

Molecular Structure and Phylogenetic Analyses of Complete Chloroplast Genomes of Two *Aristolochia* Medicinal Species

Jianguo Zhou , Hui Yao , Xinlian Chen , Ying Li , Jingyuan Song , Shilin Chen , Yonghua Li , Yingxian Cui
Institute of Medicinal Plant Development, Chinese Academy of Medical Sciences & Peking Union Medical College, Beijing 100193, China

Abstract

The family Aristolochiaceae, comprising about 600 species of eight genera, is a unique plant family containing aristolochic acids (AAs). The complete chloroplast genome sequences of *Aristolochia debilis* and *Aristolochia contorta* are reported here. The results show that the complete chloroplast genomes of *A. debilis* and *A. contorta* comprise circular 159,793 and 160,576 bp-long molecules, respectively and have typical quadripartite structures. The GC contents of both species were 38.3% each. A total of 131 genes were identified in each genome including 85 protein-coding genes, 37 tRNA genes, eight rRNA genes and one pseudogene (*ycf1*). The simple-sequence repeat sequences mainly comprise A/T mononucleotide repeats. Phylogenetic analyses using maximum parsimony (MP) revealed that *A. debilis* and *A. contorta* had a close phylogenetic relationship with species of the family Piperaceae, as well as Laurales and Magnoliales. The data obtained in this study will be beneficial for further investigations on *A. debilis* and *A. contorta* from the aspect of evolution, and chloroplast genetic engineering.

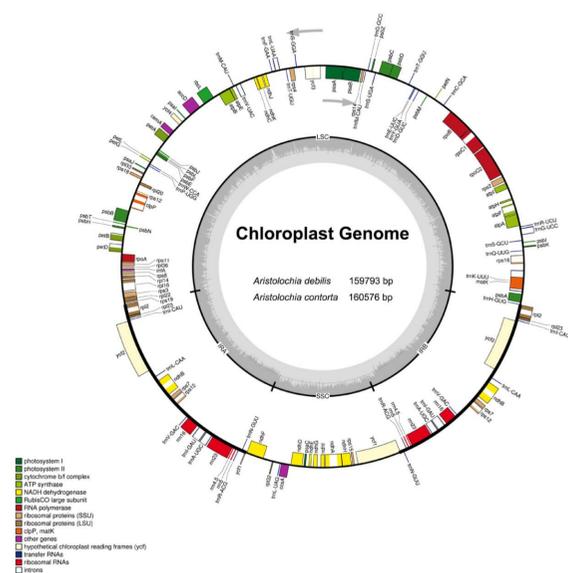
Introduction

The traditional Chinese medicine plants, *Aristolochia debilis* and *Aristolochia contorta*, are herbaceous climbers in the family Aristolochiaceae. These two species have been recorded as traditional herbal medicines which can clear lung-heat to stop coughing and activate meridians to stop pain. Modern pharmacology studies have shown that the primary chemical constituents of the two species are aristolochic acid analogues including aristolochic acids (AAs) and aristolactams (ALs).

Results

The complete chloroplast genome of *A. debilis* is a circular molecule of 159,793 bp in length comprising a large single-copy (LSC) region of 89,609 bp and a small single-copy (SSC) region of 19,834 bp separated by a pair of inverted repeats (IRs), each 25,175 bp in length. The complete chloroplast genome of *A. contorta* is 160,576 bp in length, which is divided into one LSC (89,781 bp), one SSC (19,877 bp) and two IRs, each 25,459 bp in length.

A total of 131 genes were identified from each genome including 85 protein-coding genes, 37 tRNAs, eight rRNAs, and one pseudogene (*ycf1*). The functional *ycf1* copy existed encompassing IR-SSC boundary and the other pseudogene *ycf1* copy was on the other IR region. Six protein-coding genes, seven tRNA genes, and all rRNA genes were duplicated in the IR regions. Coding regions including protein-coding genes (CDS), tRNAs, and rRNAs constituted 56.7% and 56.4% in the chloroplast genomes of *A. debilis* and *A. contorta*, respectively; while the non-coding regions including introns, pseudogenes, and intergenic spacers constituted 43.3% and 43.6% of the genome, respectively.



Gene maps of the complete chloroplast genomes of *A. debilis* and *A. contorta*.

Genes on the inside of the circle are transcribed clockwise, while those outside are transcribed counter clockwise. The darker gray in the inner circle corresponds to GC content, whereas the lighter gray corresponds to AT content.

References

- Schmeiser, R.H.; Janssen, J.W.; Lyons, J.; et al. Aristolochic acid activates RAS genes in rat tumors at deoxyadenosine residues. *Cancer Res.* 1990, 50, 5464–5469.
- Chen, L.; Mei, N.; Yao, L.; et al. Mutations induced by carcinogenic doses of aristolochic acid in kidney of big blue transgenic rats. *Toxicol. Lett.* 2006, 165, 250–256.
- Grollman, A.P.; Shibutani, S.; Moriya, M.; et al. Aristolochic acid and the etiology of endemic (Balkan) nephropathy. *Proc. Natl. Acad. Sci. USA* 2007, 104, 12129–12134.
- Lord, G.M.; Tagore, R.; Cook, T.; et al. Nephropathy caused by Chinese herbs in the UK. *Lancet* 1999, 354, 481–482.
- Wang, Y.; Zhan, D.F.; Jia, X.; et al. Complete chloroplast genome sequence of *Aquilaria sinensis* (lour.) gilg and evolution analysis within the Malvales order. *Front. Plant Sci.* 2016, 7, 280.
- Shinozaki, K.; Ohme, M.; Tanaka, M.; et al. The complete nucleotide sequence of the tobacco chloroplast genome. *EMBO J.* 1986, 4, 111–148.

With further research, current evidence from studies of AAs has demonstrated that AAs can cause nephrotoxicity, carcinogenicity, and mutagenicity, especially after prolonged low-dose or shortdated high-dose intake. In this study, we determined the complete chloroplast genome sequences of *A. debilis* and *A. contorta*, which are the first two sequenced members of the family Aristolochiaceae.

Gene contents in the chloroplast genomes of *A. debilis* and *A. contorta*.

No.	Group of Genes	Gene names	Amount
1	Photosystem I	<i>psaA, psaB, psaC, psaI, psaJ</i>	5
2	Photosystem II	<i>psbA, psbB, psbC, psbD, psbE, psbF, psbH, psbI, psbJ, psbK, psbL, psbM, psbN, psbT, psbZ</i>	15
3	Cytochrome b/f complex	<i>petA, petB*, petD*, petG, petL, petN</i>	6
4	ATP synthase	<i>atpA, atpB, atpE, atpF*, atpH, atpI</i>	6
5	NADH dehydrogenase	<i>ndhA*, ndhB*(×2)¹, ndhC, ndhD, ndhE, ndhF, ndhG, ndhH, ndhI, ndhJ, ndhK</i>	12(1)
6	RubisCO large subunit	<i>rbcL</i>	1
7	RNA polymerase	<i>rpoA, rpoB, rpoC1*, rpoC2</i>	4
8	Ribosomal proteins (SSU)	<i>rps2, rps3, rps4, rps7(×2), rps8, rps11, rps12**(×2), rps14, rps15, rps16*, rps18, rps19</i>	14(2)
9	Ribosomal proteins (LSU)	<i>rpl2*(×2), rpl14, rpl16*, rpl20, rpl22, rpl23(×2), rpl32, rpl33, rpl36</i>	11(2)
10	Proteins of unknown function	<i>ycf1, ycf2(×2), ycf3**, ycf4</i>	5(1)
11	Transfer RNAs	37 tRNAs (6 contain an intron, 7 in the IRs)	37(7)
12	Ribosomal RNAs	<i>rrn4.5(×2), rrn5(×2), rrn16(×2), rrn23(×2)</i>	8(4)
13	Other genes	<i>accD, clpP**, matK, ccsA, cemA, infA</i>	6

Data revealed the presence of 18 genes containing introns in each chloroplast genome, including *atpF, rpoC1, ycf3, rps12, rpl2, rpl16, clpP, petB, petD, rps16, ndhA, ndhB*, and six tRNA genes. All the protein-coding genes were composed of 26,239 and 26,255 codons in the chloroplast genomes of *A. debilis* and *A. contorta*, respectively.

A total of 129 and 156 simple sequence repeats (SSRs) were identified using the microsatellite identification tool (MISA) in the chloroplast genomes of *A. debilis* and *A. contorta*, respectively. In these SSRs, mononucleotide repeats were largest in number, which were found 81 and 96 times in *A. debilis* and *A. contorta*, respectively. A/T mononucleotide repeats (96.3% and 94.8%, respectively) were the most common.

The comparative genomic analysis showed that the two IR regions were less divergent than the LSC and SSC regions. The four rRNA genes were the most conserved, while the most divergent coding regions were *ndhF, rpl22, ycf1, rpoC2* and *ccsA*.

The phylogenetic tree were constructed using the Maximum parsimony (MP) method based on 60 protein-coding genes commonly present in 37 species, and the result illustrated that two *Aristolochia* species were sister taxa with respect to four Piper species (Piperaceae), and these species were grouped with four species from Laurales and five species from Magnoliales.

Jiao, Y.; Jia, H.; Li, X.; et al. Development of simple sequence repeat (SSR) markers from a genome survey of Chinese bayberry (*Myrica rubra*). *BMC Genomics* 2012, 13, 201.

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

Yang, A.H.; Zhang, J.J.; Yao, X.H.; Huang, H.W. Chloroplast microsatellite markers in *Liriodendron tulipifera* (Magnoliaceae) and cross-species amplification in *L. chinense*. *Am. J. Bot.* 2011, 98, 123–126.

Acknowledgments:

This work was supported by Chinese Academy of Medical Sciences (CAMS) Innovation Fund for Medical Sciences (CIFMS) (NO. 2016-I2M-3-016), Major Scientific and Technological Special Project for “Significant New Drugs Creation” (No. 2014ZX09304307001) and The Key Projects in the National Science and Technology Pillar Program (NO. 2011BAI07B08).