GMR

Characterization of the complete Chloroplast genome of *Correa carmen*, a valuable winter-flowering shrub

Minghui Chen, Yuxia Zong, Shiping Cheng*, Zhilu Zhang

Pingdingshan University, Pingdingshan 467000, Henan Province, People's Republic of China

Corresponding author: Shiping Cheng

E-mail: shipingcheng@163.com

Genet. Mol. Res. 16 (4): gmr16039815

Received September 29, 2017

Accepted October 17, 2017

Published October 21, 2017

DOI http://dx.doi.org/10.4238/gmr16039815

ABSTRACT. *Correa carmen* is considered important because of its considerable ornamental and economic value. The most striking characteristic of *C. carmen* flowers is their long winter-flowering period. In the present study, we generated the first complete *C. carmen* chloroplast genome sequence based on Illumina paired-end sequencing data. The entire chloroplast genome comprises a circular molecule with 156,759 bp that forms a quadripartite organization with two inverted repeats (26,981 bp) separated by large (84,887 bp) and small (17,910 bp) single copy sequences. The *C. carmen* genome includes 95 protein-coding genes, 31 transfer RNA genes, and eight

ribosomal RNA genes. Additionally, the base composition of the genome is biased (30.38% A, 18.92% C, 19.63% G, and 31.07% T) with an overall GC content of 38.55%. The results of a phylogenetic analysis are consistent with the traditional taxonomic framework of the family Rutaceae, and *C. carmen* is closely related to Phellodendron amurense.

Keywords: *Correa carmen*, chloroplast genome, Illumina sequencing, exploitation

SHORT COMMUNICATION

Correa carmen is an evergreen shrub in the genus *Correa* and the family Rutaceae. It is considered important because of its considerable ornamental and economic value. The most striking feature of *C. carmen* flowers is their long winter-flowering period. This plant species originated in Australia, and has been widely used as an ornamental plant in China since it was introduced in 2008 (personal communication). A thorough characterization of its genetic diversity is essential for formulating efficient strategies to manage and exploit *C. carmen* cultivation and clarify its taxonomic classification.

Chloroplast genome sequences are of great phylogenetic, conservation genetic, and population genetic value due to their relatively conserved structure and comparatively high substitution rates (Ravi et al. 2008; Li et al. 2016). Accordingly, we assembled the complete chloroplast (cp) genome using high-throughput Illumina sequencing data. To the best of our knowledge, this is the first report describing a genetic resource for the genus *Correa* (GenBank accession number: 150172998648951). It is reasonable to speculate that the cp genome not only represents a useful resource that may be exploited, it may also be relevant for inferring phylogenetic relationships between *C. carmen* and other related species.

Fresh leaves were collected from an adult *C. carmen* plant grown in Aletai county (Xinjiang, China; 47.70°N, 88.62°E). Total genomic DNA was isolated using an improved CTAB method (Doyle and Doyle 1987) and the DNA concentration and quality was quantifed using a Nanodrop spectrophotometer (Thermo Scientifc, Carlsbad, CA, USA). Then the checked DNA then sequenced with the HiSeq 2500 platform (Illumina, San Diego, CA, USA). The paired-end reads were assembled using the SOAPdenovo software (Luo et al. 2012) with a k-mer value of 64.

The obtained contigs were then filtered with a customized python script. A reference-guided assembly was then completed to reconstruct the chloroplast genome, with sequences generated by a BLAST search of closely related species (*Phellodendron amurense, Zanthoxylum schinifolium, Z. bungeanum*) applied as references. Additionally, the GapCloser program (http://soap.genomics.org.cn/index.html) and CpGAVAS (Liu et al. 2012) were used to fill the gaps and annotate the cp genome, respectively. The assembled genome was annotated by Geneious R10 (Biomatters Ltd., Auckland, New Zealand) and manually checked for start and stop codons and intron/exon boundaries. The transfer RNA (tRNA) sequences were confirmed using the online tools tRNA scan-

SE Search Service (Schattner et al. 2005). The final genome map was generated with OrganellarGenomeDRAW (http://ogdraw.mpimp-golm.mpg.de/; Lohse et al. 2013).

The complete *C. carmen* cp genome is 156,759 bp long and consists of a typical quadripartite structure, with one large single-copy (LSC, 84,887 bp) region, one small single-copy (SSC, 17,910 bp) region, and a pair of inverted repeat (IR, 26,981 bp) regions (Figure 1).

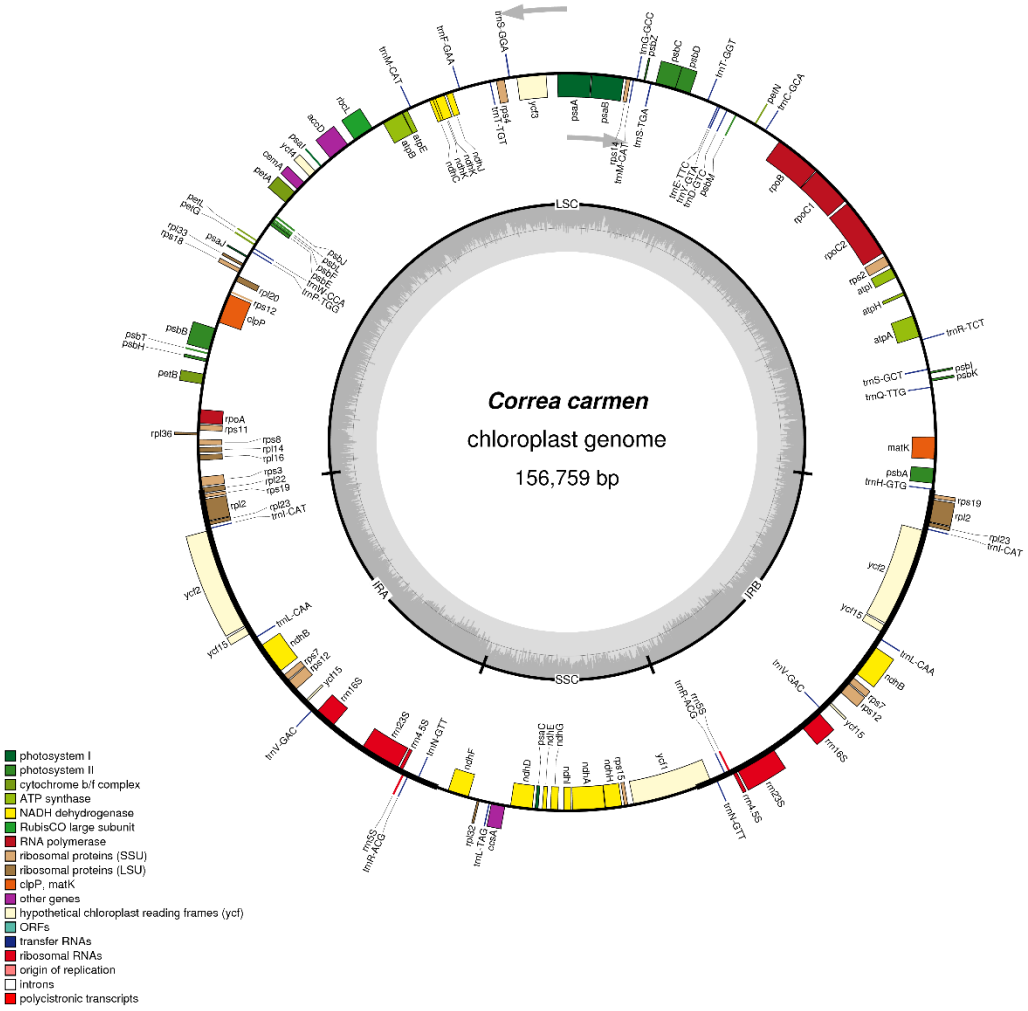

Figure 1. Physical map of the *Correa carmen* chloroplast genome. Genes shown outside the outer circle are transcribed in the clockwise direction, whereas those inside are transcribed in the counterclockwise direction. The *colored bars* indicate known protein-coding genes, tRNA and rRNA. Areas *dashed light and darker gray* in the inner circle indicates the A+T and G+C contents of the genome, respectively. LSC, large single-copy; SSC, small single-copy; IR, inverted repeat.

The cp genome encodes 134 genes, of which 95 are protein-coding genes (PCGs), 31 are tRNA genes, and eight are ribosomal RNA (rRNA) genes (Table 1).

Table 1 Genes present in chloroplast genome of Correa carmen (134 genes in total).

Functions	Family Name	Code	List of Genes
Self-replication	Small subunit of ribosome	rps	*rps2*, *rps14*, *rps4*, *rps18*, *rps11*, *rps8*, *rps3*, *rps19*, *rps7*, *rps15*, *rps7*, *rps19*, *rps12*, *rps12*
	rRNA Genes	rrn	*rrn16S*, *rrn23S*, *rrn4.5S*, *rrn5S*, *rrn5S*, *rrn4.5S*, *rrn23S*, *rrn16S*
	Large subunit of ribosome	rpl	*rpl33*, rpl20, *rpl36*, *rpl14*, *rpl16*, *rpl22*, *rpl2*, *rpl23*, *rpl32*, *rpl23*, *rpl2*
	DNA dependent RNA polymerase	rpo	*rpoC2*, *rpoC1*, *rpoB*, *rpoA*
	tRNA Genes	trn	trnH-*GTG*, *trnQ-TTG*, *trnS-GCT*, *trnR-TCT*, *trnC-GCA*, *trnD-GTC*, *trnY-GTA*, *trnE-TTC*, *trnT-GGT*, *trnS-TGA*, *trnF-GAA*, *trnG-GCC*, *trnM-CAT*, *trnS-GGA*, *trnT-TGT*, *trnF-GAA*, *trnM-CAT*, *trnW-CCA*, *trnP-TGG*, *trnI-CAT*, *trnL-CAA*, *trnV-GAC*, *trnR-ACG*, *trnN-GTT*, *trnL-TAG*, *trnN-GTT*, *trnR-ACG*, *trnV-GAC*, *trnL-CAA*, *trnI-CAT*
Photosynthesis	Subunits of ATP synthase	atp	*atpA*, *atpH*, *atpI*, *atpE*, *atpB*
	Subunits of protochlorophyllide reductase	chl	
	Subunits of NADH-dehydrogenase	ndh	*ndhJ*, *ndhK*(pseudogene), *ndhK*, *ndhC*, *ndhB*(pseudogene), *ndhF*, *ndhD*, *ndhE*, *ndhG*, *ndhI*, *ndhA*, *ndhH*, *ndhB*(pseudogene)
	Subunits of cytochrome b/f complex	pet	*petN*, *petA*, *petL*, *petG*, *petB*
	Subunits of photosystem I	psa	*psaB*, *psaA*, *psaI*, *psaJ*, *psaC*
	Subunits of photosystem II	psb	*psbA*, *psbK*, psbI, *psbM*, *psbD*, *psbC*, *psbZ*, *psbJ*, *psbL*, *psbF*, *psbE*, *psbB*, *psbT*, *psbH*
	Subunit of rubisco	rbc	*rbcL*
Other genes	Subunit of Acetyl-CoA-carboxylase	acc	*accD*
	Envelop membrane protein	cem	*cemA*
	c-type cytochrom synthesis gene	ccs	*ccsA*
	Protease	clp	*clpP*(pseudogene)
	Translational initiation factor	inf	
	Maturase	mat	*matK*
	Elongation factor	tuf	
Unkown function	Conserved open reading frames	ycf	*ycf3*, *ycf4*, *ycf2*, *ycf15*, *ycf15*, *ycf1*, *ycf15*, *ycf15*, *ycf2*

Among these genes, 12 (*trnF-GAA*, *rpoC1*, *psaA*, *rpl2*, *ycf2*, *ycf15*, *ndhA*, *ycf15*, *ycf2*, *rpl2*, *rps12*, and *rps12*) contain one intron and only one (*ycf3*) carries two introns. The majority of the genes occur as a single copy, while 18 genes occur as two copies, including seven PCG genes (*rps19*, *rpl2*, *rp123*, *ycf2*, *ycf15*, *ccsA*, *rps7*, and *rps12*), seven tRNA genes (*trnF-GAA*, *trnM-CAT*, *trnI-CAT*, *trnL-CAA*, *trnV-GAC*, *trnR-ACG*, and *trnN-GTT*), and all four rRNA genes (*rrn16S*, *rrn23S*, *rrn4.5S*, and *rrn5S*). Moreover, there are four copies of *ycf15*, which is a PCG gene. These 20 genes are completely or partially located within the IR regions. Furthermore, the sequenced cp genome has a biased nucleotide composition (30.38% A, 18.92% C, 19.63% G,

and 31.07% T). The overall A + T content (61.45%) is higher than that of the IR regions (56.98%), but lower than those of the LSC (63.18%) and SSC (66.69%) regions.

To ascertain the phylogenetic position of *C. carmen* within the order Rutaceae, a neighbor-joining phylogenetic tree was reconstructed with the MEGA7 program using the concatenated sequences of cp PCGs for 26 species. The phylogenetic relationships uncovered here are consistent with the morpho-taxonomy of the order Sapindales (Figure 2).

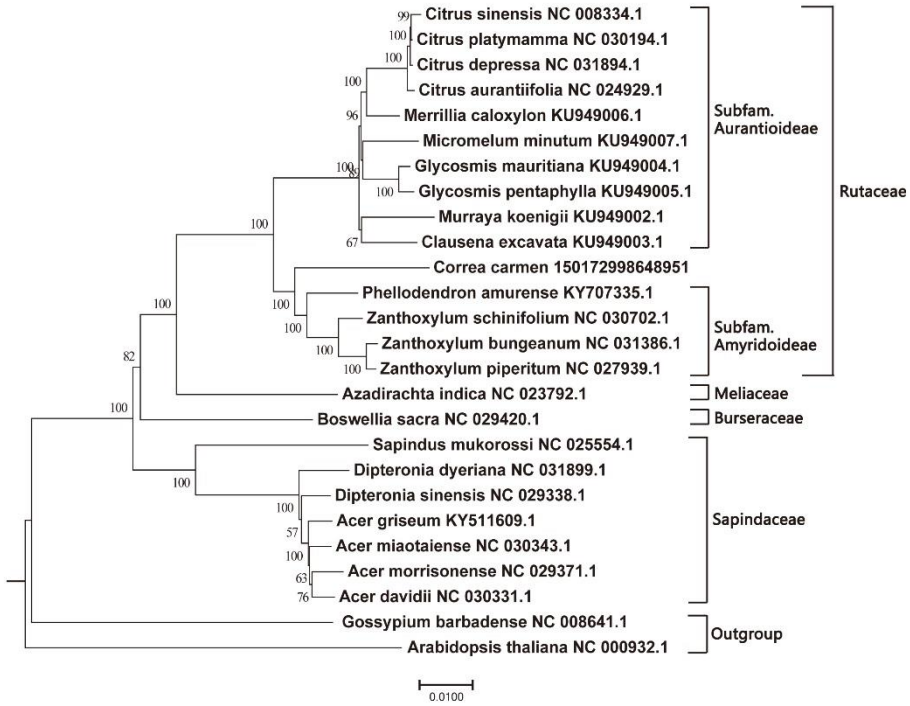

Figure 2. Maximum likelihood phylogenetic tree of *Correa carmen* with 25 other species based on complete chloroplast genome sequences using *Arabidopsis thaliana* and *Gossypium barbadense* as outgroup. Numbers on the nodes are bootstrap values with 1000 replicates. Accession numbers are listed as below: *Citrus sinensis* (NC 008334.1), *Citrus platymamma* (NC 030194.1), *Citrus depressa* (NC 031894.1), *Citrus aurantiifolia* (NC 024929.1), *Merrillia caloxylon* (KU949006.1), *Micromelum minutum* (KU949007.1), *Glycosmis mauritiana* (KU949004.1), *Glycosmis pentaphylla* (KU949005.1), *Murraya koenigii* (KU949002.1), Clausena *excavata* (KU949003.1), *Correa carmen* (150172998648951), *Phellodendron amurense* (KY707335.1), *Zanthoxylum schinifolium* (NC 030702.1), *Zanthoxylum bungeanum* (NC 031386.1), *Zanthoxylum piperitum* (NC 027939.1), *Azadirachta indica* (NC 023792.1), *Boswellia sacra* (NC 029420.1), *Sapindus mukorossi* (NC 025554.1), *Dipteronia dyeriana* (NC 031899.1), *Dipteronia sinensis* (NC 029338.1), *Acer griseum* (KY511609.1), *Acer miaotaiense* (NC 030343.1), *Acer morrisonense* (NC 029371.1), *Acer davidii* (NC 030331.1), *Gossypium barbadense* (NC 008641.1), *Arabidopsis thaliana* (NC 000932.1). Family- and subfamily-level taxonomy is presented for each taxon.

Specifically, the 15 taxa within the family Rutaceae are further clustered into two monophyletic subclades with high bootstrap support. Moreover, a close relationship was observed between *C. carmen* and *P. amurense*, which belonged to the subfamily Amyridoideae. Our results confirm earlier findings on the phylogeny within Rutaceae family (Chen 2017; Liu and Shi 2017), suggesting two relatively distinct subfamilies Auranitioideae and Amyridoideae within family Rutaceae. The complete cp genome may be useful for population genomic studies of *C. carmen*. The resulting data and information may be important for formulating new potential conservation and management strategies for this species.

CONFLICT OF INTEREST STATEMENT

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

ACKNOWLEDGMENTS

CM and CS conceived and designed the experiments. CM and CS together revised and approved the final manuscript. This work was jointly supported by Scientific Research Foundation for the introduction of talent of Pingdingshan University (No. PXY-BSQD2016009), Key Research Project of Colleges and Universities of Henan Province (172102110111, 16A220004 and 162120110070) and College Students Science and Technology Innovation Project (S & TIF2017147).

REFERENCES

Chen KK (2017). Characterization of the complete chloroplast genome of the Tertiary relict tree *Phellodendron amurense* (Sapindales: Rutaceae) using Illumina sequencing technology Conservation Genetics Resources:1-4. https://doi.org/10.1007/s12686-017-0761-x

Doyle JJ, Doyle JL (1987). A rapid DNA isolation procedure for small quantities of fresh leaf tissue. Phytochem Bull 19:11-15 https://doi.org/10.1016/0026-265x(63)90057-2

Li B, Li Y, Cai Q (2016). Development of chloroplast genomic resources for Akebia quinata (Lardizabalaceae). Conserv Genet Resour 8:447–449 https://doi.org/10.1007/s12686-016-0593-0

Liu C, Shi LC, Zhu YJ (2012). CpGAVAS, an integrated web server for the annotation, visualization, analysis, and GenBank submission of completely sequenced chloroplast genome sequences. BMC Genom 13:715 https://doi.org/10.1186/1471-2164-13-715

Liu J, Shi C (2017). The complete chloroplast genome of wild shaddock, Citrus maxima (Burm.) Merr. Conservation Genet Resour. 1-3 https://doi.org/10.1007/s12686-017-0733-1

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. Nucleic Acids Res 41:575-581. https://doi.org/10.1093/nar/gkt289

Luo R, Liu B, Xie Y, Li Z, et al. (2012). SOAPdenovo2: an empirically improved memory-efcient short-read de novo assembler. Gigascience 1(1):1 https://doi.org/10.1186/2047-217x-1-18

Ravi V, Khurana JP, Tyagi AK, Khurana P (2008). An update on chloroplast genomes. Plant Syst Evol 271:101-122

Schattner P, Brooks AN, Lowe TM (2005) The tRNAscan-SE, snoscan and snoGPS web servers for the detection of tRNAs and snoRNAs. Nucleic Acids Res 33:686-689. https://doi.org/10.1093/nar/gki366