MITOCHONDRIAL DNA PART B: RESOURCES, 2016
VOL. 1, NO. 1, 734–736
http://dx.doi.org/10.1080/23802359.2016.1233469

Taylor & Francis
Taylor & Francis Group

MITOGENOME ANNOUNCEMENT

OPEN ACCESS

The first complete plastome sequence from the family Sapotaceae, *Pouteria campechiana* (Kunth) Baehni

Sangjin Jo, Hoe-Won Kim, Young-Kee Kim, Se-Hwan Cheon and Ki-Joong Kim

School of Life Sciences, Korea University, Seoul, Korea

ABSTRACT

In this study, we determined the complete plastome sequence of *Pouteria campechiana* (Kunth) Baehni (Sapotaceae) (NCBI acc. no. KX426215). This is the first time a plastome from the Sapotaceae has been sequenced. The gene order and structure of the *P. campechiana* plastome are collinear with those of the typical plastome of land plants. The complete plastome size is 157,922 bp in length and consists of a large single-copy region of 87,122 bp and a small single-copy region of 18,559 bp, which are separated by a pair of 26,120 bp-long inverted repeat regions. The overall A–T content of the plastome sequence is 63.2%. The plastome contains 113 genes, of which 79 are protein-coding genes, 30 are tRNA genes, and 4 are rRNA genes. Sixteen genes contain one intron and two genes have two introns. A total of 91 simple sequence repeat loci were identified within the genome. Phylogenetic analysis revealed that *P. campechiana* is a sister group of the Primulaceae-Ebenaceae clade with 100% bootstrap support.

ARTICLE HISTORY
Received 29 August 2016
Accepted 3 September 2016

KEYWORDS
Plastome; tropical fruit; Sapotaceae; *Pouteria campechiana*

Pouteria campechiana (Kunth) Baehni, commonly referred to as canistel or yellow sapote, is a widely cultivated tropical fruit that originated in southern Mexico and Central America (Kim 2011). The edible flesh of *P. campechiana* fruit has a sweet taste and boiled egg yolk-like texture. It belongs to the family Sapotaceae in the order Ericales (APG IV 2016). Sapotaceae consist of three subfamilies, 54 genera and ~1,273 species (Christenhusz & Byng 2016). The genus *Pouteria* consist of approximately 235 species, including several edible fruit species such as canistel (*P. campechiana*), green sapote (*P. viridis),* lucuma (*P. lucuma*), and mamey sapote (*P. sapota*). To the best of our knowledge, no plastome from a species within the Sapotaceae has previously been sequenced. Accordingly, the complete plastome sequence of *P. campechiana* determined in the present study will be useful not only for phylogenetic study of the group, but also in the development of molecular markers for this tropical fruit species.

The leaves of *P. campechiana* used in this study were collected from the Korea University greenhouse, where we grew the plants from seeds that were originally collected in Thailand. The plants flowered and fruited in the greenhouse. A voucher specimen was deposited in the Korea University Herbarium (KUS acc. no. 2014-0238). Fresh leaves were ground into powder in liquid nitrogen and total DNAs were extracted using the CTAB method (Doyle & Doyle 1987). The DNAs were further purified by ultracentrifugation and dialysis (Palmer 1986). The genomic DNAs are deposited in the Plant DNA Bank in Korea (PDBK acc. no. 2014-0238). The complete plastome sequence was generated using an Illumina HiSeq 2000 system (Illumina Inc., San Diego, CA). Annotations were performed using the National Center for Biotechnology Information (NCBI) BLAST, DOGMA (Wyman et al. 2004) and tRNAscan-SE programs (Lowe & Eddy 1997). For the phylogenetic analysis, we selected and downloaded 25 complete plastome sequences based on the APG IV system (APG IV 2016) from the NCBI database. All 27 plastome sequences were from super-asterid plants.

The gene order and structure of the *P. campechiana* plastome are similar to those of typical angiosperms (Shinozaki et al. 1986; Kim & Lee 2004; Yi & Kim 2012). The complete plastome is 157,922 bp in length, and consists of a large single-copy (LSC) region of 87,122 bp and a small single-copy (SSC) region of 18,560 bp, which are separated by two inverted repeats (IR) of 26,120 bp. The plastome comprises 113 unique genes (79 protein-coding genes, 30 tRNA genes and four rRNA genes). Seven protein-coding, seven tRNA and four rRNA genes are duplicated in the IR regions. The average A–T content of the plastome is 63.2%, whereas that in the LSC, SSC and IR regions is 65.3%, 69.9% and 57.2%, respectively. Sixteen genes contain one intron and two genes, *ycf3* and *clpP*, have two introns. A total of 91 simple sequence repeat (SSR) loci are distributed throughout the plastome. Among these, 75, 11 and 5 are mono-SSR, di-SSR and tri-SSR loci, respectively. Some of these loci will be useful in identifying cultivars of *P. campechiana*.

CONTACT Ki-Joong Kim kimkj@korea.ac.kr School of Life Sciences, Korea University, Seoul 02841, Korea

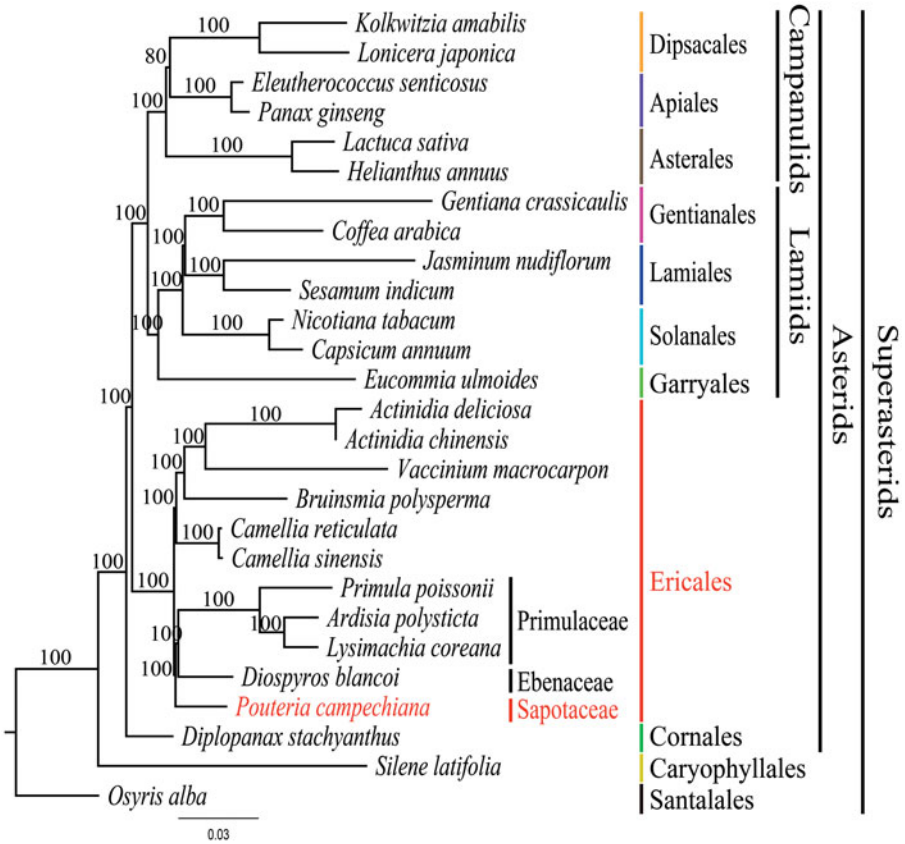

Figure 1. Maximum Likelihood (ML) tree based on 79 protein-coding and 4 rRNA genes from 27 plstomes as determined by RAxML(-ln $L = -439773.826827$). The numbers at each node indicate the ML bootstrap values. Genbank accession numbers of taxa are shown: *Actinidia chinensis* (NC_026690), *Actinidia deliciosa* (NC_026691), *Ardisia polysticta* (NC_021121), *Bruinsmia polysperma* (NC_030180), *Camellia reticulata* (NC_024663), *Camellia sinensis* (NC_020019), *Capsicum annuum* (NC_018552), *Coffea arabica* (NC_008535), *Diospyros blancoi* (KX426216), *Diplopanax stachyanthus* (NC_029750), *Eleutherococcus senticosus* (NC_016430), *Eucomia ulmoides* (KU204775), *Gentiana crassicaulis* (NC_027442), *Helianthus annuus* (NC_007977), *Jasminum nudiflorum* (NC_008407), *Kolkwitzia amabilis* (NC_029874), *Lactuca sativa* (NC_007578), *Lonicera japonica* (NC_026839), *Lysimachia coreana* (NC_026197), *Nicotiana tabacum* (NC_001879), *Osyris alba* (NC_027960), *Panax ginseng* (NC_006290), *Pouteria campechiana* (KX426215), *Primula poissonii* (NC_024543), *Sesamum indicum* (NC_016433), *Silene latifolia* (NC_016730), and *Vaccinium macrocarpon* (NC_019616).

To validate the phylogenetic relationships of *P. campechiana* in asterids, we constructed a maximum likelihood tree using 27 super-asterid taxa. Phylogenetic analysis was performed on a data set that included 79 protein-coding genes and 4 rRNA genes from the 27 selected taxa using RAxML v. 7.7.1 (Stamatakis et al. 2008). The 83 gene sequences (82,565 bp in length) were aligned with the MUSCLE program using Geneious v. 6.1.8 (Biomatters Ltd.; Kearse et al. 2012). The resulting tree shows that *P. campechiana* forms a clade with the species of Ebenaceae and Primulaceae with a 100% bootstrap value (Figure 1). The Sapotaceae represented by *P. campechiana* form a sister group to the Primulaceae–Ebenaceae clade within the Ericales. The close relationship of Sapotaceae to Lecythidaceae was suggested by a previous study (Anderberg et al. 2002); however, our tree is consistent with the phylogenetic tree recently reported by the Angiosperm Phylogeny Group (www.mobot.org/MOBOT/research/APweb), which also shows Sapotaceae as a sister group to the Primulaceae–Ebenaceae clade.

Disclosure statement

The authors report no conflicts of interest, and are independently responsible for the content and writing of the paper.

Funding

This work was supported by the National Research Foundation of Korea (NRF) under grant nos. NRF-2015M3A9B8030588 and NRF-2015M3A9B8047398 to KJK.

References

Anderberg AA, Rydin C, Källersjö M. 2002. Phylogenetic relationships in the order Ericales sl: analyses of molecular data from five genes from the plastid and mitochondrial genomes. Am J Bot. 89:677–687.

APG IV. 2016. An update of the angiosperm phylogeny group classification for the orders and families of flowering plants: APG IV. Bot J Linn Soc. 181:1–20.

Christenhusz MJ, Byng JW. 2016. The number of known plants species in the world and its annual increase. Phytotaxa. 261:201–217.

Doyle JJ, Doyle JL. 1987. A rapid DNA isolation procedure for small quantities of fresh leaf tissue. Phytochem Bull. 19:11–15.

Kearse M, Moir R, Wilson A, Stones-Havas S, Cheung M, Sturrock S, Buxton S, Cooper A, Markowitz S, Duran C. 2012. Geneious basic: an integrated and extendable desktop software platform for the organization and analysis of sequence data. Bioinformatics. 28:1647–1649.

Kim K-J, 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.

Kim K-J. 2011. Tropical fruit resources. Seoul: Geobook.

Lowe TM, Eddy SR. 1997. tRNAscan-SE: a program for improved detection of transfer RNA genes in genomic sequence. Nucleic Acids Res. 25:955–964.

Palmer JD. 1986. Isolation and structural analysis of chloroplast DNA. Method Enzymol. 118:167–186.

Shinozaki K, Ohme M, Tanaka M, Wakasugi T, Hayashida N, Matsubayashi T, Zaita N, Chunwongse J, Obokata J, Yamaguchi-Shinozaki K. 1986. The complete nucleotide sequence of the tobacco chloroplast genome: its gene organization and expression. EMBO J. 5:2043–2049.

Stamatakis A, Hoover P, Rougemont J. 2008. A rapid bootstrap algorithm for the RAxML web servers. Syst Biol. 57:758–771.

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

Yi D-K, Kim K-J. 2012. Complete chloroplast genome sequences of important oilseed crop *Sesamum indicum* L. PLoS One. 7:e35872.