeric indel	–	T	1	Deletion
<i>psbI-trnS</i>	Intergenic	Homopolymeric indel	–	T	1	Deletion
<i>psbI-trnS</i>	Intergenic	Homopolymeric indel	–	A	1	Deletion
<i>rbcL-accD</i>	Intergenic	Homopolymeric indel	–	A	1	Deletion
<i>rps18-rpl20</i>	Intergenic	Homopolymeric indel	T	–	1	Insertion
<i>trnE-trnT</i>	Intergenic	Homopolymeric indel	–	A	1	Deletion
<i>trnK-rps16</i>	Intergenic	Homopolymeric indel	A	–	1	Insertion
<i>trnK-rps16</i>	Intergenic	Homopolymeric indel	A	–	1	Insertion
<i>trnL</i>	Intron	Homopolymeric indel	–	A	1	Deletion
<i>trnL</i>	Intron	Homopolymeric indel	A	–	1	Insertion
<i>trnL</i>	Intron	Homopolymeric indel	–	T	1	Deletion
<i>trnR-aptA</i>	Intergenic	Homopolymeric indel	–	T	1	Deletion
<i>atpH-atpI</i>	Intergenic	Indel	TTTATT	–	5	Insertion
<i>clpP-psbB</i>	Intergenic	Indel	–	GTCTT	5	Deletion
<i>petL-petG</i>	Intergenic	Indel	–	G	1	Deletion
<i>rpoB-trnC</i>	Intergenic	Indel	–	TGTATT	5	Deletion
<i>rpoB-trnC</i>	Intergenic	Indel	TACAA	–	5	Insertion
<i>rrn23</i>	Coding	Indel	–	AATTAA	6	Deletion
<i>rrn23</i>	Coding	Indel	–	TTAATT	6	Deletion

Notes.

^aThe chloroplast genome of *H. ammodendron* was used as a standard.
HA, *H. ammodendron*; HP, *H. persicum*.

included six Ts and nine Tv. Ten nonsynonymous substitutions occurred simultaneously in seven genes (Table 5).

Repeat structure feature

Simple sequence repeats (SSRs) are also called microsatellites. Within the cp genomes of HA and HP, 59 different SSR loci were detected. Of these, 44 loci are mono-nucleotide repeats, three are di-nucleotide repeats, one is a tri-nucleotide repeat, and 11 are tetra-nucleotide repeats; penta-nucleotide repeats or those containing a higher number of nucleotide repeats were not detected. Among the SSR loci detected, the most frequently observed repeats were A/T and AT/TA, accounting for 77.97% of the total number of SSR loci (Table 6). By comparison, in the cp genomes of *M. yunnanensis* and *M. balansae*, 36 SSR loci were identified (Song et al., 2015).

Inversions

Inversions are important events in the evolution of plant cp genomes. Smaller inversions are less frequent in these genomes, and they are generally associated with hairpins (Fig. 5). Most inversions are found in spacers and introns, and in most cases, the presence/absence of

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

Region	Location	<i>H. ammodendron</i>	<i>H. persicum</i>
<i>atpA</i>	Coding	G	A
<i>atpI</i>	Coding	T	C
<i>matK</i>	Coding	C	A
<i>ndhF</i>	Coding	C	T
<i>ndhI</i>	Coding	G	T
<i>psbC</i>	Coding	A	C
<i>rpoB</i>	Coding	C	T
<i>rpoC2</i>	Coding	C	A
<i>rpoC2</i>	Coding	C	G
<i>rpoC2</i>	Coding	G	T
<i>rps15</i>	Coding	A	G
<i>rps3</i>	Coding	T	G
<i>ycf1</i>	Coding	A	G
<i>ycf1</i>	Coding	G	C
<i>ycf1</i>	Coding	G	T
<i>atpB-rbcL</i>	Intergenic	A	C
<i>atpF-atpH</i>	Intergenic	G	C
<i>atpH-atpI</i>	Intergenic	G	A
<i>ndhF-rpl32</i>	Intergenic	G	T
<i>psaJ-rpl33</i>	Intergenic	C	T
<i>psaJ-rpl33</i>	Intergenic	T	A
<i>psbE-petL</i>	Intergenic	C	A
<i>psbM-trnD</i>	Intergenic	A	G
<i>rpl14-rpl16</i>	Intergenic	T	G
<i>rpl20-rps12</i>	Intergenic	G	T
<i>rpl33-rps18</i>	Intergenic	T	C
<i>rpoA-rps11</i>	Intergenic	A	G
<i>rpoA-rps11</i>	Intergenic	T	C
<i>rpoB-trnC</i>	Intergenic	G	T
<i>rpoB-trnC</i>	Intergenic	T	G
<i>rps18-rpl20</i>	Intergenic	T	G
<i>rps8-rpl14</i>	Intergenic	G	A
<i>trnG-trnR</i>	Intergenic	A	C
<i>trnH-psbA</i>	Intergenic	T	G
<i>trnK-matK</i>	Intergenic	A	C
<i>trnK-rps16</i>	Intergenic	A	C
<i>trnP-psaJ</i>	Intergenic	C	T
<i>trnP-psaJ</i>	Intergenic	C	T
<i>clpP</i>	Intron	T	G
<i>ndhA</i>	Intron	T	C

(continued on next page)

Table 4 (continued)

Region	Location	<i>H. ammodendron</i>	<i>H. persicum</i>
<i>rpl16</i>	Intron	T	C
<i>rps16</i>	Intron	T	G
<i>trnV</i>	Intron	T	C
<i>ycf3</i>	Intron	T	C

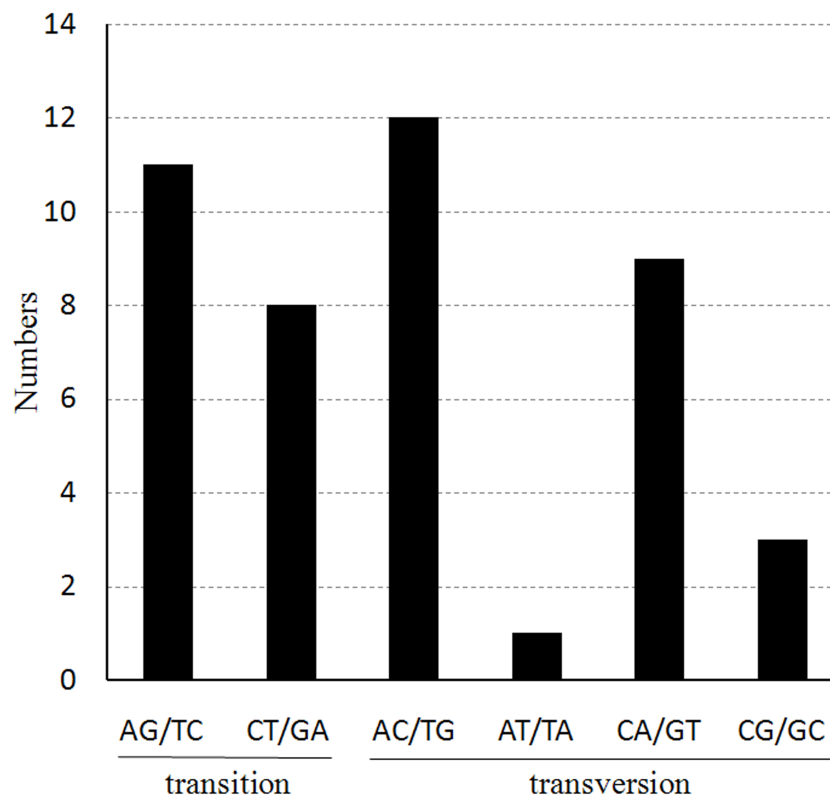


Figure 4 The nucleotide substitution patterns in the two *Haloxylon* chloroplast genomes. The patterns were divided into six types, as indicated by the six non-strand-specific base-substitution types (i.e., numbers of G to A and C to T sites for each respective set of associated mutation types). The *H. ammodendron* chloroplast genome was used as a standard.

inversions is highly homoplastic during cp genome evolution (Kim & Lee, 2005; Catalano, Saidman & Vilardi, 2009), even at the population level (Quandt & Stech, 2004). A sequence alignment of the *Haloxylon* cp genomes revealed that an inversion event of 14 bp and one of 3 bp occur in the *petA-psbJ* intergenic region and in the *rpl16* intron, respectively. The two inverted sequences are predicted to form secondary hairpin structures, with repeat sequences of 27 bp and 19 bp at the two ends, respectively (Fig. 5).

Pseudogenes

Pseudogenes have been defined as nonfunctional regions of genomic DNA that originally derived from functional genes (Balakirev & Ayala, 2003). These are evolutionary relics of functional components in the genome that provide important information regarding the

Table 5 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*.

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

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

No.	Location	Motif	No. of repeats	SSR start	SSR end
1	<i>trnK-matK</i>	A	11	1,658	1,668
2	<i>trnK-rps16</i>	A	12	4,210	4,221
3	<i>rps16-trnQ</i>	A	10	6,461	6,470
4	<i>trnQ-psbK</i>	A	10	6,957	6,966
5	<i>psbK-psbI</i>	A	10	7,578	7,587
6	<i>psbI-trnS</i>	A	12	7,854	7,865
7	<i>atpF</i> intron	A	10	12,476	12,485
8	<i>rpoC1</i> intron	A	10	22,386	22,395
9	<i>trnE-trnT</i>	A	10	31,169	31,178
10	<i>trnL</i> -intron	A	12	47,464	47,475
11	<i>trnF-ndhJ</i>	A	10	48,982	48,991
12	<i>rbcL-accD</i>	A	12	57,323	57,334
13	<i>accD-psaI</i>	A	10	59,584	59,593
14	<i>psbF</i>	A	10	64,309	64,318
15	<i>clpP</i> intron	A	10	71,717	71,726
16	<i>petB</i> intron	A	18	75,505	75,522
17	<i>ndhI-ndhA</i>	A	10	118,705	118,714
18	<i>psaA</i>	C	10	40,165	40,174
19	<i>trnK-rps16</i>	T	10	4,464	4,473
20	<i>psbI-trnS</i>	T	10	7,745	7,754
21	<i>trnR-atpA</i>	T	11	9,948	9,958
22	<i>atpA-atpF</i>	T	10	11,532	11,541
23	<i>atpF</i> intron	T	11	12,457	12,467
24	<i>rps2-rpoC2</i>	T	11	15,957	15,967

(continued on next page)

Table 6 (continued)

No.	Location	Motif	No. of repeats	SSR start	SSR end
25	<i>rps2-rpoC2</i>	T	11	18,156	18,166
26	<i>rpoB</i>	T	10	25,865	25,874
27	<i>trnD-trnY</i>	T	10	30,323	30,332
28	<i>trnL-trnF</i>	T	10	48,029	48,038
29	<i>ndhJ-ndhK</i>	T	10	49,646	49,655
30	<i>trnV</i> intron	T	15	52,214	52,228
31	<i>trnM-atpE</i>	T	10	52,658	52,667
32	<i>rbcL-accD</i>	T	14	57,377	57,390
33	<i>petL-petG</i>	T	10	66,141	66,150
34	<i>psaJ-rpl33</i>	T	12	67,499	67,510
35	<i>rps18-rpl20</i>	T	10	68,447	68,456
36	<i>rpoA</i>	T	10	78,219	78,228
37	<i>rps11-rpl36</i>	T	12	79,577	79,588
38	<i>rpl32-trnL</i>	T	11	11,2371	11,2381
39	<i>ndhA</i> intron	T	12	119,581	119,592
40	<i>ndhA</i> intron	T	10	119,793	119,802
41	<i>ycf1</i>	T	12	125,285	125,296
42	<i>ycf1</i>	T	10	125,890	125,899
43	<i>ycf1</i>	T	14	126,895	126,908
44	<i>ycf1</i>	T	10	127,195	127,204
45	<i>rps16-trnQ</i>	AT	5	6,277	6,286
46	<i>trnS-trnG</i>	AT	5	8,177	8,186
47	<i>trnS-trnG</i>	AT	5	8,300	8,309
48	<i>trnN-ndhF</i>	TAA	4	109,380	109,391
49	<i>psbA-trnK</i>	TTGT	3	1,522	1,533
50	<i>matK-trnK</i>	TTCT	3	3,873	3,884
51	<i>atpI-rps2</i>	ATTA	3	15,121	15,132
52	<i>trnE-trnY</i>	ATTA	3	31,084	31,095
53	<i>accD-psaI</i>	TAAT	4	59,721	59,736
54	<i>rps18-rpl20</i>	TTTA	3	68,474	68,485
55	<i>clpP</i> intron	TTTC	3	71,598	71,609
56	<i>rrn23</i>	AGGT	3	104,481	104,492
57	<i>trnL-ccsA</i>	AACC	3	113,312	113,323
58	<i>ycf1</i>	TAAT	3	124,297	124,308
59	<i>rrn23</i>	CTAC	3	131,310	131,321

history of the gene and genome evolution (*Balakirev & Ayala, 2003; Zou et al., 2009; Choi & Park, 2015*). The *rpl22* and *rps18* genes are putative pseudogenes in the Paeoniaceae (*Dong et al., 2013b*), whereas the *atpB* gene is a pseudogene in *Aster spathulifolius*. Conversely, the *rpl22*, *rps18*, and *atpB* genes are predicted to be normal and functional in the *Haloxylon* species, whereas *rpl23* is present as a pseudogene in the *Haloxylon* cp genomes (*Fig. 1* and *Table S1*).

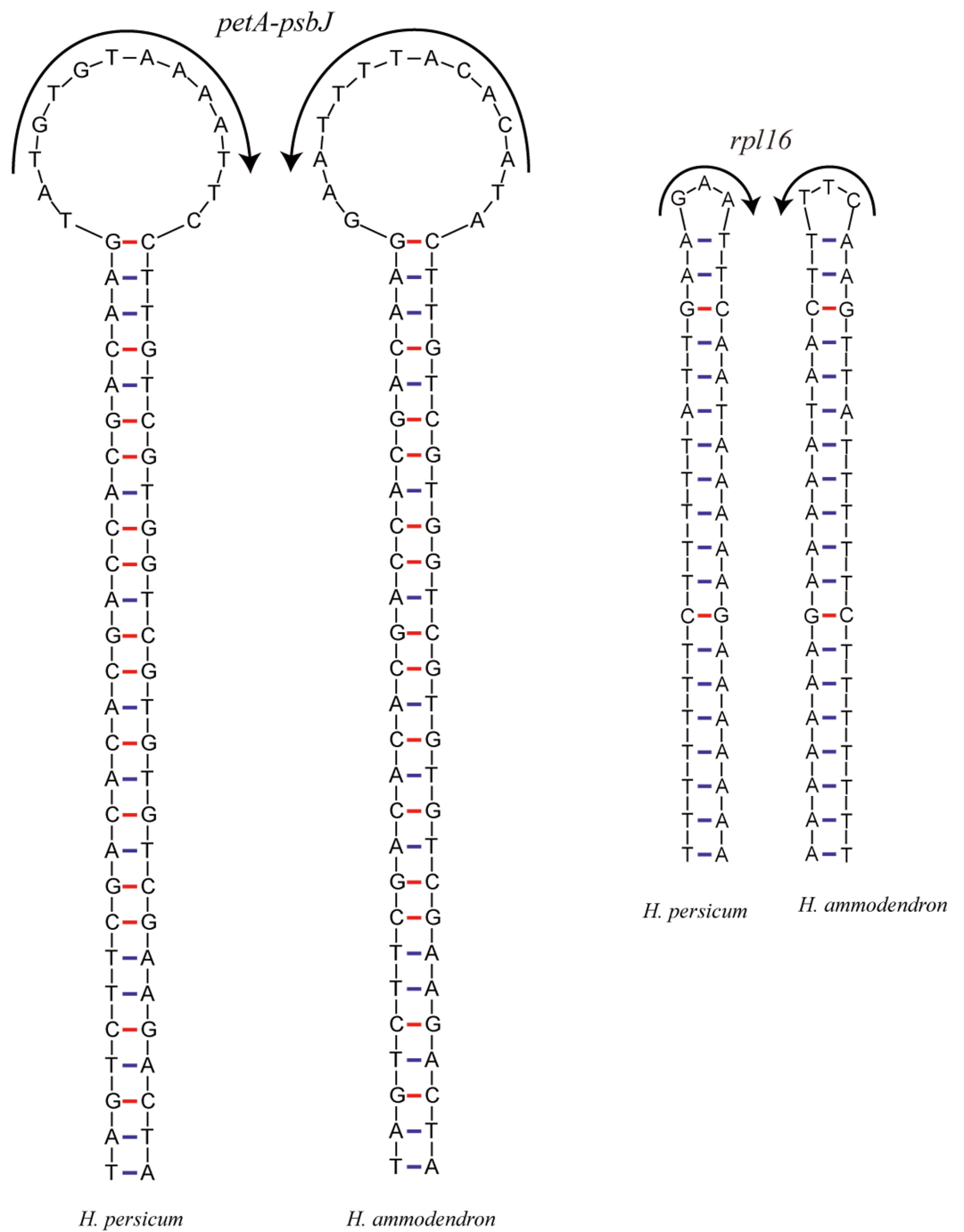


Figure 5 The hairpin loops predicted to be formed by inversions in the *Haloxylon* chloroplast genomes.

CONCLUSIONS

Two *Haloxylon* cp genomes were sequenced and characterized for the first time, and we found that they share the same overall organization and gene content found in most angiosperm cp genomes, including that of the closely related *Spinacia* and *Beta* species. The location and distribution of repeat sequences and differing nucleotide mutation sites between the two cp genomes were identified. The LSC/IRB/SSC/IRA boundary regions of the Amaranthaceae cp genomes were compared, and lightly intense variations were identified within the genus *Haloxylon*. The complete *Haloxylon* cp genome sequences reported here enhance the genomic information available for the Amaranthaceae family and further contribute to the study of germplasm diversity. These data represent a valuable source of markers for future research on *Haloxylon* population genetics.

ACKNOWLEDGEMENTS

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

ADDITIONAL INFORMATION AND DECLARATIONS

Funding

This study was financially supported by the China National Science and Technology Plan Project (2012BAD16B0101), the “948 Program” (No. 2008-4-47), the Identification and Screening of *Haloxylon* Germplasm Resources (80117B1001), the 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), the joint projects No. 70009C1020 and 70009C1036, and the National Forest Genetic Resources Platform (2005DKA21003). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Grant Disclosures

The following grant information was disclosed by the authors:

China National Science and Technology Plan Project: 2012BAD16B0101.

948 Program: 2008-4-47.

Identification and Screening of *Haloxylon* Germplasm Resources: 80117B1001.

State Forestry Administration “Public Welfare Research Foundation”: 201004010.

Chinese Ministry of Science and Technology: 2006BAD26B0101.

National Natural Science Foundation of China: 30972412.

Joint projects: 70009C1020, 70009C1036.

National Forest Genetic Resources Platform: 2005DKA21003.

Competing Interests

The authors declare there are no competing interests.

Author Contributions

- Wenpan Dong conceived and designed the experiments, performed the experiments, analyzed the data, wrote the paper, prepared figures and/or tables, reviewed drafts of the paper.
- Chao Xu analyzed the data, wrote the paper, prepared figures and/or tables, reviewed drafts of the paper.
- Delu Li contributed reagents/materials/analysis tools, reviewed drafts of the paper.
- Xiaobai Jin wrote the paper, reviewed drafts of the paper.
- Ruili Li prepared figures and/or tables, reviewed drafts of the paper.
- Qi Lu conceived and designed the experiments, contributed reagents/materials/analysis tools, wrote the paper, reviewed drafts of the paper.
- Zhili Suo conceived and designed the experiments, performed the experiments, analyzed the data, contributed reagents/materials/analysis tools, wrote the paper, reviewed drafts of the paper.

Field Study Permissions

The following information was supplied relating to field study approvals (i.e., approving body and any reference numbers):

Minqin Eremophytes Botanical Garden, Gansu Province, China (under the leadership of Gansu Desert Control Research Institute, Gansu, China).

The approval numbers are 2012BAD16B0101 and 80117B1001.

DNA Deposition

The following information was supplied regarding the deposition of DNA sequences:

The chloroplast genome sequences have been deposited to GenBank with the following accession numbers: [KF534478](#) for *Haloxylon ammodendron* and [KF534479](#) for *Haloxylon persicum*.

Supplemental Information

Supplemental information for this article can be found online at <http://dx.doi.org/10.7717/peerj.2699#supplemental-information>.

REFERENCES

- Akhani H, Edwards G, Roalson EH. 2007.** Diversification of the old world Salsoleae *s.l.* (Chenopodiaceae): molecular phylogenetic analysis of nuclear and chloroplast data sets and a revised classification. *International Journal of Plant Sciences* **168**:931–956 DOI [10.1086/518263](#).
- APG III. 2009.** An update of the Angiosperm Phylogeny Group classification for the orders and families of flowering plants: APG III. *Botanical Journal of the Linnean Society* **161**:105–121 DOI [10.1111/j.1095-8339.2009.00996.x](#).
- Balakirev ES, Ayala FJ. 2003.** Pseudogenes: are they junk or functional DNA? *Annual Review of Genetics* **37**:123–151 DOI [10.1146/annurev.genet.37.040103.103949](#).

- Benson G. 1999.** Tandem repeats finder: a program to analyze DNA sequences. *Nucleic Acid Research* **27**(2):573–580 DOI [10.1093/nar/27.2.573](https://doi.org/10.1093/nar/27.2.573).
- Catalano SA, Saidman BO, Vilardi JC. 2009.** Evolution of small inversions in chloroplast genome: a case study from a recurrent inversion in angiosperms. *Cladistics* **25**:93–104 DOI [10.1111/j.1096-0031.2008.00236.x](https://doi.org/10.1111/j.1096-0031.2008.00236.x).
- Choi KS, Park SJ. 2015.** The complete chloroplast genome sequence of *Aster spathulifolius* (Asteraceae): genomic features and relationship with Asteraceae. *Gene* **572**:214–221 DOI [10.1016/j.gene.2015.07.020](https://doi.org/10.1016/j.gene.2015.07.020).
- Chung HJ, Jung JD, Park HW, Kim JH, Cha HW, Min SR, Jeong WJ, Liu JR. 2006.** The complete chloroplast genome sequences of *Solanum tuberosum* and comparative analysis with *Solanaceae* species identified the presence of a 241-bp deletion in cultivated potato chloroplast DNA sequence. *Plant Cell Reports* **25**:1369–1379 DOI [10.1007/s00299-006-0196-4](https://doi.org/10.1007/s00299-006-0196-4).
- Conant GC, Wolfe KH. 2008.** GenomeVx: simple web-based creation of editable circular chromosome maps. *Bioinformatics* **24**(6):861–862 DOI [10.1093/bioinformatics/btm598](https://doi.org/10.1093/bioinformatics/btm598).
- Dong WP, Liu H, Xu C, Zuo YJ, Chen ZJ, Zhou SL. 2014.** A chloroplast genomic strategy for designing taxon specific DNA mini-barcodes: a case study on ginsengs. *BMC Genetics* **15**:138 DOI [10.1186/s12863-014-0138-z](https://doi.org/10.1186/s12863-014-0138-z).
- Dong WP, Liu J, Yu J, Wang L, Zhou SL. 2012.** Highly variable chloroplast markers for evaluating plant phylogeny at low taxonomic levels and for DNA barcoding. *PLoS ONE* **7**(4):e35071 DOI [10.1371/journal.pone.0035071](https://doi.org/10.1371/journal.pone.0035071).
- Dong WP, Xu C, Cheng T, Lin K, Zhou SL. 2013a.** Sequencing angiosperm plastid genomes made easy: a complete set of universal primers and a case study on the phylogeny of Saxifragales. *Genome Biology and Evolution* **5**(5):989–997 DOI [10.1093/gbe/evt063](https://doi.org/10.1093/gbe/evt063).
- Dong WP, Xu C, Cheng T, Zhou SL. 2013b.** Complete chloroplast genome of *Sedum sarmentosum* and chloroplast genome evolution in Saxifragales. *PLoS ONE* **8**(10):e77965 DOI [10.1371/journal.pone.0077965](https://doi.org/10.1371/journal.pone.0077965).
- Downie SR, Olmstead RG, Zurawski G, Soltis DE, Soltis PS, Watson JC, Palme JD. 1991.** Six independent losses of the chloroplast DNA *rpl2* intron in dicotyledons: molecular and phylogenetic implications. *Evolution* **45**:1245–1259 DOI [10.2307/2409731](https://doi.org/10.2307/2409731).
- Frazer KA, Pachter L, Poliakov A, Rubin EM, Dubchak I. 2004.** VISTA: computational tools for comparative genomics. *Nucleic Acids Research* **32**:W273–W279 DOI [10.1093/nar/gkh458](https://doi.org/10.1093/nar/gkh458).
- Goremykin VV, Hirsch-Ernst KI, Wolf S, Hellwig FH. 2003.** Analysis of the *Amborella trichopoda* chloroplast genome sequence suggests that *Amborella* is not a basal angiosperm. *Molecular Biology and Evolution* **20**:1499–1505 DOI [10.1093/molbev/msg159](https://doi.org/10.1093/molbev/msg159).
- Jia ZQ, Lu Q. 2004.** *Haloxylon* Bunge. Beijing: China Environmental Science Press (in Chinese).

- Kim KJ, Lee HL. 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 YF, Wang JH, Qu SP, Huang XQ. 2014.** The complete chloroplast genome sequence of sugar beet (*Beta vulgaris* ssp. *vulgaris*). *Mitochondrial DNA* **25**:209–211 DOI [10.3109/19401736.2014.883611](https://doi.org/10.3109/19401736.2014.883611).
- Librado P, Rozas J. 2009.** DnaSP v5: a software for comprehensive analysis of DNA polymorphism data. *Bioinformatics* **25**:1451–1452 DOI [10.1093/bioinformatics/btp187](https://doi.org/10.1093/bioinformatics/btp187).
- Logacheva MD, Samigullin TH, Dhingra A, Penin AA. 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](https://doi.org/10.1186/1471-2229-8-59).
- Long Y, Zhang J, Tian X, Wu SS, Zhang Q, Zhang JP, Dang ZH, Pei XW. 2014.** De novo assembly of the desert tree *Haloxylon ammodendron* (CA Mey.) based on RNA-Seq data provides insight into drought response, gene discovery and marker identification. *BMC Genomics* **15**:1111 DOI [10.1186/1471-2164-15-1111](https://doi.org/10.1186/1471-2164-15-1111).
- Masood MS, Nishikawa T, Fukuoka S, Njenga PK, Tsudzuki T, Kadowaki K. 2004.** The complete nucleotide sequence of wild rice (*Oryza nivara*) chloroplast genome: first genome wide comparative sequence analysis of wild and cultivated rice. *Gene* **340**:133–139 DOI [10.1016/j.gene.2004.06.008](https://doi.org/10.1016/j.gene.2004.06.008).
- Millen RS, Olmstead RG, Adams KL, Palmer JD, Lao NT, Heggie L, Kavanaghd TA, Hibberde JM, Graye JC, Mordenf CW, Calieg PJ, Jermini LS, Wolfe KH. 2001.** Many parallel losses of *inf A* from chloroplast DNA during angiosperm evolution with multiple independent transfers to the nucleus. *The Plant Cell* **13**:645–658 DOI [10.1105/tpc.13.3.645](https://doi.org/10.1105/tpc.13.3.645).
- Ni LH, Zhao ZL, Xua HX, Chen SL, Dorje G. 2016.** The complete chloroplast genome of *Gentiana straminea* (Gentianaceae), an endemic species to the Sino-Himalayan subregion. *Gene* **577**:281–288 DOI [10.1016/j.gene.2015.12.005](https://doi.org/10.1016/j.gene.2015.12.005).
- Pyankov VI, Artyusheva EG, Edwards GE, Black CCJR, Soltis PS. 2001.** Phylogenetic analysis of tribe Salsoleae (Chenopodiaceae) based on ribosomal ITS sequences: implications for the evolution of photosynthesis types. *American Journal of Botany* **88**(7):1189–1198 DOI [10.2307/3558329](https://doi.org/10.2307/3558329).
- Quandt D, Stech M. 2004.** Molecular evolution and phylogenetic utility of the chloroplast *trnT-trnF* region in bryophytes. *Plant Biology* **6**:545–554 DOI [10.1055/s-2004-821144](https://doi.org/10.1055/s-2004-821144).
- Raman G, Park S. 2016.** The complete chloroplast genome sequence of Ampelopsis: gene organization, comparative analysis, and phylogenetic relationships to other angiosperms. *Frontiers in Plant Science* **7**:341 DOI [10.3389/fpls.2016.00341](https://doi.org/10.3389/fpls.2016.00341).
- Rambaut A. 1996.** *Se-Al: sequence alignment editor*. Version 2.0. Oxford: University of Oxford, Department of Zoology.
- Schmitz-Linneweber C, Maier RM, Alcaraz JP, Cottet A, Herrmann RG, Mache R. 2001.** The plastid chromosome of spinach (*Spinacia oleracea*): complete nucleotide sequence and gene organization. *Plant Molecular Biology* **45**:307–315 DOI [10.1023/A:1006478403810](https://doi.org/10.1023/A:1006478403810).

- Shaw J, Lickey EB, Schilling EE, Small RL. 2007.** Comparison of whole chloroplast genome sequences to choose non-coding regions for phylogenetic studies in angiosperms: the tortoise and the hare III. *American Journal of Botany* **94**:275–288 DOI [10.3732/ajb.94.3.275](https://doi.org/10.3732/ajb.94.3.275).
- Sheng Y, Zheng WH, Quan PK, Ma KP. 2004.** Population genetic structure of a dominant desert tree, *Haloxylon ammodendron* (Chenopodiaceae), in the southeast Gurbantunggut desert detected by RAPD and ISSR markers. *Acta Botanica Sinica* **46**:675–681 (in Chinese with English Abstract).
- Sheng Y, Zheng WH, Quan PK, Ma KP. 2005.** Genetic variation within and among populations of a dominant desert tree *Haloxylon ammodendron* (Amaranthaceae) in China. *Annals of Botany* **96**:245–252 DOI [10.1093/aob/mci171](https://doi.org/10.1093/aob/mci171).
- Song Y, Dong W, Liu B, Xu C, Yao X, Gao J, Corlett RT. 2015.** Comparative analysis of complete chloroplast genome sequences of two tropical trees *Machilus yunnanensis* and *Machilus balansae* in the family Lauraceae. *Frontiers in Plant Science* **6**:662 DOI [10.3389/fpls.2015.00662](https://doi.org/10.3389/fpls.2015.00662).
- Song CS, Jia KF. 2000.** *Scientific survey of Wulate Haloxylon ammodendron forest nature reserve (The series of nature reserve)*. Beijing: China Forestry Publishing House (in Chinese with English Overview).
- Su HJ, Hogenhout SA, Al-Sadi AM, Kuo CH. 2014.** Complete chloroplast genome sequence of omani lime (*Citrus aurantiifolia*) and comparative analysis within the Rosids. *PLoS ONE* **9**:e113049 DOI [10.1371/journal.pone.0113049](https://doi.org/10.1371/journal.pone.0113049).
- Suo ZL, Chen LN, Pei D, Jin XB, Zhang HJ. 2015.** A new nuclear DNA marker from ubiquitin ligase gene region for genetic diversity detection of walnut germplasm resources. *Biotechnology Reports* **5**:40–45 DOI [10.1016/j.btre.2014.11.003](https://doi.org/10.1016/j.btre.2014.11.003).
- Suo ZL, Jia ZQ, Lu Q, Pan BR, Jin XB, Xu G, Peng XQ, Sun HB, Tao YH. 2012a.** Distinguishing *Haloxylon persicum* and *H. ammodendron* (*Haloxylon* Bunge, Amaranthaceae) using DNA Marker. *AASRI Procedia* **1**:305–310.
- Suo ZL, Li WY, Jin XB, Zhang HJ. 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. *Journal of Microbial & Biochemical Technology* **8**:266–271 DOI [10.4172/1948-5948.1000296](https://doi.org/10.4172/1948-5948.1000296).
- Suo ZL, Zhang CH, Zheng YQ, He LX, Jin XB, Hou BX, Li JJ. 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 DOI [10.1007/s00299-012-1330-0](https://doi.org/10.1007/s00299-012-1330-0).
- Tamura K, Stecher G, Peterson D, Filipski A, Kumar S. 2013.** MEGA6: molecular evolutionary genetics analysis version 6.0. *Molecular Biology and Evolution* **30**:2725–2729 DOI [10.1093/molbev/mst197](https://doi.org/10.1093/molbev/mst197).
- Thompson JD, Gibson TJ, Plewniak F, Jeanmougin F, Higgins DG. 1997.** The CLUSTAL_X windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools. *Nucleic Acids Research* **25**:4876–4882 DOI [10.1093/nar/25.24.4876](https://doi.org/10.1093/nar/25.24.4876).

- Wang XM, Yang DY, Tian YZ, Zhang BW, Tu PF, Sun QS, Li XB. 2009.** Inter-simple sequence repeats analysis of *Haloxylon ammodendron* from seeds carried back by Shenzhou No. 4 spaceship. *Journal of Northwest University* **39**:259–263 (in Chinese with English Abstract).
- Worberg A, Quandt D, Barniske AM, Lohne C, Hilu KW, Borsch T. 2007.** Phylogeny of basal eudicots: insights from non-coding and rapidly evolving DNA. *Organisms Diversity & Evolution* **7**:55–77 DOI [10.1016/j.ode.2006.08.001](https://doi.org/10.1016/j.ode.2006.08.001).
- Wyman SK, Jansen RK, Boore JL. 2004.** Automatic annotation of organellar genomes with DOGMA. *Bioinformatics* **20**:3252–3255 DOI [10.1093/bioinformatics/bth352](https://doi.org/10.1093/bioinformatics/bth352).
- Zhang P, Dong YZ, Wei Y, Hu CZ. 2006a.** ISSR analysis of genetic diversity of *Haloxylon ammodendron* (CA Mey.) Bunge in Xinjiang. *Acta Botanica Boreali-Occidentalia Sinica* **26**:1337–1341 (in Chinese with English Abstract).
- Zhang P, Dong YZ, Wei Y, Hu CZ. 2006b.** Analysis of genetic diversity of *Haloxylon persicum* (Chenopodiaceae) in Xinjiang by ISSR. *Acta Botanica Yunnanica* **28**:359–362 (in Chinese with English Abstract).
- Zhu GL, Mosyakin SL, Clemants SE. 2004.** *Haloxylon* Bunge (Chenopodiaceae). In: Flora of China Editorial Committee, ed. *Flora of China*. Vol. 5. St. Louis: Science Press, Beijing/Missouri Botanic Garden Press, 395–396.
- Zou SH, Lehti-Shiu MD, Thibaud-Nissen F, Prakash T, Buell CR, Shiu SH. 2009.** Evolutionary and expression signatures of pseudogenes in *Arabidopsis* and rice. *Plant Physiology* **151**:3–15 DOI [10.1104/pp.109.140632](https://doi.org/10.1104/pp.109.140632).
- Zuker M. 2003.** Mfold web server for nucleic acid folding and hybridization prediction. *Nucleic Acids Research* **31**:3406–3415 DOI [10.1093/nar/gkg595](https://doi.org/10.1093/nar/gkg595).