

## The complete chloroplast genome of *Curcuma longa* L. (Zingiberaceae)

Jie Luo<sup>a</sup>, Mingzhi Li<sup>b</sup> and Guo Yang<sup>c</sup>

<sup>a</sup>Academy of Yuanpei, Shaoxing University, Shaoxing, China; <sup>b</sup>Independent Researcher, Guangzhou, China; <sup>c</sup>Academy of Life Science, Shaoxing University, Shaoxing, China

### ABSTRACT

*Curcuma longa* L. is a widely distributed medicinal plant in Southeast Asia, especially in China. In this study, we reported and characterized the complete chloroplast genome sequence of *C. longa*. The chloroplast genome was determined to be 162,255 bp in length. It contained large single-copy (LSC) and small single-copy (SSC) regions of 86,975 bp and 15,776 bp, respectively, which were separated by a pair of 29,752 bp inverted repeat (IR) regions. The genome is predicted to contain 133 genes, including 87 protein-coding genes, 38 tRNA genes, and eight rRNA genes. The overall GC content of the genome is 36.2%. A phylogenetic tree reconstructed by 23 chloroplast genomes reveals that *C. longa* is mostly related to *Curcuma roscoeana*, with bootstrap support values of 100%.

### ARTICLE HISTORY

Received 27 May 2019  
Accepted 1 August 2019

### KEYWORDS

*Curcuma longa*; complete chloroplast genome; phylogenetic analysis

The genus *Curcuma* comprises more than 50 species and distributed widely in Southeast Asia (Fang et al. 1999). There are about five species of this genus in China. *Curcuma longa* is a widely distributed medicinal plant in South and Southwest China. But, to date, there is still no complete cp genome was characterized for *C. longa*. Here, we characterized the complete chloroplast (cp) genome sequence of *C. longa*; the first complete chloroplast genome sequence from the genus *C. longa*, based on the genome skimming sequencing data.

The total genomic DNA was extracted from the fresh leaves of *C. longa*. (South China Botanical Garden, N23°10'54.84", E113°21'46.61") using the Plant DNA Mini Kit (Genepioneer, Nanjing, China). Specimens were stored in the Herbarium of Shaoxing University (accession number: SXU-20190415JH01). The whole genome sequencing was conducted by Nanjing Genepioneer Biotechnologies Inc. (Nanjing, China) on the Illumina HiSeq platform. The filtered sequences were assembled using the program SPAdes assembler 3.10.0 (Bankevich et al. 2012). Annotation was performed using the DOGMA (Wyman et al. 2004) and BLAST searches.

The cp genome of *C. longa* was determined to comprise a 162,255 bp double stranded, circular DNA (NCBI acc. no. MK965541), and it contained two inverted repeat (IR) regions of 29,752 bp, separated by large single-copy (LSC) and small single-copy (SSC) regions of 86,975 bp and 15,776 bp, respectively. The cp genome was predicted to contain 133 genes, including 87 protein-coding genes, 38 tRNA genes, and eight rRNA genes. Seven protein-coding genes, eight tRNA genes, and four rRNA genes were duplicated in IR regions. Nineteen genes contained one intron and two genes (clpP and ycf3) contained two introns. The overall GC content of *C. longa* cp

genome is 36.2% and the corresponding values in LSC, SSC, and IR regions are 34.0%, 29.5%, and 41.1%, respectively.

To investigate its taxonomic status, alignment was performed on the 23 chloroplast genome sequences using MAFFT v7.307 (Katoh and Standley 2013), and a maximum-likelihood (ML) tree (*Sabal domingensis* and *Typha latifolia* were used as the outgroup) was constructed using FastTree version 2.1.10 (Price et al. 2010). As expected *C. longa* is mostly related to *Curcuma roscoeana*, with bootstrap support values of 100% (Figure 1), which in agreement with previously published phylogenetic studies (Sass et al. 2016). The complete cp genome sequence of *C. longa* will provide a useful resource for the conservation genetics of this species as well as for the phylogenetic studies of Zingiberaceae.

### Disclosure statement

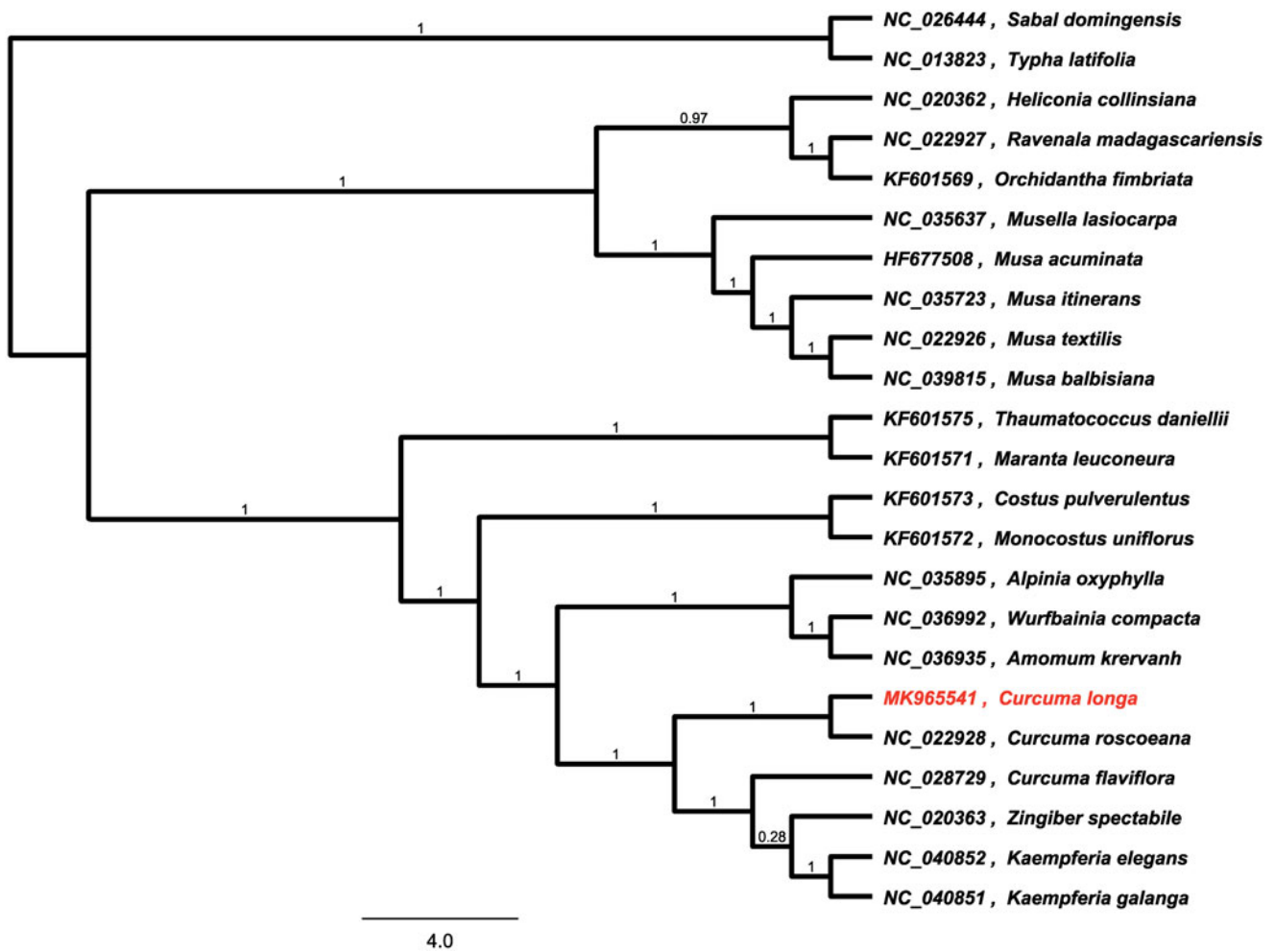
No potential conflict of interest was reported by the authors.

### Funding

The work was supported by the National Natural Science Foundation of China [31700476], Shaoxing Science and Technology Project [2015B70023], Shaoxing University Project [2016LG1013], and Xinmiao Talents Program of Zhejiang Province [2017R428015].

### References

- Bankevich A, Nurk S, Antipov D, Gurevich AA, Dvorkin M, Kulikov AS, Lesin VM, Nikolenko SI, Pham S, Prjibelski AD, et al. 2012. SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing. *J Comput Biol.* 19:455–477.



**Figure 1.** Phylogenetic tree inferred using maximum-likelihood (ML) method based on the complete chloroplast genome of 23 representative species. *Sabal domingensis* and *Typha latifolia* were used as the outgroup. A total of 1000 bootstrap replicates were computed and the bootstrap support values are shown at the branches. Genbank accession numbers were shown.

Fang ZF, Zhao SD, Skvortsov AK. 1999. Flora of China (English version). Beijing: Science Press.

Katoh K, Standley DM. 2013. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Mol Biol Evol.* 30:772–780.

Price MN, Dehal PS, Arkin AP. 2010. FastTree 2—approximately maximum-likelihood trees for large alignments. *PLoS One.* 5:e9490.

Sass C, Iles WJD, Barrett CF, Smith SY, Specht CD. 2016. Revisiting the Zingiberales: using multiplexed exon capture to resolve ancient and recent phylogenetic splits in a charismatic plant lineage. *PeerJ.* 4: e1584.

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