

Characterization of the complete chloroplast genome of the Chinese cherry *Prunus pseudocerasus* (Rosaceae)

Ying Feng¹ · Tao Liu¹ · Xiao-Yu Wang¹ · Bin-Bin Li¹ · Cheng-Lin Liang¹ · Yu-Liang Cai¹

Received: 3 May 2017 / Accepted: 8 May 2017 / Published online: 18 May 2017
© The Author(s) 2017. This article is an open access publication

Abstract The Chinese cherry *Prunus pseudocerasus* is a fruit tree species with high economic and ornamental values. Many of its wild populations are under threat or even on the verge of extinction. Here, its complete chloroplast genome was assembled using next-generation sequencing technology. The circular genome is 157,834 bp long, and contains a pair of inverted repeat (IR) regions of 26,398 bp each, separated by a large single-copy (LSC) region of 85,954 bp and a small single-copy (SSC) region of 19,084 bp. It encodes a total of 131 genes, including 86 protein-coding genes (78 PCG species), 37 tRNA genes (29 tRNA species) and eight rRNA genes (four rRNA species). Its base composition is biased (31.2% A, 18.7% C, 18.0% G & 32.1% T) with an overall A + T content of 63.3%. Phylogenetic analysis indicated that *P. pseudocerasus* is closely related to the congeners *P. maximowiczii*, *P. serrulata*, *P. subhirtella* and *P. yedoensis*.

Keywords Cherry · Chloroplast genome · Illumina sequencing · MITObim · *Prunus pseudocerasus*

The Chinese cherry *Prunus pseudocerasus* is a species of fruit tree within the family Rosaceae (Li and Bartholomew 2003). It originates in Southwest China, but now is widely dispersed in the temperate zone of Northern Hemisphere, mostly occurring on sunny mountain slopes or on the sides of ravines with an elevation of 300–1200 m (Chen et al.

2016; Li and Bartholomew 2003; Li et al. 2009). *P. pseudocerasus* possesses high economic and ornamental values. As a traditional fruit with peculiar flavor, its cultivation history can date back to approximately 3000 years ago in China (Liu and Liu 1993). Its fruit contains a variety of nutritional ingredients and trace elements, e.g. carotene, vitamin C, proteins, saccharides, iron and phosphorus (Yu and Li 1986). It has also long been used as the rootstock for sweet cherry ever since the latter's introduction into China (Zhang and Gu 2016). Besides, the past decades has witnessed the significant role of Chinese cherry landscape in the booming rural tourism industry (Chen et al. 2016).

However, as indicated by recent field surveys (Chen et al. 2016; Li et al. 2009), many wild populations of *P. pseudocerasus* are under threat or even on the verge of extinction largely due to anthropogenic activities (e.g. road construction, deforestation, grazing and thoroughbred replacement). Urgent preservation and restoration practices have become necessary for this valuable germplasm. A good knowledge of its genetic diversity would be essential to the formulation of efficient strategies for its conservation, management and exploitation. To facilitate such purposes, its complete chloroplast (cp) genome was assembled from high-throughput Illumina sequencing data in this study. The annotated genomic sequence is available from GenBank with the accession number KX255667.

Total genomic DNA was extracted from silica-dried leaves of an individual with the DNeasy Plant Mini Kit (Qiagen, CA, USA), and used for the shotgun library preparation following the manufacturer's protocol for the Illumina NextSeq 500 Sequencing System (Illumina, CA, USA). In all, 8.63 M of 150-bp paired raw reads were obtained, quality-trimmed with Trimmomatic v0.35 (Bolger et al. 2014), and used for the assembly of cp genome with MITObim v1.8 (Hahn et al. 2013). The cp

✉ Yu-Liang Cai
cylxlc0673@sina.com

¹ College of Horticulture, Northwest A&F University,
3 Taicheng Road, Yangling 712100, Shaanxi,
People's Republic of China

genome of *Prunus persica* (HQ336405) (Jansen et al. 2011) was used as the initial reference as well as for the purpose of genome annotation. A physical map of the genome was drawn with the web-based tool OrganellarGenomeDraw (OGDRAW) (<http://ogdraw.mpimp-golm.mpg.de/>) (Lohse et al. 2013).

The cp genome of *P. pseudocerasus* was successfully assembled with an average coverage of 235-fold. It is 157,834 bp in length, and contains a pair of inverted repeat (IR) regions of 26,398 bp each, separated by a large single-copy (LSC) region of 85,954 bp and a small single-copy (SSC) region of 19,084 bp (Fig. 1). It harbors 131 genes, including 86 protein-coding genes (78 PCG species), 37 tRNA genes (29 tRNA species) and eight rRNA genes (four rRNA species). The majority of

the gene species occur as a single copy, whereas 20 gene species occur in double copies, including eight PCG species (*ndhB*, *rpl2*, *rpl23*, *rps7*, *rps12*, *rps19*, *ycf1* & *ycf2*), eight tRNA species (*trnA-UGC*, *trnG-GCC*, *trnI-CAU*, *trnI-GAU*, *trnL-CAA*, *trnN-GUU*, *trnR-ACG* & *trnV-GAC*) and all four rRNA species (*rrn4.5*, *rrn5*, *rrn16* & *rrn23*). Except for *trnG-GCC* which resides within the LSC region, all the other 19 duplicated gene species are partially or completely located within the IR regions. Ten PCG species (*atpF*, *ndhA*, *ndhB*, *petB*, *petD*, *rpl2*, *rpl16*, *rpoC1*, *rps12* & *rps16*) and six tRNA species (*trnA-UGC*, *trnG-GCC*, *trnI-GAU*, *trnK-UUU*, *trnL-UAA* & *trnV-UAC*) harbor a single intron, while two other PCG species (*clpP* & *ycf3*) have a couple of introns. This cp genome has an biased base composition (31.2% A, 18.7%

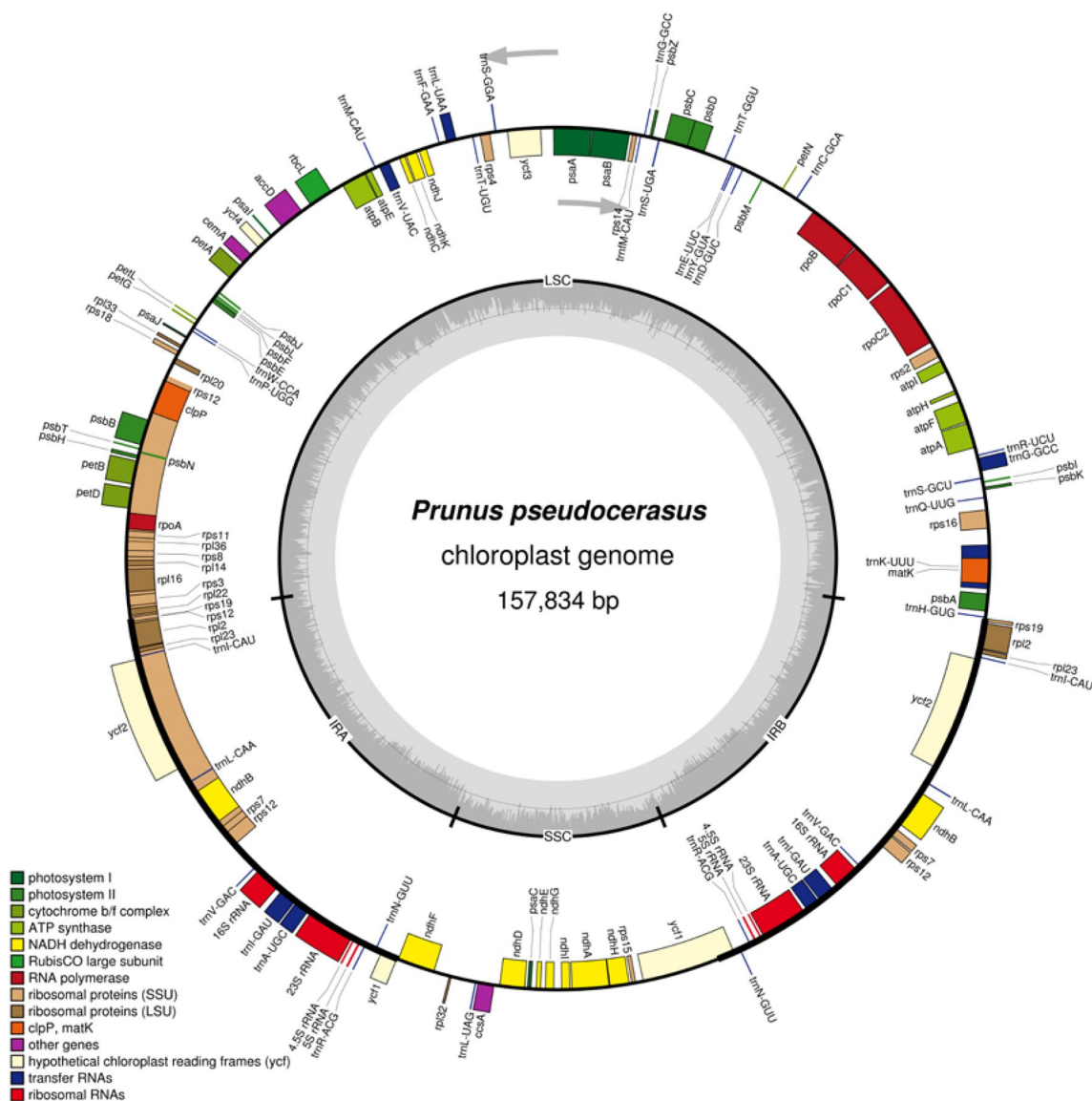
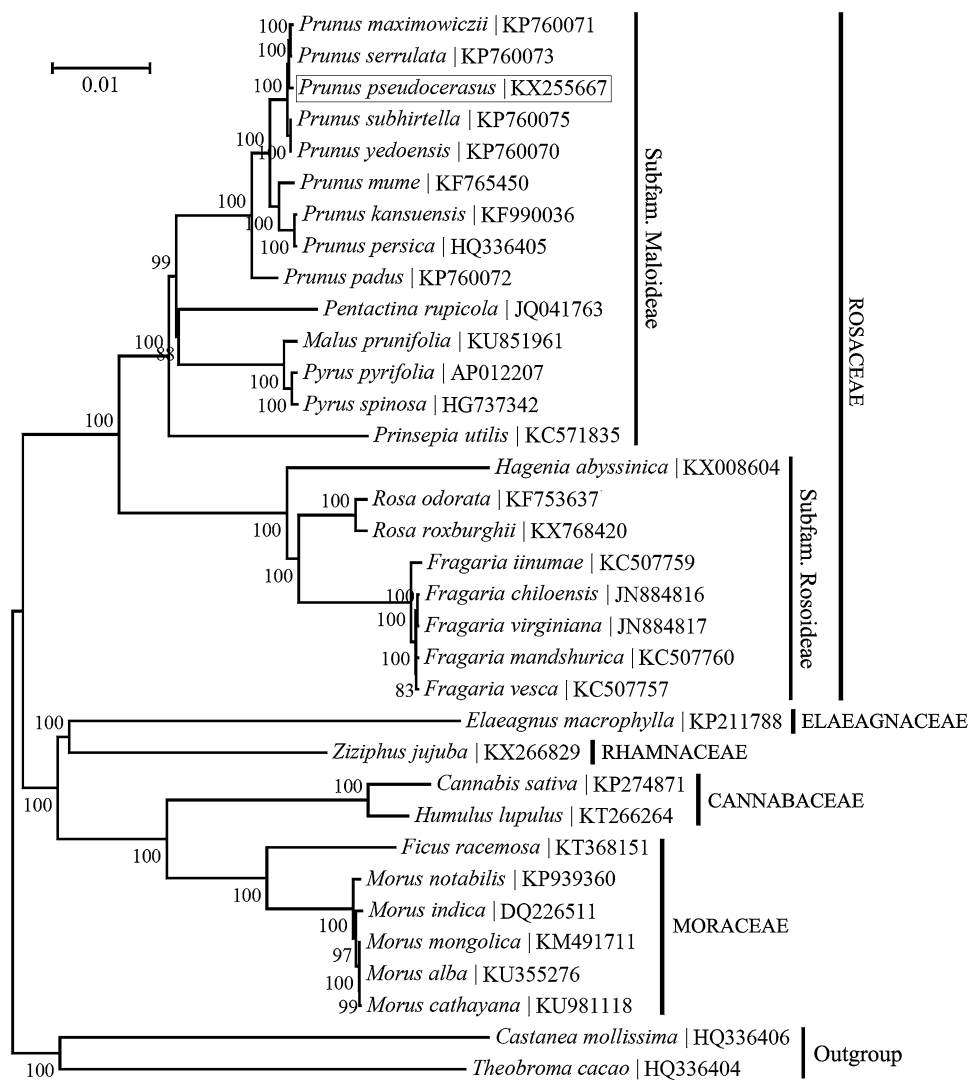


Fig. 1 Physical map of the chloroplast genome of *Prunus pseudocerasus*

Fig. 2 Phylogeny of 32 species within the order Rosales based on the neighbor-joining (NJ) analysis of the concatenated coding sequences of chloroplast PCGs. The bootstrap values were based on 500 resamplings, and are indicated next to the branches. The tree was rooted with *Castanea mollissima* and *Theobroma cacao*



C, 18.0% G & 32.1% T) with an overall A+T content of 63.3%. The A+T contents of the LSC, SSC and IR regions are 65.4, 69.7 and 57.5%, respectively.

A good knowledge of its genetic relationship with related taxa would provide valuable background information for broadening the genetic basis of rootstock breeding programs (Zhang and Gu 2016). Thus, we further investigated its phylogenetic relationships with another 31 taxa with publicly available cp genomes within the order Rosales (Fig. 2). A neighbor-joining (NJ) phylogeny was reconstructed using the concatenated coding sequences of cp PCGs with MEGA6 (Tamura et al. 2013). The phylogenetic analysis corroborated the traditional taxonomy of the order Rosales with high bootstrap support. Specifically, the 22 species within the family Rosaceae were clustered into two groups, corresponding to the two distinct subfamilies: Maloideae and Rosoideae. *P. pseudocerasus* was found to be closely related to the four congeners *P. maximowiczii*, *P. serrulata*, *P. subhirtella* and *P. yedoensis*.

Acknowledgements This study was financially supported by the Chinese Government Forestry Technology Popularization and Demonstration Project (SLTG[2016]15-2) and the International Sci-Tech Cooperation Project between China and Hungary (2016YFE0130900).

Open Access This article is distributed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made.

References

- Bolger AM, Lohse M, Usadel B (2014) Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* 30:2114–2120. doi:10.1093/bioinformatics/btu170
- Chen T, Huang X-J, Zhang J, Chen Q, Liu Y, Tang H-R, Pan D-M, Wang X-R (2016) Genetic diversity and population

- structure patterns in Chinese cherry (*Prunus pseudocerasus* Lindl) landraces. *Plant Mol Biol Report* 34:440–453. doi:[10.1007/s11105-015-0934-2](https://doi.org/10.1007/s11105-015-0934-2)
- Hahn C, Bachmann L, Chevreux B (2013) Reconstructing mitochondrial genomes directly from genomic next-generation sequencing reads—a baiting and iterative mapping approach. *Nucl Acids Res* 41:e129. doi:[10.1093/nar/gkt371](https://doi.org/10.1093/nar/gkt371)
- Jansen RK, Sasaki C, Lee S-B, Hansen AK, Daniell H (2011) Complete plastid genome sequences of three rosids (*Castanea*, *Prunus*, *Theobroma*): evidence for at least two independent transfers of *rpl22* to the nucleus. *Mol Biol Evol* 28:835–847. doi:[10.1093/molbev/msq261](https://doi.org/10.1093/molbev/msq261)
- Li C, Bartholomew B (2003) *Cerasus* miller. *Flora China* 9:404–420
- Li M-M, Cai Y-L, Qian Z-Q, Zhao G-F (2009) Genetic diversity and differentiation in Chinese sour cherry *Prunus pseudocerasus* Lindl., and its implications for conservation. *Genet Resour Crop Evol* 56:455–464. doi:[10.1007/s10722-008-9378-y](https://doi.org/10.1007/s10722-008-9378-y)
- Liu C-J, Liu M-Z (1993) Identification of seed remains in the copper casting sites in Houma, Shanxi Province. Cultural Relics Press, Beijing
- Lohse M, Drechsel O, Kahlau S, Bock R (2013) OrganellarGenomeDRAW—a suite of tools for generating physical maps of plastid and mitochondrial genomes and visualizing expression data sets. *Nucl Acids Res* 41:W575–W581. doi:[10.1093/nar/gkt289](https://doi.org/10.1093/nar/gkt289)
- Tamura K, Stecher G, Peterson D, Filipski A, Kumar S (2013) MEGA6: molecular evolutionary genetics analysis version 6.0. *Mol Biol Evol* 30:2725–2729. doi:[10.1093/molbev/mst197](https://doi.org/10.1093/molbev/mst197)
- Yu D-J, Li C-L (1986) *Flora of China*, vol 38. Science Press, Beijing
- Zhang Q, Gu D (2016) Genetic relationships among 10 *Prunus* root-stock species from China, based on simple sequence repeat markers. *J Amer Soc Hort Sci* 141:520–526. doi:[10.21273/jashs03827-16](https://doi.org/10.21273/jashs03827-16)