

## Complete sequencing of the chloroplast genomes of two *Medicago* species

Yan Weihong, Shi Wengui, Liu Lei, Ma Yubao, Chen Libo, Wang Zhaolan and Hou Xiangyang

Ministry of Agriculture, Institute of Grassland Research of CAAS/Key Laboratory of Grassland Resources and Utilization, Inner Mongolia, Hohhot, China

### ABSTRACT

In this study, we aimed to determine the complete nucleotide sequences of the chloroplast genomes of *Medicago sativa* L. and *Medicago falcata* L. and perform their phylogenomic analysis. The chloroplast genomes of *M. Sativa* and *M. falcata* were found to be 125,095 and 125,810 bp in length, respectively. Additionally, the AT contents of *M. Sativa* and *M. falcata* are 66.1% and 66.2%, respectively. Each chloroplast genome contained the same 109 unique genes, including 29 transfer RNA genes, 4 ribosomal RNA genes, and 76 protein-coding genes. Phylogenomic analysis based on complete chloroplast genomes from the *Medicago* showed that the geographical environment has no significant effect on their genomic composition.

### ARTICLE HISTORY

Received 24 February 2017  
Accepted 27 April 2017

### KEYWORDS

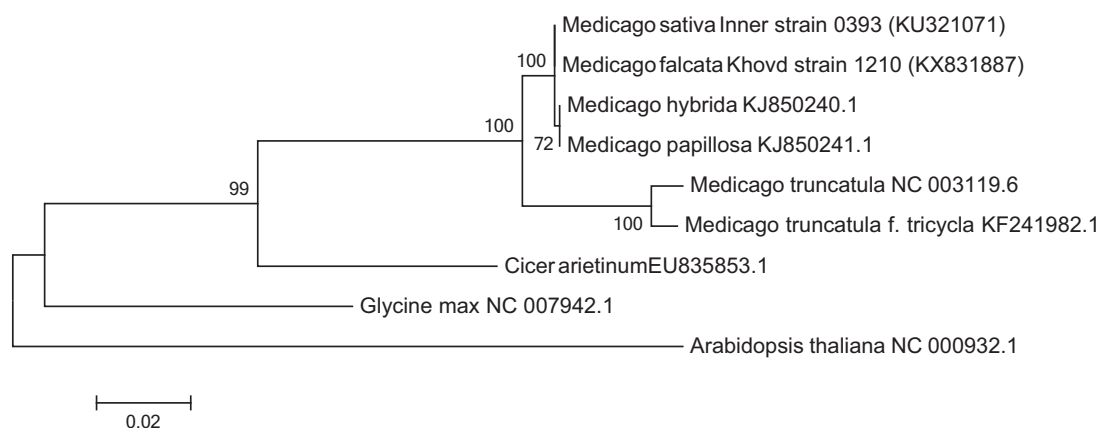
*Medicago*; chloroplast genome; phylogenomic analysis

*Medicago* plants are one of the most widely cultivated perennial forage legumes worldwide. The plant materials originating in different geographical, ecological, and climatic regions were used in this study. *Medicago sativa* seeds were obtained from Inner Mongolia, China (42°02'N, 111°35'E), and *M. falcata* seeds from Khovd, Mongolia (46°49'N, 91°45'E). The seeds were not deposited into the National Library and were cultivated and spread planting at Taipusiqi test sites, the Institute of Grassland, Chinese Academy of Agricultural Sciences (Inner Mongolia, China). After allowing the plants to grow for 12 weeks in a greenhouse at 22 °C with a photoperiod of 16 h, their leaves were collected for further experiments.



Total genomic DNA was extracted from 150 mg of actively growing fresh leaves using a modified CTAB method (Doyle 1987). Nine novel universal primer pairs (Yang et al. 2014)

were used to amplify *Medicago* chloroplast genomes. Long-range PCR method was performed to amplify target fragments, and purified DNA was used to construct short-insert (350–550 bp) libraries. Solexa high-throughput sequencing system (Illumina HiSeq2500, San Diego, CA) was used to generate raw sequence reads. DNA from each individual was indexed using tags and pooled together for sequencing at Beijing Yuanquanyike Biological Technology Co., Ltd (Beijing, China).

*Medicago sativa* and *M. falcata* were sequenced individually to produce 8,325,635 and 7,689,970 clean paired-end reads, respectively, with 96 bp in average length. The amplified fragments were sequenced by Sanger-sequencing using primers designed on the basis of the reference genome. First, various quality control checks on the short reads were



**Figure 1.** ML phylogenetic tree is rooted with *Arabidopsis thaliana*. The most basal lineage was *Glycine max* (L.) Merr. followed by the *Cicer arietinum* L. The next branch included *Medicago*, of which *M. sativa* and *M. falcata* showed higher similarity.

**CONTACT** Hou Xiangyang  Houxy16@126.com  Ministry of Agriculture, Institute of Grassland Research of CAAS/Key Laboratory of Grassland Resources and Utilization, Inner Mongolia, Hohhot, 010010, China

© 2017 The Author(s). Published by Informa UK Limited, trading as Taylor & Francis Group.

This is an Open Access article distributed under the terms of the Creative Commons Attribution License (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

performed using solexa QA to filter the Illumina data for high-quality and vector- and adaptor-free reads (Cox et al. 2010). High-quality short reads were assembled into contigs using SOAP denovo (Luo et al. 2012). Subsequently, we identified highly similar genome sequences by Blast with the default search parameters. The GapCloser software (University of Hong Kong, Pokfulam, Hong Kong) was used to fill the gaps with high-quality short sequences, and then the gaps were verified by Sanger sequencing. Finally, to avoid assembly errors, some pairs of primers were designed for PCR amplifications. The final complete chloroplast genome sequences were deposited into GenBank with accession numbers of KU321071 (*M. sativa*) and KX831887 (*M. falcata*).

In order to examine the phylogenetic utility of different regions, phylogenetic analyses were performed based on two data partitions (complete cpDNA sequences and 72 protein-coding exons) from five Leguminosae plants' cp genomes to construct a Maximum-Likelihood phylogenetic tree. Phylogenetic relationships with bootstrap values and posterior probabilities are presented in Figure 1. Phylogeny among the two newly sequenced *Medicago* chloroplasts and other chloroplasts was constructed on the basis of the genome sequences and minor allele frequencies (MAFs) in the polymorphism sites. As clear differences were observed between chloroplast and the newly sequenced chloroplasts, *thaliana* chloroplast was used as an outgroup directly.

We expect the two complete *Medicago* chloroplast genome sequences and the genomic variations might provide some valid information for further studies on the genetic relationship of this genus.

### Disclosure statement

None of the co-authors has any conflict of interest to declare. The authors alone are responsible for the content and writing of the paper.

### Funding

This work was supported by the National Natural Science Foundation of China under Grant [No.31302017].

### References

- Doyle JJ. 1987. A rapid DNA isolation procedure for small quantities of fresh leaf tissue. *Phytochem Bull.* 19:11–15.
- Cox MP, Peterson DA, Biggs PJ. 2010. SolexaQA: at-a-glance quality assessment of Illumina second-generation sequencing data. *BMC Bioinformatics.* 11:485.
- Luo R, Liu B, Xie Y, Li Z, Huang W, Yuan J, He G, Chen Y, Pan Q, Liu Y, et al. 2012. SOAP denovo2: an empirically improved memory-efficient short-read de novo assembler. *Gigascience.* 1:18.
- Yang JB, Li DZ, Li HT. 2014. Highly effective sequencing whole chloroplast genomes of angiosperms by nine novel universal primer pairs. *Mol Ecol Resour.* 14:1024–1031.