

## MITOGENOME ANNOUNCEMENT

**The complete chloroplast genome sequence of Safflower  
(*Carthamus tinctorius* L.)**Chaolong Lu<sup>1,2</sup>, Qi Shen<sup>1,3</sup>, Jun Yang<sup>1</sup>, Bo Wang<sup>1,4</sup>, and Chi Song<sup>1</sup><sup>1</sup>Institute of Chinese Materia Medica, China Academy of Chinese Medical Sciences, Beijing, P.R. China, <sup>2</sup>Maize Research Institute, Sichuan Agricultural University, Wenjiang, China, <sup>3</sup>Guizhou Rapeseed Institute, Guizhou Province of Academy of Agricultural Sciences, Guiyang, China, and <sup>4</sup>Institute of Medicinal Plant Development, Chinese Academy of Medical Science & Peking Union Medical College, Beijing, P.R. China**Abstract**

Safflower (*Carthamus tinctorius* L.) is a traditional medical plants of Asia. In this study, the complete chloroplast genome of safflower was presented. The total genome size was 153,675 bp in length, containing a pair of inverted repeats (IRs) of 25,407 bp, separated by large single copy (LSC) and small single copy (SSC) of 83,606 bp and 19,156 bp, respectively. Overall GC content of the genome was 37.4%. The chloroplast genome harbored 127 annotated genes, including 89 protein coding genes, 30 tRNA genes and 8 rRNA genes. A total of 7 of these genes were duplicated in the inverted repeat regions. Twelve genes contained one intron.

**Keywords**

Chloroplast, genome sequence, medicinal plant, safflower

**History**

Received 14 January 2015

Accepted 24 January 2015

Published online 5 March 2015

The Asteraceae is a larger family in angiosperms, comprising 24,000 to 30,000 species from 1600 to 2000 genera (Garnatj et al., 2011). Safflower (*Carthamus tinctorius* L.) is one of valuable medicinal crops of Asteraceae, which is extensively cultivated in west Asia and China. Safflower could yield a variety of products, including the numerous secondary metabolites of its flowers and the unsaturated fatty acids in its seeds, many of which are very beneficial to human health (Li et al., 2012; Lulin et al., 2012).

Chloroplast was hypothesized to have evolved from ancient endosymbiotic cyanobacteria. Its genomes sequences recently had been viewed as valuable information for taxonomic classification and phylogenetic reconstruction of angiosperms (Huang et al., 2014). At present, complete chloroplast genomes of *Helianthus annuus*, *Artemisia frigida* and some of influential Asteraceae family plants had been reported (Liu et al., 2013; Nie et al., 2012; Timme et al., 2007). In the article, a complete chloroplast genome of safflower was sequenced and presented (Figure 1).

Total genome DNA of safflower was extracted to be used for constructing 180 bp small-size fractionated libraries according to NEB Next Ultra DNA library Prep Kit for Illumina. Whole genome sequences were executed using Illumina HiSeq 1500 Genome Analyzer (rapid run model configuration for pair-end 101 reads) at institute of Chinese materia medica in China. Raw reads were filtered by standard method (Patel & Jain, 2012), then 3,827,770 filtered reads were assembled into the complete

chloroplast genome by SOAPdenovo software (Luo et al., 2012) with *Artemisia frigida* chloroplast genome sequence as reference genome (Qian et al., 2013). The annotation of chloroplast genome was conducted accorded to methods described by Liu et al. (2012). In brief, the most homologous protein, cDNA and rRNA from GenBank were allocated on safflower chloroplast genome with the combination of Blastx, Blastn, protein2genome and est2genome programs (Liu et al., 2012). The inverted repeats were identified using the vmatch software tool (Abouelhoda et al., 2004). And tRNAs were identified using tRNAscan (Lowe & Eddy, 1997). The complete chloroplast genome sequence together with gene annotations were submitted to the GenBank under the accession number of KP404628.

Similarly to typical chloroplast genome structure in higher plants, it has a large single copy (LSC) region of 83,606 bp and a small single copy (SSC) region of 19,156 bp, was separated by a pair of inverted repeats (IRa and IRb) of 25,407 bp (Figure 1). The GC content of the whole chloroplast genome was 37.4% with the IRs region has a higher GC content (42.7%) than LSC (35.7%) and SSC (30.7%) region due to the presence of GC-rich rRNA genes. The gene coding regions account for 62.2% of the genome length including 127 genes, composed of 89 protein-coding genes, 30 tRNA genes and 8 rRNA genes. Among these genes, 12 genes contain one intron. The chloroplast genome contains 30 different tRNA genes and 10 of them were present in the IRs region (Figure 1).

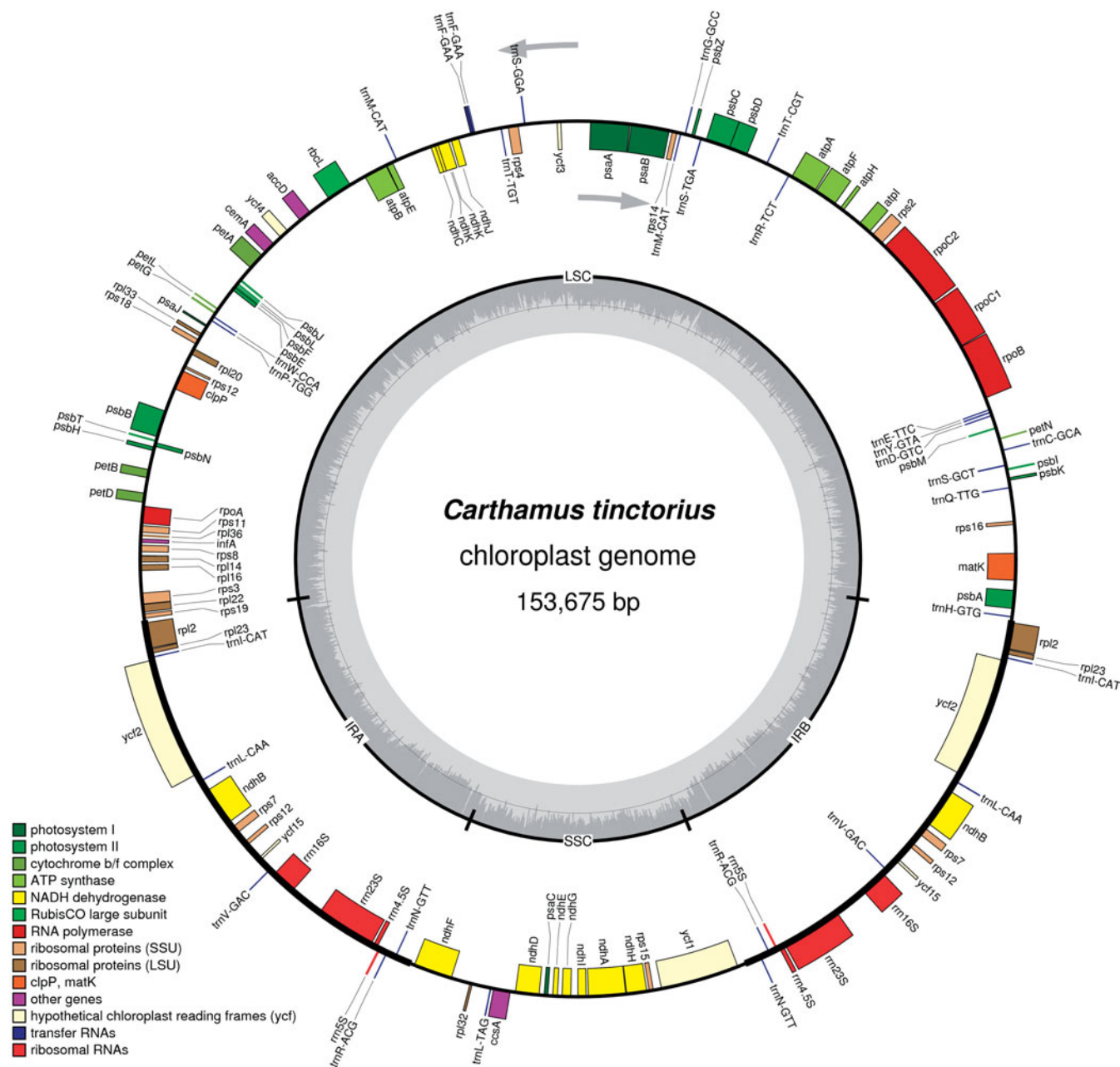


Figure 1. Gene map of *C. tinctorius* chloroplast genome. Genes drawn inside the circle are transcribed clockwise, while those drawn outside the circle are counter clockwise. Genes belonging to different functional groups are color-coded. Darker grey in the inner circle corresponds to the GC content.

## Declaration of interest

This work was supported by China Postdoctoral Science Foundation (2014M550135) and National Natural Science Foundation of China (31360067).

## References

- Abouelhoda MI, Kurtz S, Ohlebusch E. (2004). Replacing suffix trees with enhanced suffix arrays. *J Dis Algor* 2:53–86.
- Garnatje T, Canela MÁ, Garcia S, Hidalgo O, Pellicer J, Sánchez-Jiménez I, Siljak-Yakovlev S, et al. (2011). GSAD: A genome size in the Asteraceae database. *Cytom Part A* 79:401–4.
- Huang H, Shi C, Liu Y, Mao SY, Gao LZ. (2014). Thirteen *Camellia* chloroplast genome sequences determined by

high-throughput sequencing: Genome structure and phylogenetic relationships. *BMC Evol Biol* 14:151.

- Li H, Dong Y, Yang J, Liu X, Wang Y, Yao N, Guan L, et al. (2012). De novo transcriptome of safflower and the identification of putative genes for oleosin and the biosynthesis of flavonoids. *PLoS One* 7:e30987.

- Liu C, Shi L, Zhu Y, Chen H, Zhang J, Lin X, Guan X. (2012). CpGAVAS, an integrated web server for the annotation, visualization, analysis, and GenBank submission of completely sequenced chloroplast genome sequences. *BMC Genomics* 13:715.

- Liu Y, Huo N, Dong L, Wang Y, Zhang S, Young HA, Feng X, Gu YQ. (2013). Complete chloroplast genome sequences of Mongolia medicine *Artemisia frigida* and phylogenetic relationships with other plants. *PLoS One* 8:e57533.

- Lowé TM, Eddy SR. (1997). tRNAscan-SE: A program for improved detection of transfer RNA genes in genomic sequence. *Nucleic Acids Res* 25:0955–64.

- Lulin H, Xiao Y, Pei S, Wen T, Shangqin H. (2012). The first Illumina-based *de novo* transcriptome sequencing and analysis of safflower flowers. *PLoS One* 7:e388653.
- Luo RB, Liu BH, Xie YL, Li ZY, Huang WH, Yuan JY, He GH, et al. (2012). SOAPdenovo2: An empirically improved memory-efficient short-read *de novo* assembler. *GigaScience* 1:18.
- Nie X, Lv S, Zhang Y, Du X, Wang L, Biradar SS, Tan X, et al. (2012). Complete chloroplast genome sequence of a major invasive species, crofton weed (*Ageratina adenophora*). *PLoS One* 7:e36869.
- Patel RK, Jain M. (2012). NGS QC Toolkit: A toolkit for quality control of next generation sequencing data. *PLoS One* 7:e30619.
- Qian J, Song J, Gao H, Zhu Y, Xu J, Pang X, Yao H, et al. (2013). The complete chloroplast genome sequence of the medicinal plant *Salvia miltiorrhiza*. *PLoS One* 8:e57607.
- Timme RE, Kuehl JV, Boore JL, Jansen RK. (2007). A comparative analysis of the *Lactuca* and *Helianthus* (Asteraceae) plastid genomes: Identification of divergent regions and categorization of shared repeats. *Am J Bot* 94:302–12.