

The complete chloroplast genome sequence of the medicinal and economic plant woad *Isatis indigotica* (Brassicaceae)

Shu Yang and Zhezhi Wang

Key Laboratory of the Ministry of Education for Medicinal Resources and Natural Pharmaceutical Chemistry, National Engineering Laboratory for Resource Development of Endangered Crude Drugs in Northwest China, College of Life Sciences, Shaanxi Normal University, Xi'an, Shaanxi, P.R. China

ABSTRACT

Chloroplast (cp) genome sequences provide a valuable source for phylogenetic analysis. It becomes a popular useful tool for population and phylogeny in recently report. Here, the complete chloroplast genome of the *Isatis indigotica* has been reconstructed from the whole-genome Illumina sequencing data. The circular genome is 156,670 bp in size and comprises a pair of inverted repeat (IR) regions of 26,995 bp each, a large single-copy (LSC) region of 84,907 bp and a small single-copy (SSC) region of 17,773 bp. The total GC content is 36.5%, while the corresponding values of the LSC, SSC and IR region are 34.2%, 29.7% and 42.3%, respectively. The chloroplast genome contains 140 genes, including 94 protein-coding genes. The phylogenetic analysis showed a strong sister relationship with *Raphanus sativus* in Brassicaceae. These findings provide a foundation for further investigation of cp genome evolution in *I. indigotica* and other higher plants.

ARTICLE HISTORY

Received 3 July 2017
Accepted 26 July 2017

KEYWORDS

Isatis indigotica; Illumina sequencing; chloroplast genome; MITObim

The chloroplast is a plant-specific organelle that carries out photosynthesis and other crucial biosynthesis, which becoming valuable phylogenetic tool for low recombination rate (Saski et al. 2007; Dong et al. 2012). It is hot spot that researches on sequencing chloroplast of valued medicine plants such as *Salvia miltiorrhiza*, *Magnolia officinalis* and *Ginkgo biloba* (Li et al. 2012; Lin et al. 2012; Qian et al. 2013).

Isatis indigotica Fort. (Brassicaceae) is native to the steppe and desert zones of the Caucasus as a source of blue dye. In China, *I. indigotica* is widely distributed and used as a popular herbal medicine for treatment of colds, fever and influenza (Xu et al. 2010; Shan et al. 2014). The gene resource of *I. indigotica* is abundant (Tang et al. 2014; Zhang et al. 2016); however, few data are available regarding the *I. indigotica* chloroplast genome. Here, we collected the samples of *I. indigotica* from Qinling Mountains (Shaanxi Province) and stored in our lab and then reconstructed the complete chloroplast genome sequence from the whole-genome Illumina sequencing data. Our data are valuable for future genetic and phylogenetic studies on *I. indigotica*.

A subset of 17.4 M trimmed reads were used for reconstructing the chloroplast genome with MITObim v1.7 (Hahn et al. 2013), with that of its congener *Arabidopsis thaliana* (GenBank: NC_000932) as the initial reference sequence. A total of 1,524,092 individual chloroplast reads were

achieved with an average coverage of 1223.6-fold. The chloroplast genome was annotated in GENEIOUS R8 (Biomatters Ltd., Auckland, New Zealand) by aligning with that of *A. thaliana* (NC_000932).

The whole chloroplast genome of *I. indigotica* is a double-stranded circular DNA molecule with a length of 156,670 bp with the accession number KT939360. It shows a typical quadripartite structure which comprises a pair of inverted repeat (IR) regions of 26,995 bp each, separated by a large single-copy (LSC) region of 84,907 bp and a small single-copy (SSC) region of 17,773 bp. The total GC content is 36.5%, while the corresponding values of the LSC, SSC and IR region are 34.2%, 29.7% and 42.3%, respectively.

This chloroplast genome harbours 137 functional genes, including 90 protein-coding genes (PCGs), 37 tRNA genes and 8 rRNA genes. Eight PCGs, seven tRNA genes and all rRNA genes, most of which related to photosynthesis, are duplicated in the IR regions. The LSC region possesses 61 PCGs and 22 tRNA genes, while the SSC region contains 13 PCGs and one tRNA gene. 32 PCGs, 19 tRNA genes and 4 rRNA genes are located in the forward strand while others are located in the reverse strand.

We performed a phylogenetic tree using 16 cp genome sequences in Brassicaceae family. The neighbour-joining tree showed the position of *I. indigotica* was situated as the sister of *Raphanus sativus* which was in good agreement with

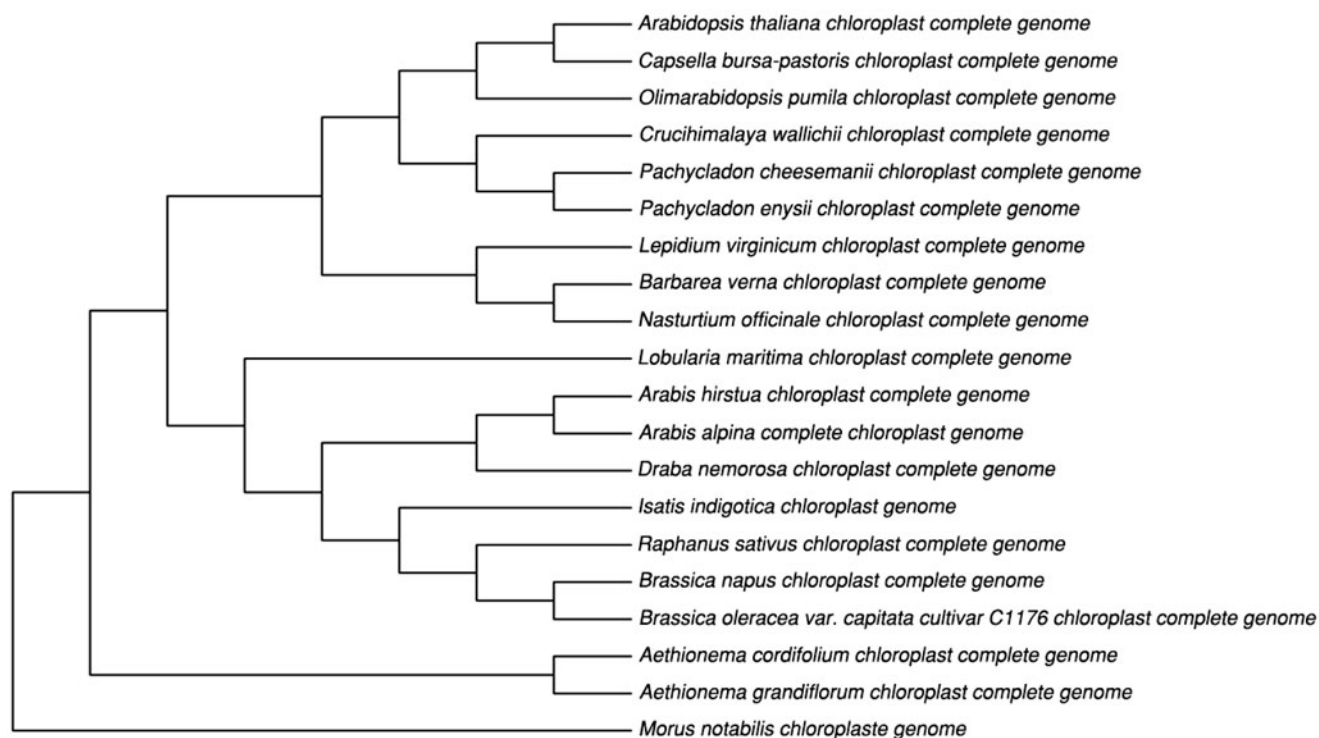


Figure 1. Phylogenetic of 16 species within the family Brassicaceae based on the neighbour-joining analysis of the whole cp genome sequences using 500 bootstrap replicates and setting *Morus notabilis* (Moraceae) as out-group.

taxonomical classification (Figure 1). This study provides the reference to phylogenetic and evolutionary study.

Funding

The authors were financially supported by the Fundamental Research Funds for National Natural Science Foundation of China No. 31670299.

Disclosure statement

The authors report no conflicts of interest and alone are responsible for the content and writing of the paper.

References

- Dong W, Liu J, Yu J, Wang L, Zhou S. 2012. Highly variable chloroplast markers for evaluating plant phylogeny at low taxonomic levels and for DNA barcoding. *PLoS One*. 7:e35071.
- Hahn C, Bachmann L, Chevreur B. 2013. Reconstructing mitochondrial genomes directly from genomic next-generation sequencing reads – a baiting and iterative mapping approach. *Nucleic Acids Res*. 41:e129
- Li X, Hu Z, Lin X, Li Q, Gao H, Luo G, Chen S. 2012. [High-throughput pyrosequencing of the complete chloroplast genome of *Magnolia officinalis* and its application in species identification]. *Yao Xue Xue Bao = Acta Pharmaceutica Sinica*. 47:124–130.
- Lin CP, Wu CS, Huang YY, Chaw SM. 2012. The complete chloroplast genome of *Ginkgo biloba* reveals the mechanism of inverted repeat contraction. *Genome Biol Evol*. 4:374–381.
- Qian J, Song J, Gao H, Zhu Y, Xu J, Pang X, Yao H, Sun C, Li X, Li C, et al. 2013. The complete chloroplast genome sequence of the medicinal plant *Salvia miltiorrhiza*. *PLoS One*. 8:e57607.
- Saski C, Lee SB, Fjellheim S, Guda C, Jansen RK, Luo H, Tomkins J, Rognli OA, Daniell H, Clarke JL. 2007. Complete chloroplast genome sequences of *Hordeum vulgare*, *Sorghum bicolor* and *Agrostis stolonifera*, and comparative analyses with other grass genomes. *Theor Appl Genet*. 115:571–590.
- Shan J, Sun G, Ren J, Zhu T, Jia P, Qu W, Li Q, Wu J, Ma H, Wen S. 2014. An α -glucan isolated from root of *Isatis indigotica*, its structure and adjuvant activity. *Glycoconj J*. 31:317–326.
- Tang X, Xiao Y, Lv T, Wang F, Zhu Q, Zheng T, Yang J. 2014. High-throughput sequencing and *de novo* assembly of the *Isatis indigotica* transcriptome. *PLoS One*. 9:e102963.
- Xu Y, Sun J, He S. 2010. Effect of three kinds of *Radix isatidis* preparation on the expression of nucleoprotein of influenza virus. *Shandong Med J*. 50:8–10.
- Zhang L, Chen J, Li Q, Chen W. 2016. Transcriptome-wide analysis of basic helix-loop-helix transcription factors in *Isatis indigotica* and their methyl jasmonate responsive expression profiling. *Gene*. 576:150–159.