Physiol Mol Biol Plants (January–March 2017) 23(1):35–41
DOI 10.1007/s12298-016-0395-6

RESEARCH ARTICLE

# The complete chloroplast genome sequence and phylogenetic analysis of *Chuanminshen* (*Chuanminshenviolaceum* Sheh et Shan)

**Can Yuan[1] · Wenjuan Zhong[1] · Fangsheng Mou[1] · Yiyun Gong[1] · Deqiang Pu[1] · Peicheng Ji[1] · Haiyan Huang[1] · Zehu Yang[1] · Chao Zhang[1]**

Received: 10 July 2016 / Revised: 17 October 2016 / Accepted: 18 November 2016 / Published online: 3 December 2016


**Abstract** Chloroplast genome sequences are very useful for species identification and phylogenetics. *Chuanminshen* (*Chuanminshen violaceum* Sheh et Shan) is an important traditional Chinese medicinal plant, for which the phylogenetic position is still controversial. In this study, the complete chloroplast genome of *Chuanminshen violaceum* Sheh et Shan was determined. The total size of *Chuanminshen* chloroplast genome was 154,529 bp with 37.8% GC content. It has the typical quadripartite structure, a large single copy (17,800 bp) and a small single copy (84,171 bp) and a pair of inverted repeats (26,279 bp). The whole genome harbors 132 genes, which includes 85 protein coding genes, 37 tRNA genes, eight rRNA genes, and two pseudogenes. Thirty-nine SSR loci, 32 tandem repeats and 49 dispersed repeats were found. Phylogenetic analyses results with the help of MEGA showed a new insight for the *Chuanminshen* phylogenetic relationship with the reported chloroplast genomes in *Apiales* plants.

**Keywords** *Chuanminshen violaceum* Sheh et Shan · Chloroplast genome · Genome features · Phylogenetic analysis

## Introduction

Chloroplast plays an important role in the plant photosynthesis and carbon fixation (Neuhaus and Emes 2000). The size of the major angiosperm chloroplast genome is 110–165 kb and contains 90–110 unigenes (Sugiura 1992). They consist of four parts, a large single copy region (LSC), a small single copy region (SSC), and two inverted repeats (IRs; Jansen et al. 2005). Chloroplast genomes are highly conserved in sequence and structure due to their non-recombinant, haploid and uniparentally (i.e., maternally) inherited nature (Birky 2001; Wicke et al. 2011). Therefore, chloroplast genome sequence was widely used in the phylogenetic analyses, organelle-scale barcodes research and evolutionary studies. Presently, chloroplast sequences of 30 species of *Apiales* plants were reported in NCBI (http://www.ncbi.nlm.nih.gov/genome/organelle/).

*Chuanminshen* is a Chinese endemic genera of the *Apiaceae* plant family which has only one species. *Chuanminshen violaceum* Sheh et Shan is a typical species of the *Apiaceae* family (Flora 1979), and is considered a 'medicine food homology' plant. The rhizome of the plant is popularly used as a traditional Chinese medicine in China. *Chuanminshen* has significant effects in reducing phlegm, anti-cough and nourishing yin. Modern pharmacognosy research showed that the roots of *Chuanminshen* are rich in bioactive components such as *Chuanminshen violaceum* polysaccharides (CVPs; Lei and Zhang 2012). These components function as antioxidants, antimutations, antifatigue (Chen and Peng 2011), and antivirals (Song et al. 2013), and have expectorant and antitussive properties as well, along with reports of physique enhancement (Feng et al. 2015; Zhang et al. 2007). The latest study also found that polysaccharides obtained from *Chuanminshen* could improve the immune responses of foot-and-mouth disease vaccine in mice (Feng et al. 2015).

Previously, research of *Chuanminshen* mostly focused on the field of pharmacognosy, but very rarely on the phylogenetics. To date, the phylogenetic analysis of *Chuanminshen* (She and Shan 1980; Song et al. 2014; Tao

Can Yuan, Wenjuan Zhong and Fangsheng Mou have contributed equally to this work.

✉ Chao Zhang
jychaozhang@163.com

[1] Industrial Crop Research Institute, Sichuan Academy of Agricultural Sciences, Chengdu 610300, China

et al. 2008) is still controversial. In the current study, we sequenced the *Chuanminshen* chloroplast genome, and made a survey of its general features. Furthermore, we investigated the phylogenetic relationships by *Chuanminshen* and all other *Apiales* plants whose chloroplast genome have been reported.

## Materials and methods

### Plant materials

The sequenced plant was originated from our own breeding parents, the inbred line CMS-1. The plant was confirmed by Shangqin Hu (Director of Chinese herbal medicine planting research center of Sichuan province, China) through plant phenotype indentification.

### Methods

Total genomic DNA was extracted from fresh, clean *Chuanminshen* leaves by plant genomic DNA extraction kit ™. cpDNA amplification was carried out through nine universal effective primer pairs by long-range PCR (Yang et al. 2014). The PCR products were fragmented to construct 500 bp short-insert libraries according to the Illumina manufacturer's manual. Each DNA library was labeled with a barcode and pooled together in one lane and the sequencing was executed using Illumina Hiseq 2000 in Kunming Institute of Botany, Chinese Academy of Sciences. Raw data were filtered using the Next Generation Sequencing (NGS) QC toolkit (Patel and Jain 2012), high-quality short reads were assembled into the complete chloroplast genome using SOAPdenovo software (Luo et al. 2015) andthe chloroplast genome was annotated using the Dual Organellar GenoMe Annotator (DOGMA) tool (Wyman et al. 2004) and CpGAVAS (Liu et al. 2012). The start/stop codons and intron/exon boundaries were manually corrected. The perl script MicroSAtellite (MISA; Parida et al. 2010) was used to analyze the distributing frequency of chloroplast genome simple sequence repeat (SSR; parameter setting: the minimum repeat number of each unit are as follows, 1–10, 2–6, 3–5, 4–5, 5–5, 6–5). Tandem repeat finder (Benson 1999; the default parameters were used) and REPuter (Kurtz and Schleiermacher 1999; the setting parameters of REPuter to the minimal repeat size of 30 bp, hamming distance to 3) was used to analyze the repeat incidents in *Chuanminshen* cp DNA. In order to avoid the influences of the IR regions, we used only a single IR region, meanwhile, the redundant results of REPuter were manually removed. *Chuanmingshen* cp genome, 30 *Apiales* chloroplast genome and an outgroup chloroplast genome was downloaded from NCBI, 80 common proteins were used to carry out the maximum likelihood (ML) analysis based on the JTT matrix-based model and 1000 bootstraps by MEGA7 (Kumar et al. 2016), all positions containing gaps and missing data were eliminated. There were a total of 12,409 positions in the final dataset, the branch lengths measured in the number of substitutions per site.

## Results and discussion

### Cp genome general features of *Chuanminshen*

The full length of chloroplast genome sequence for *Chuanminshen* (with the accession number in GenBank is KU921430) was 154,529 bp, constructed with a quadripartite structure (Fig. 1). The four parts were LSC with 17,800 bp, SSC with 84,172 bp, IR regions (IRa and IRb) with 26,279 bp. The whole *Chuanminshen* cp genome contained 132 genes, including 85 protein-coding genes, 37 transfer RNAs, eight ribosomal RNAs and two pseudogenes (*ycf1* and *rps19*). Whereas only 112 unigenes were harbored in *Chuanminshen* cp genome due to a multi-copy of six protein-coding genes, four rRNAs and 10 tRNAs. The total GC content in *Chuanminshen* cp genome was 37.8%. Of these, the LSC region, SSC region and IR region was 35.9, 31.5 and 42.9%, respectively. Most genes in *Chuanminshen* cp genome contained only one or no intron. In addition, three genes had two introns, which are *rps12* (ribosomalproteinS12), *clpP* (clp protease proteolytic subunit), *ycf3* (hypothetical chloroplast RF34).

Compared with other cp genome sequences, *Chuanminshen* cp genome was smaller than *Panax ginseng* 156,318 bp (Kim and Lee 2004), *Nicotiana tabacum* 155,943 bp (Shinozaki et al. 1986), and much longer than *Salvia miltiorrhiza* Bge151328 bp (Qian et al. 2013), a model plant of Chinese herbal medicine (Zhang et al. 2015). Statistical analysis showed that *Chuanminshen* is the sixth largest cp genome in size and has the highest GC content among the 15 chloroplast genome sequences (including *Chuanminshen*) of *Apiaceae* plant family. Similar to the *Sesamum indicum* L (Yi and Kim 2012), *Epimedium* (Zhang et al. 2016), *Daucus carota* (Ruhlman et al. 2006), it is rich in GC content in IR regions owing to four rRNA genes. Through the annotation, all the genes were classified into four categories, including the gene for self replication such as rRNA genes and tRNA genes, genes for photosynthesis such as *psaA* and *ndhA*, other genes such as *clpP*, and genes of unknown functions such as *ycf1*.

### Cp SSR and repeat sequence analysis

SSR markers are widely used in phylogenetic analysis, population genetics and ecological studies (Cavalier-Smith 2002).

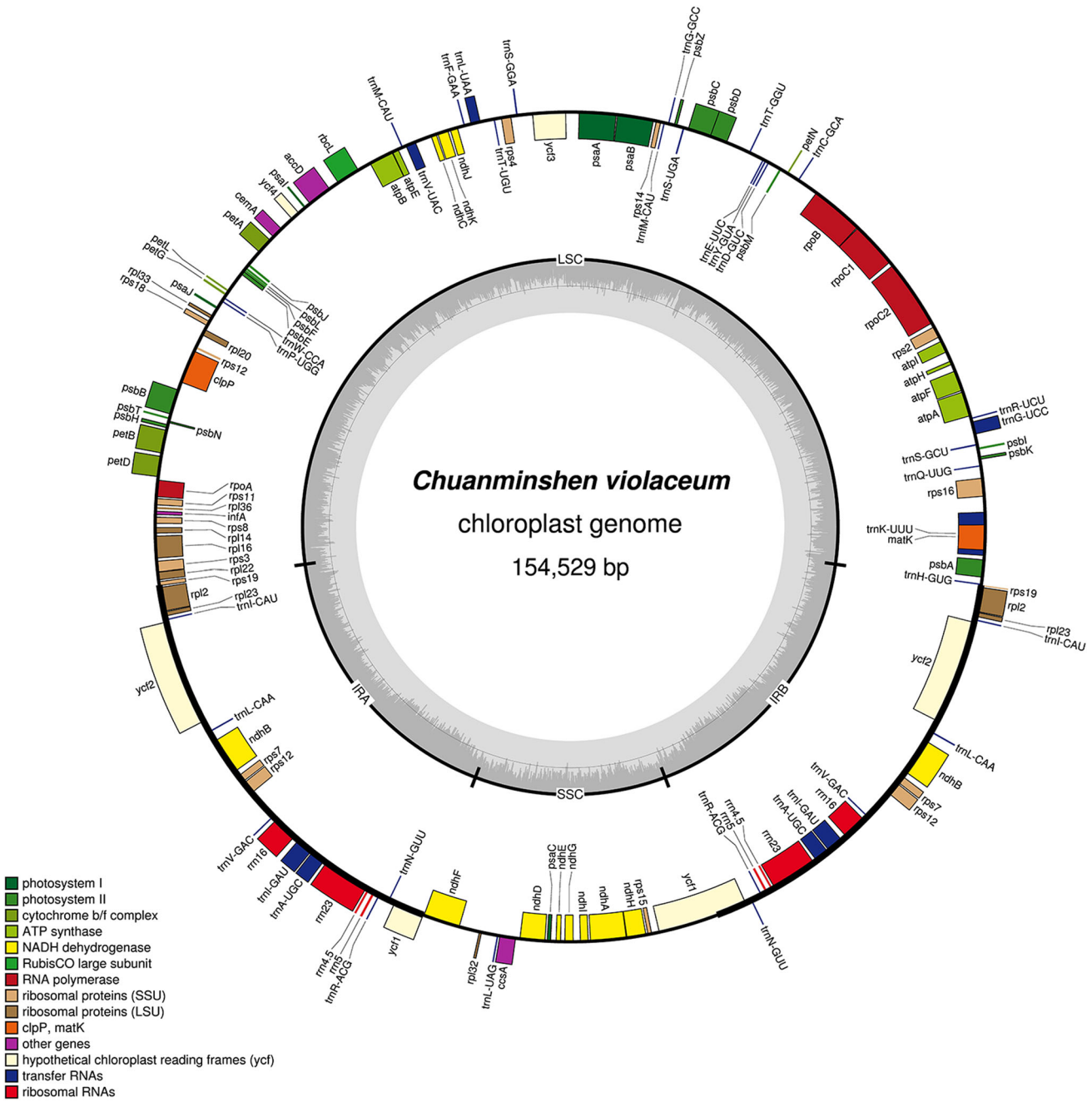


**Fig. 1** Gene map of the *Chuanminshen violaceum* Sheh et Shan chloroplast genome: *Outside of the outer circle* was drawn with the counter clockwise genes, and *inside* was drawn with transcribed clockwise genes, *various colors* indicated the genes in different functional groups. *LSC* The large single copy region, *SSC* the small single copy region, *IRa* and *IRb* the inverted repeats. The *two innermost gray area*: The *darker gray* represented GC content and the *lighter* represented AT content

Using MISA, 39 SSR loci were identified in the *Chuanmingshen* cp genome. Three compound formations of SSR were presented, 10.3% di-nucleotide (all was AT/AT) and 89.7% mono-nucleotide (84.6% A/T, 5.1% G/C; Fig. 2). Eleven repeats of A/T motif is the most frequent and included up to 12. The second abundant is 10 repeats of A/T motif, which was found in eight loci. Once in eight repeats and thrice in six repeats were found in AT/AT motif. Obtaining SSR from the cp genome was common in the study, because it exhibited highly polymorphic results due to the diversity levels in repeat unit copy numbers among the same species (Grassi et al. 2002; Powell et al. 1995). In accordance with most cp SSR research results, *Chuanminshen* cp SSR was rich in homopolymers as well. Surprisingly, we did not find any tri-nucleotide repeats or larger repeat units in the *Chuanmingshen* cp genome, we

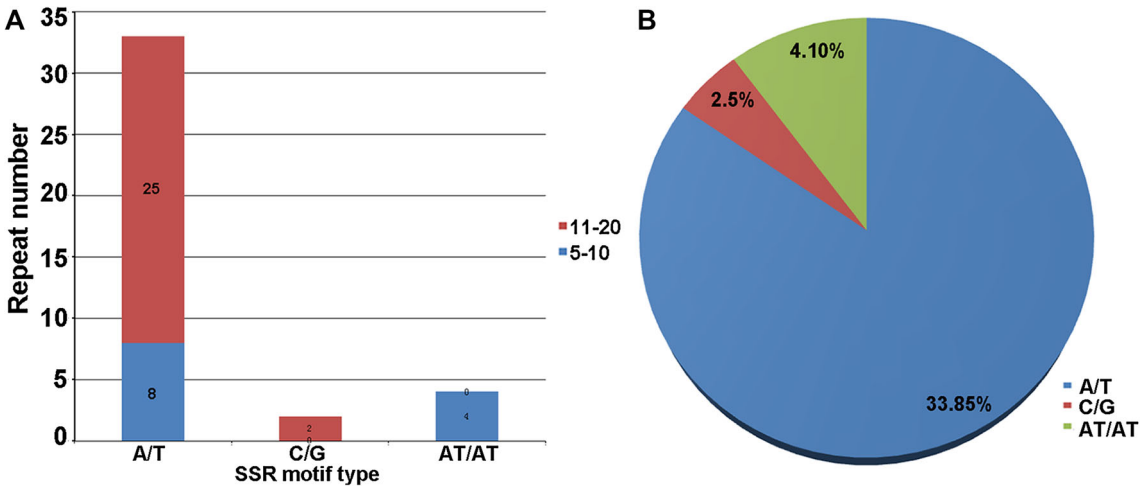


**Fig. 2** The distribution information of SSR motif repeat in *Chuanminshen* cp genome. **a** The number of distributions of SSR motif repeat. x-axis SSR repeat types; y-axis Number of SSR types; *red bars* stand for 11–20 motif repeats, *blue bars* stand for 5–10 motif repeats. **b** The percentage distribution of SSR motif type. *Blue sector* represented A/T, *green sector* represented AT/AT, *red sector* represented G/C

suspected that this result was relevant due to the usage of the default MISA tool. Subsequently, we analyzed the *Daucus carota*, *Salvia miltiorrhiza* and *Epimedium* cp genome based on our own set of parameters. In contrast to the previous SSR analysis reported in the literature, the statistical results confirmed our speculation that only two types of the repeat units were observed.

In most angiosperm plants, repeat regions regularly occurred in non-coding regions and frequent variation occurred due to illegitimate recombination and slipped-strand mispairing (Asano et al. 2004; Timme et al. 2007). The tandem repeats and dispersed repeat sequence analysis were realized by Tandem repeat finder and REPuter respectively in this study. Thirty-two tandem repeats were found, ranging from 9 to 40 bp. Twenty-one bp repeats were most common within five times, which was thrice in LSC and twice in IRs. The major distribution of repeat size was 11–21 bp (Fig. 3a). The number of tandem repeats in *Chuanminshen* cp genome was equivalent to that of crofton weed (Nie et al. 2012) and slightly more than bamboo (Zhang et al. 2011) whereas the longest tandem repeat size was shorter than the crofton weed with 85 bp and bamboo with 65 bp. To investigate the dispersed repeats in *Chuanminshen* cp genome, we adjusted the setting parameters of the REPuter described above. In total, 19 repeats were found and were divided into three categories, eight forward repeats, nine palindromic repeats and two reverse repeats. The maximum size of repeat was 64 bp. Repeat unit length in 30–40 bp was 16 including two reverse repeats, 41–50 bp was two, and the rest was one with 64 bp (Fig. 3).

The results present here in regarding to the SSR and the repeat sequences promoted the identification of *Chuanminshen* species which laid the foundation for further phylogenetic studies and diversity analyses.

## Phylogenetic analysis

Chloroplast genome was successfully applied to phylogenetics in several angiosperms (Samson et al. 2007). Previous studies had two viewpoints in the phylogenetic analysis of *Chuanminshen*. In the first report, the phylogenetic position of *Chuanminshen* was placed in the *Peucedaneae* Drude based on the microstructural features of the fruit (She and Shan 1980). Beyond the traditional classification, by using the molecular markers ITS, RAPD and ISSR, *Chuanminshen* was clustered closely to *Changium smyrnioides* (Tao et al. 2008) and belonged to the genera *Smyrnieae* Koch which was supported based on the variation of psbA-trnH sequence (Song et al. 2014). In order to provide more evidence to solve the dispute throughout the current study, we performed phylogenetic analysis based on 80 common protein coding sequences, 30 plants of *Araliaceae* and *Apiaceae* with *Nicotiana tabacum* as an outgroup (Fig. 4). The phylogenetic tree constructed with Maximum Likelihood (ML) algorithms was carried out using MEGA7, the results suggested that the closest relationship to *Chuanminshen* was *Bupleurum falcatum* which belonged to *Ammineae* Koch, *Anthriscus cerefolium* and *Daucus carota* belonging to the *Scandicineae* DC. The three closer species are *Carum carvi*, *Anethum graveolens*, and *Foeniculum vulgare* which also were classified in the *Ammineae* Koch. Our analysis results suggested that *Chuanminshen*'s phylogenetic position was distant from the *Peucedaneae* Drude, whereas only the chloroplast genome represents maternal genetic code, so further

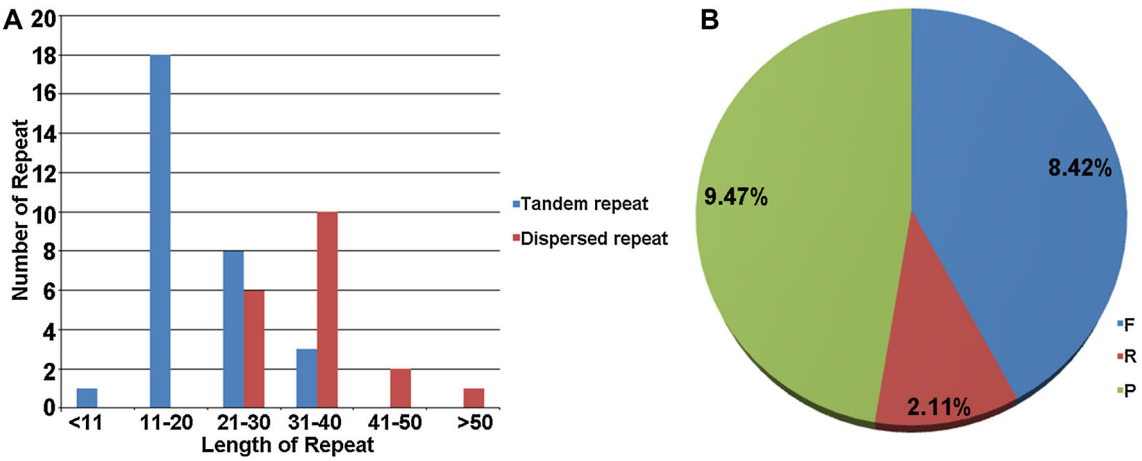


**Fig. 3** Distribution frequency of disperse and tandem repeats in *Chuanminshen* cp genome. **a** The distribution frequency of repeats, x-axis Length of repeats; y-axis Number of repeats; *Red bars* stand for dispersed repeats, *blue bars* stand for tandem repeats. **b** The percentage distribution of dispersed repeat type. *F* Forward, *P* palindromic, *R* reverse

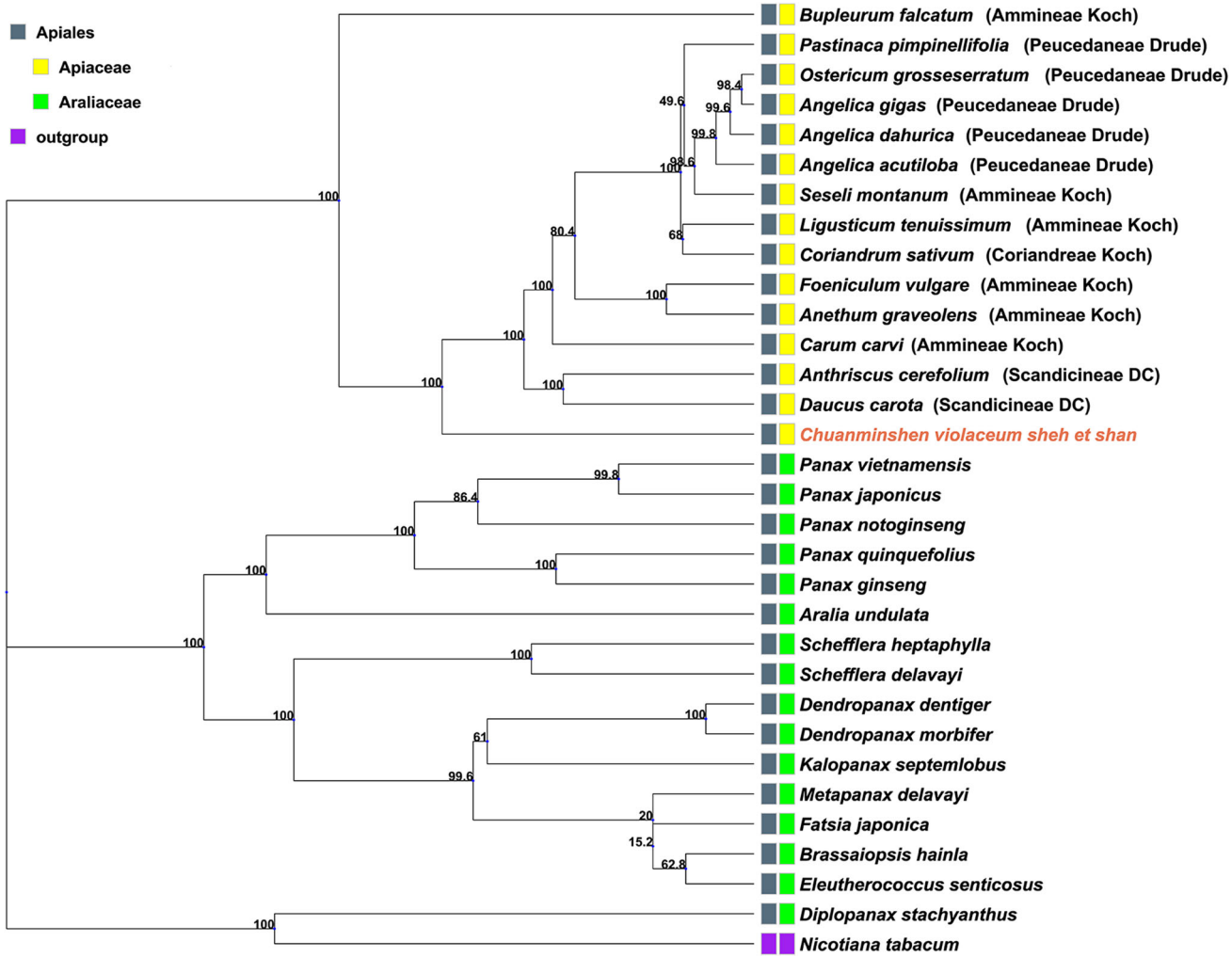


**Fig. 4** ML phylogenetic tree of 31 plants of *Apiales* based on 80 protein coding sequence of cp genomes. Numbers above each node are bootstrap support values from 1000 replicates. The *gray box* indicates the taxa that belong to the *Apliales*, the *yellow box* indicates the taxa that belongs to *Apiaceae*, the *green box* indicates the taxa that belongs to *Araliaceae*, the *purple box* indicates the taxa that belongs to the outgroup

confirmation of the phylogenetic relationship is still needed to combine the nuclear genome information.

**Acknowledgements** This work was supported by the Science and Technology Support Program of Sichuan Province, China (No. 2014FZ0053). The authors would like to thank Shangqing Hu director (Chinese herbal medicine planting research center of Sichuan province, China) for the *Chuanminshen* species identification.

**Compliance with ethical standards**

**Conflict of interest** The authors declare that they have no actual or potential conflicts of interest.

## References

Asano T, Tsudzuki T, Takahashi S, Shimada H, Kadowaki K (2004) Complete nucleotide sequence of the sugarcane (*Saccharum officinarum*) chloroplast genome: a comparative analysis of four monocot chloroplast genomes. DNA Res 11:93–99

Benson G (1999) Tandem repeats finder: a program to analyze DNA sequences. Nucleic Acids Res 27:573–580

Birky CW (2001) The inheritance of genes in mitochondria and chloroplasts: laws, mechanisms, and models. Annu Rev Genet 35:125–148

Cavalier-Smith T (2002) Chloroplast evolution: secondary symbiogenesis and multiple losses. Curr Biol 12:R62–R64

Chen D, Peng C (2011) Study on anti-fatigue and anti-oxidant effects of *Chuanminshen violaceum*. Res Pract Chin Med 25:28–30

Feng H, Fan J, Qiu H, Wang Z, Yan Z, Yuan L et al (2015) *Chuanminshen violaceum* polysaccharides improve the immune responses of foot-and-mouth disease vaccine in mice. Int J Biol Macromol 78:405–416

Flora C (1979) China flora editorial board. Academic Press by Beijing science, CHN, ISBN, Beijing

Grassi F, Labra M, Scienza A, Imazio S (2002) Chloroplast SSR markers to assess DNA diversity in wild and cultivated grapevines. Vitis 41:157–158

Jansen RK, Raubeson LA, Boore JL, dePamphilis CW, Chumley TW, Haberle RC et al (2005) Methods for obtaining and analyzing whole chloroplast genome sequences. Methods Enzymol 395:348–384

Kim KJ, Lee HL (2004) Complete chloroplast genome sequences from Korean ginseng (Panax schinseng Nees) and comparative analysis of sequence evolution among 17 vascular plants. DNA Res 11:247–261

Kumar S, Stecher G, Tamura K (2016) MEGA7: molecular evolutionary genetics analysis version 7.0 for bigger datasets. Mol Biol Evol 33:1870–1874

Kurtz S, Schleiermacher C (1999) REPuter: fast computation of maximal repeats in complete genomes. Bioinformatics 15:426–427

Lei X, Zhang M (2012) Materia Medica. Pharm Clin Chin 3:34–38

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

Luo R, Liu B, Xie Y, Li Z, Huang W, Yuan J et al (2015) Erratum: SOAPdenovo2: an empirically improved memory-efficient short-read de novo assembler. Gigascience 4:30

Neuhaus HE, Emes MJ (2000) Nonphotosynthetic metabolism in plastids. Annu Rev Plant Physiol Plant Mol Biol 51: 111–140

Nie X, Lv S, Zhang Y, Du X, Wang L, Biradar SS et al (2012) Complete chloroplast genome sequence of a major invasive species, crofton weed (*Ageratina adenophora*). PLoS ONE 7:e36869

Parida SK, Yadava DK, Mohapatra T (2010) Microsatellites in *Brassica unigenes*: relative abundance, marker design, and use in comparative physical mapping and genome analysis. Genome 53:55–67

Patel RK, Jain M (2012) NGS QC Toolkit: a toolkit for quality control of next generation sequencing data. PLoS ONE 7:e30619

Powell W, Morgante M, McDevitt R, Vendramin GG, Rafalski JA (1995) Polymorphic simple sequence repeat regions in chloroplast genomes: applications to the population genetics of pines. Proc Natl Acad Sci U S A 92:7759–7763

Qian J, Song J, Gao H, Zhu Y, Xu J, Pang X et al (2013) The complete chloroplast genome sequence of the medicinal plant *Salvia miltiorrhiza*. PLoS ONE 8:e57607

Ruhlman T, Lee SB, Jansen RK, Hostetler JB, Tallon LJ, Town CD et al (2006) Complete plastid genome sequence of *Daucus carota*: implications for biotechnology and phylogeny of angiosperms. BMC Genom 7:222

Samson N, Bausher MG, Lee SB, Jansen RK, Daniell H (2007) The complete nucleotide sequence of the coffee (*Coffea arabica* L.) chloroplast genome: organization and implications for biotechnology and phylogenetic relationships amongst angiosperms. Plant Biotechnol J 5:339–353

She M, Shan R (1980) *Cyclorhiza* and *Chuanminshen*: two newly proposed genera in *Umbelliferae* (*Apiaceae*). Acta Phytotax Sin 18:45–49

Shinozaki K, Ohme M, Tanaka M, Wakasugi T, Hayashida N, Matsubayashi T et al (1986) The complete nucleotide sequence of the tobacco chloroplast genome: its gene organization and expression. EMBO J 5:2043–2049

Song X, Yin Z, Li L, Cheng A, Jia R, Xu J et al (2013) Antiviral activity of sulfated *Chuanminshen violaceum* polysaccharide against duck enteritis virus in vitro. Antivir Res 98:344–351

Song C, Wu B, Zhou W, Liu Q (2014) Analyses on relationship and taxonomic position of *Chuanminshen* Sheh et Shan (*Apiaceae*) based on variation of psbA-trnH sequence. Plant Resour Environ 23:19–26

Sugiura M (1992) The chloroplast genome. Plant Mol Biol 19:149–168

Tao X, Gui X, Fu C, Qiu Y (2008) Analysis of genetic differentiation and phylogenetic relationship between *Changium smyrnioides* and *Chuanminshen violaceum* using molecular markers and ITS sequences. J Zhejiang Univ (Agric Life Sci) 34:473–481

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 94:302–312

Wicke S, Schneeweiss GM, dePamphilis CW, Muller KF, Quandt D (2011) The evolution of the plastid chromosome in land plants: gene content, gene order, gene function. Plant Mol Biol 76:273–297

Wyman SK, Jansen RK, Boore JL (2004) Automatic annotation of organellar genomes with DOGMA. Bioinformatics 20:3252–3255

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

Yi DK, Kim KJ (2012) Complete chloroplast genome sequences of important oilseed crop *Sesamum indicum* L. PLoS ONE 7:e35872

Zhang M, Yu T, Su X, Zhang H (2007) Physiochemical properties and the immunological activity of the *Chuanmingshen* polysaccharide. West China J Pharm Sci 22:396–398

Zhang YJ, Ma PF, Li DZ (2011) High-throughput sequencing of six bamboo chloroplast genomes: phylogenetic implications for temperate woody bamboos (*Poaceae*: *Bambusoideae*). PLoS ONE 6:e20596

Zhang X, Luo H, Xu Z, Zhu Y, Ji A, Song J et al (2015) Genome-wide characterisation and analysis of bHLH transcription factors related to tanshinone biosynthesis in *Salvia miltiorrhiza*. Sci Rep 5:11244

Zhang Y, Du L, Liu A, Chen J, Wu L, Hu W et al (2016) The complete chloroplast genome sequences of five *Epimedium* species: lights into phylogenetic and taxonomic analyses. Front Plant Sci 7:306