






The first complete plastome of *Mimusops coriacea* (A. DC.) Miq. (Sapotaceae)

Rafaela Jorge Trad^{1,2} , Saura Rodrigues da Silva³  and Maria do Carmo Estanislau do Amaral¹ 

¹Universidade Estadual de Campinas (UNICAMP), Instituto de Biologia, Departamento de Biologia Vegetal, Campinas, SP, Brazil.

²Universidade Federal de Minas Gerais (UFMG), Instituto de Ciências Biológicas (ICB), Laboratório de Macroecologia, Belo Horizonte, MG, Brazil.

³Universidade Estadual Paulista (UNESP), Faculdade de Ciências Agrárias e Veterinárias, Departamento de Biotecnologia Agropecuária e Ambiental, Jaboticabal, SP, Brasil.

Abstract

Mimusops coriacea (A. DC.) Miq. (Sapotoideae, Sapotaceae, Ericales) is native to Madagascar and the Comoro Islands. This species is cultivated in many countries around the world and grows on sand in coastal vegetation. Here we sequenced, assembled, and annotated the first complete chloroplast genome of *M. coriacea*. The newly assembled chloroplast was analyzed with other available chloroplasts of Sapotaceae. Our results found a general conserved structure. The complete chloroplast genome has 159,689 bp, including 133 genes distributed in four regions: a large single-copy region of 88,887 bp, a small single-copy region of 18,618 bp, and two inverted repeats of 26,092 bp each. Our maximum likelihood phylogenetic tree was generated with 80 protein-coding genes and recovered a monophyletic Sapotaceae sister to a clade formed by Ebenaceae + Primulaceae. In our analysis, *Mimusops coriacea* clustered with the other eight species of Sapotaceae included in the study.

Keywords: Ericales, complete chloroplast genome, Sapotoideae.

Received: June 14, 2021; Accepted: November 28, 2021.

The family Sapotaceae (Ericales) is represented by 53 genera and between 1,100 and 1,275 species (Stevens, 2001, onwards). In Brazil, there are 13 genera, including about 240 species, which grow mostly in the Amazon and the Atlantic Forest (Alves-Araújo *et al.*, 2020). Although the genus *Mimusops* L. has 41 species only two occur in Brazil, both cultivated: *Mimusops coriacea* (A. DC.) Miq. and *M. elengi* L. (Pennington, 1990; Alves-Araújo, 2020). *Mimusops coriacea* is native only to Madagascar and to the Comoro Islands, where it grows on the sand in coastal forests and its edible fruits can be found in local markets (Gautier *et al.*, 2012). This species has also been widely cultivated in the tropics. The flowers of *M. coriacea* are similar to those of other species in the genus and are among the most complex in the family. They have two whorls of sepals, corolla lobes with lateral appendages, and staminodes (Pennington, 1990; Gautier *et al.*, 2012). The wood of *Mimusops coriacea* is used for construction, its branches and branchlets are fermented to prepare a beverage in the Antilles, and the fruits can be eaten; it is also used in traditional medicine in Ecuador (Bustamante *et al.*, 2019), and recently its phytochemical and bioactive potential was corroborated (Bustamante *et al.*, 2021). According to Pennington (1990), the identification of genera of Sapotaceae is challenging and depends on a combination of characters, since no character is diagnostic. Quite different treatments were proposed depending on the character chosen (compare Aubreville (1964) and Baehni (1965) as an example). This

situation, added to the absence of a comprehensive treatment for the family in the Neotropics since *Flora Brasiliensis* (Miquel, 1863), resulted in considerable confusion with ill-defined generic limits and genera with numerous synonyms (Pennington, 1990). Thus, the development of tools that allow us to address phylogenetic relationships and clarify generic limits is extremely important for Sapotaceae. To our knowledge, there are published plastomes from only six genera in the family (Jo *et al.*, 2016, Khayi *et al.*, 2020, Liu *et al.*, 2019, Niu *et al.*, 2018, 2020, Tao *et al.*, 2021, Wang *et al.*, 2021). Therefore, we present the first complete plastome for a species of the genus *Mimusops*.

Leaves from the collection N. Hanazaki n. 33759 deposited in the herbarium at the University of Campinas (UEC) under the voucher number 90949 were used for DNA extraction. The leaves were grinded for 60 seconds in 2 mL tubes using TissueLyserII (Qiagen, Hilden, Germany - Cat. No. 85300) and the total DNA was extracted using a modified version of the Doyle and Doyle (1987) CTAB protocol. The DNA was purified using the Wizard® SV Gel and PCR Clean-Up System (Promega, Madison, USA - Cat. No. A9282). Whole-genome libraries were prepared by Genohub Inc. facility (Austin, USA) and paired-end reads (2 x 150 bp) were sequenced on Illumina NextSeq 500 platform (Illumina Inc., San Diego, USA), resulting in a total of 10,687,184 reads. The whole chloroplast was *de novo* assembled using NOVOPlasty v. 3.8.2 (Dierckxsens *et al.*, 2017) using the *rbcL*, accession number L01932.2, as seed. Reads were mapped to the chloroplast with Bowtie2 (Langmead and Salzberg, 2012) plugin in Geneious 9 (Kearse *et al.*, 2012) using default parameters to validate the assembly and verify the coverage. A total of 503,543 reads mapped the assembled plastome with mean coverage of 412.8.

Automatic annotation was done using GeSeq (Tillich *et al.*, 2017) implemented on the Chlorobox website. Start and stop codons were visually inspected and manually adjusted on Geneious 9 (Kearse *et al.*, 2012), and tRNAs limits were corrected based on ARAGORN (Laslett and Canback, 2004) output. Potential pseudogenes were defined by Blast following Silva *et al.* (2018). The annotated chloroplast was submitted to GenBank under the accession number MW846242. A circular map was generated with OGDRAW (Lohse *et al.*, 2013). Five plastomes from Sapotaceae had the quadripartite structure annotated, i.e., the large single copy (LSC), the small single copy (SSC), and the two inverted repeats (IR): *Madhuca hainanensis* Chun & F.C. How, *Manilkara zapota* (L.) P. Royen, *Pouteria caimito* (Ruiz & Pav.) Radlk., *Sideroxylon wightianum* Wall. and *Synsepalum dulcificum* (Schumacher & Thonn.) Daniell. These five plastomes were aligned with *Mimusops coriacea* using progressive Mauve algorithm in Mauve Plugin v. 2.3.2 (Darling *et al.*, 2004) in Geneious 9 (Kearse *et al.*, 2012) to check for structural differences such as inversions or rearrangements. The limits between the four main regions of these six plastomes were evaluated with IRscope (Amiryousefi *et al.*, 2018). Forward, palindromic, complementary, and reverse repeats were identified in REPuter (Kurtz *et al.*, 2001); parameters were set as follows: minimal size of 30 bp and Hamming distance of 3. Simple sequence repeats (SSRs) were identified in MISA (Beier *et al.*, 2017; Thiel *et al.*, 2003) with a minimum number of 7, 4, 4, 3, 3, and 3, for mono-, di-, tri-, tetra-, penta-, and hexanucleotide repeats, respectively. To determine the position of *Mimusops coriacea* within Sapotaceae we selected 16 plastomes from Ericales on the GenBank database: three from Ebenaceae, two from Primulaceae, eight from Sapotaceae, two from Styracaceae, and one from Theaceae. All species information and GenBank accession numbers are presented in Table S1. A total of 80 protein-coding and four rRNA genes were aligned with MAFFT v 7.435 (Kato and Standley, 2013). A maximum likelihood tree based on the concatenation approach considering different evolutionary models for each locus was generated with IQ-TREE 2 (Minh *et al.*, 2020).

The general structure and the gene order of the plastome of *M. coriacea* are similar to those of the other five genera

of Sapotaceae (*Madhuca hainanensis*, *Manilkara zapota*, *Pouteria caimito*, *Sideroxylon wightianum* and *Synsepalum dulcificum*) (Figure 1). The complete plastome has a total length of 159,689 bp, including a LSC of 88,887 bp, a SSC of 18,618 bp, and two IRs each of 26,092 bp. The GC content was 36.8% for the complete chloroplast, 34.6% for the LSC, 30.2% for the SSC and 42.9% for the IRs. The whole plastome and its four main regions have both similar sizes and respective GC content among all analyzed Sapotaceae species (Table 1). *Mimusops coriacea* plastome contains a total of 133 genes of which 81 are unique protein-coding genes, 30 are tRNA and four are rRNA genes (Figure 2). Seven protein-coding, seven tRNA and four rRNA genes are in the IR duplicated region. No rearrangements were observed in the chloroplast of the analyzed Sapotaceae species as occur in some other families from Ericales. However, rearrangements have been reported in the order, i.e., the *clpP* loss in Actinidiaceae (Wang *et al.*, 2016), the highly reduced and rearranged plastome of non-photosynthetic Ericaceae (Logacheva *et al.*, 2016; Braukmann *et al.*, 2017), a 20 kb inversion in two genera of Styracaceae (Cai *et al.*, 2021), and the pseudogenization of the *pt-accD* and transfer to the nucleus in *Primula sinensis* Sabine ex Lindle (Primulaceae) (Liu *et al.*, 2016).

All the analyzed Sapotaceae plastomes have the junctions between single-copy and repeated regions in the same location as those of a typical angiosperm plastome (Wang *et al.*, 2018). The LSC-IRb limit is flanked by *rpl22* and *rps19* on the LSC side and by *rpl2* on the IRb side. The IRb-SSC junction has the *ycf1*-fragment on IRb in *Madhuca hainanensis* and *Mimusops coriacea* and the *ndhF* on SSC; only in *Synsepalum dulcificum* the *ndhF* spans the junction. We believe the absence of the *ycf1*-fragment in other species is just an annotation issue. The *ycf1* spans the SSC-IRa junction and the IRa-LSC junction is flanked by *rpl2* on IRa side and by *trnH* on LSC in all species but *Synsepalum dulcificum* which has a tRNA instead of the *trnH* (Figure S1). In *Mimusops coriacea* chloroplast, 16 genes have one intron each and two genes (*ycf3* and *clpP*) have two introns each. REPuter (Kurtz *et al.*, 2001) identified 49 long repeats in *Mimusops coriacea*, being 15 palindromic, 18 forward, eight reverse and eight complementary. For the other five Sapotaceae species included, between 30 repeats

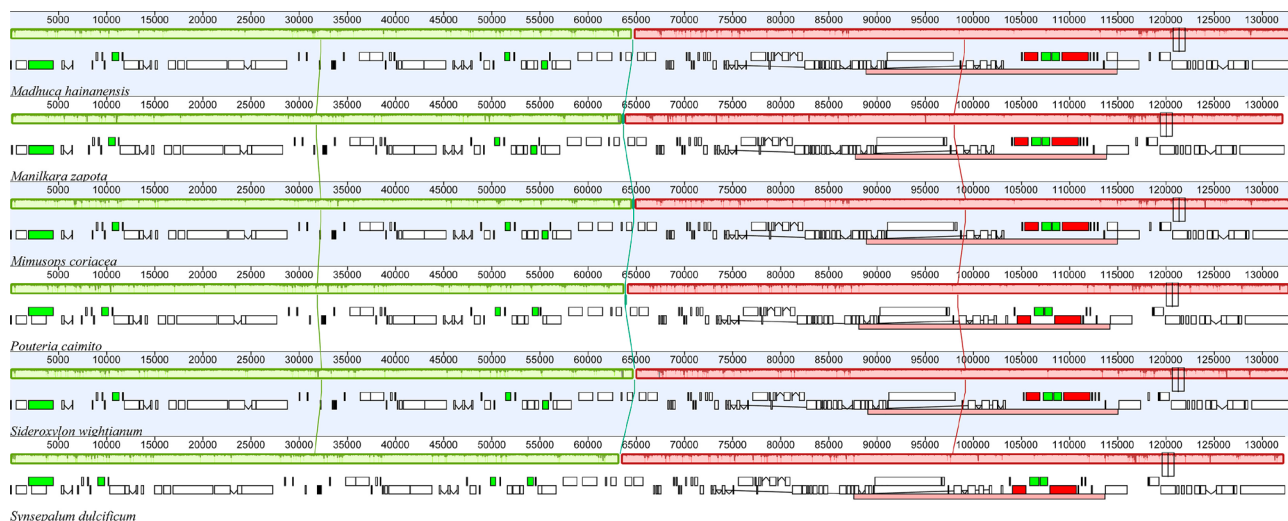


Figure 1 – Progressive Mauve alignment showing synteny and rearrangements for six Sapotaceae plastomes. Species names are indicated below each respective plastome.

Table 1 – Comparison of chloroplast genome size and GC content across three different regions (LSC, SSC, and IR) for six Sapotaceae species. LSC - large single copy; SSC - small single copy; IR - inverted repeat.

Species	LSC		SSC		IR		Complete plastome	
	bp	% GC	bp	% GC	bp	% GC	bp	% GC
<i>Madhuca hainanensis</i> Chun & F.C. How	88,846	34.6	18,598	30.1	26,093	42.9	159,630	36.8
<i>Manilkara zapota</i> (L.) P.Royen	87,745	34.9	18,443	30.2	26,099	42.9	158,386	37.0
<i>Mimusops coriacea</i> (A. DC.) Miq.	88,887	34.6	18,618	30.2	26,092	42.9	159,689	36.8
<i>Pouteria caimito</i> Radlk.	88,096	34.6	18,620	30.2	26,100	42.9	158,916	36.8
<i>Sideroxylon wightianum</i> Hook. & Arn.	89,043	34.7	18,365	30.4	26,065	42.9	159,538	36.9
<i>Synsepalum dulcificum</i> (Schumach. & Thonn.) Daniell	87,574	34.7	18,635	30.1	26,127	42.9	158,463	36.9

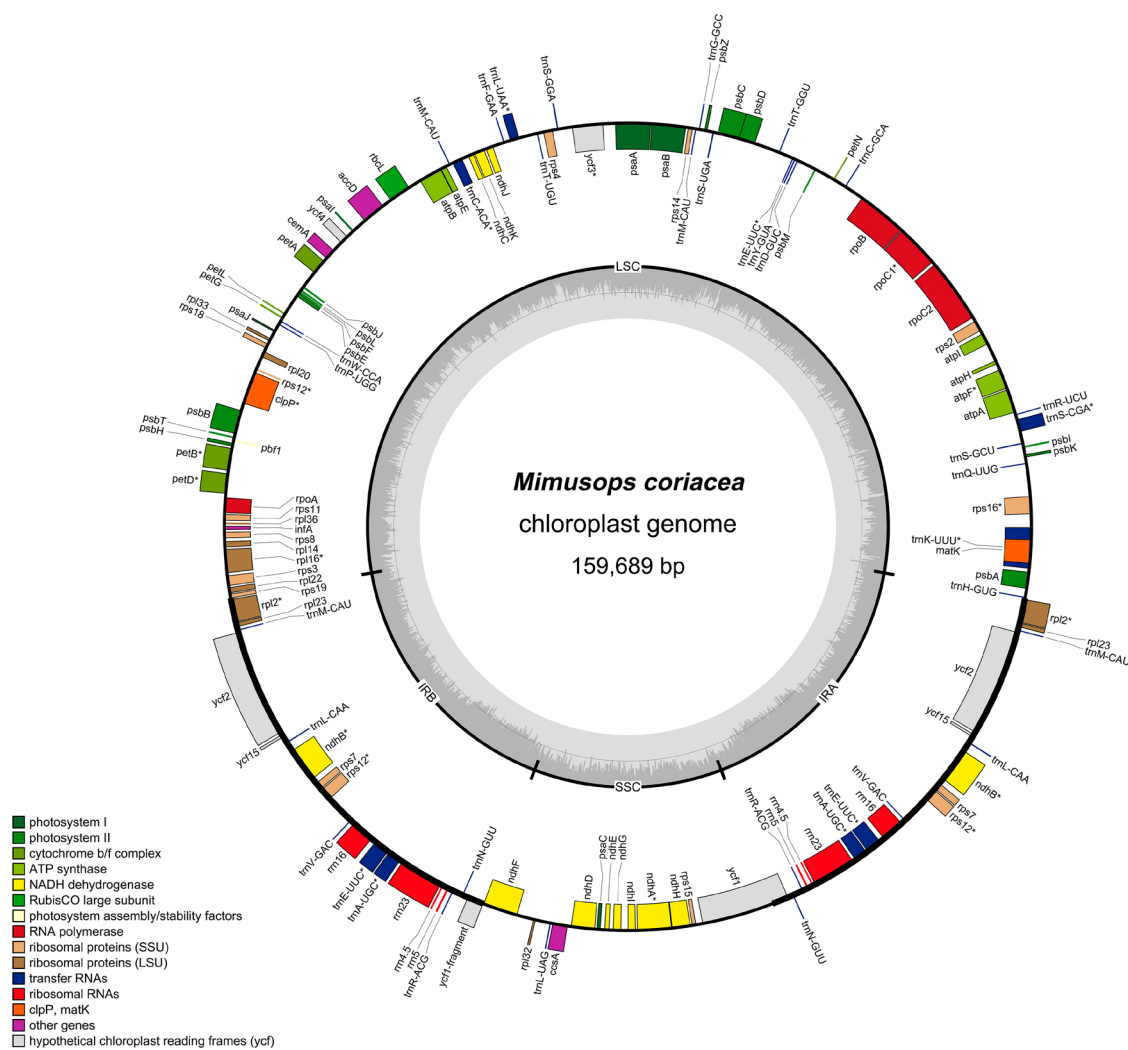


Figure 2 – Circular map of *Mimusops coriacea* complete chloroplast genome. The genes represented outside the circle are transcribed counterclockwise and those inside the outer circle are transcribed clockwise. Genes are colored according to their functional groups following the legend. The inner gray graphs indicate the GC content across the plastome.

in *Synsepalum dulcificum* and 50 repeats in both *Madhuca hainanensis* and *Sideroxylon wightianum* were found (Table S2). MISA v. 2.1 (Beier *et al.*, 2017; Thiel *et al.*, 2003) identified 344 SSRs throughout *Mimusops coriacea* plastome. Among these repeats 287 are mono-SSR and 47 are di-SSR, four are tri-SSR, five are tetra-SSR and one is penta-SSR. Similar number of repeats were observed in the other five Sapotaceae species included (Table S2).

Relationships within Ericales are not fully understood (Stevens, 2001, onwards). In our maximum likelihood tree,

Sapotaceae was recovered as monophyletic and sister to the clade Ebenaceae + Primulaceae, which agrees with Rose *et al.* (2018). On the other hand, Larson *et al.* (2020) recovered Sapotaceae as sister to Ebenaceae, and Primulaceae not close to that clade. In Sapotaceae, *Mimusops coriacea* clustered with the other eight species from the family (Figure 3). Our tree also showed a paraphyletic *Pouteria* Aubl. The complete chloroplast genome of *Mimusops coriacea* may contribute to future evolutionary studies within Sapotaceae, including a better comprehension of generic limits and interspecific relationships.

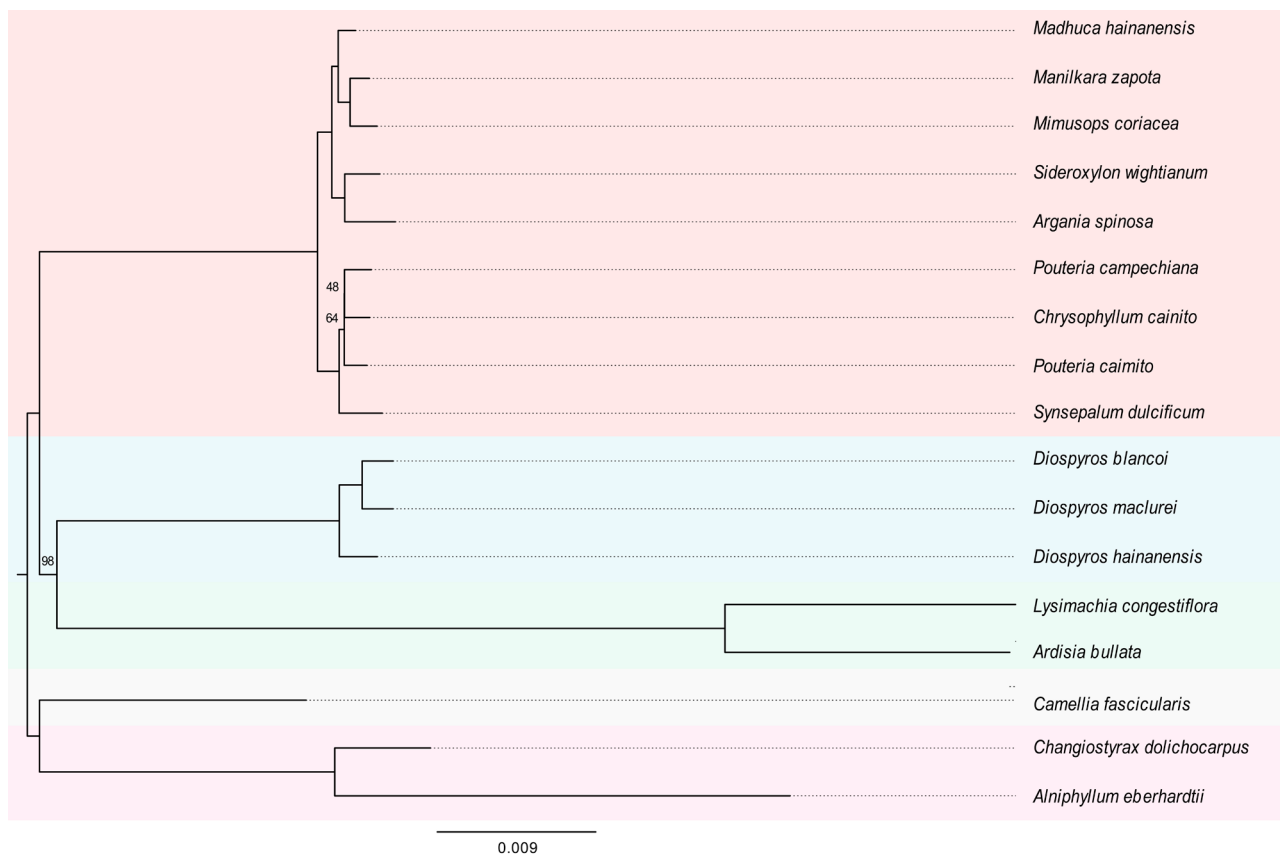


Figure 3 – Phylogenetic tree of 16 Ericales species based on 80 chloroplast protein-coding genes and four rRNAs generated by maximum likelihood method. Numbers above branches represent ultrafast bootstrap support values; nodes with support lower than 100% are indicated. Theaceae and Styracaceae were used as outgroups. The background was colored as follows: Sapotaceae – red, Ebenaceae – blue, Primulaceae – green, Theaceae – gray, Styracaceae – pink.

Acknowledgements

We would like to thank Profa. Dra. Anete P. Souza for allowing the use of CBMEG – Unicamp cluster, and Profa. Dra. Aparecida Donisete de Faria and Prof. Dr. Anderson Alves-Araújo who identified the specimen. This work was supported by Fundação de Apoio à Pesquisa do Estado de São Paulo (FAPESP) under Research Grant 2012/51781-0 to M.C.E.A. and a Postdoctoral Fellowship 2018/02285-6 to S.R.S.; Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq) under PROTAX Grant 562299/2010-6, Universal Research Grant 482988/2013-4 and a Research Grant 311744/2017-6 to M.C.E.A., and the scholarships 142484/2014-8 and 205856/2014-4 to R.J.T.; FAEPEX/Unicamp Grant 1124/13 to M.C.E.A.; and Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES – finance code 001) Doctoral Fellowship to R.J.T.

Conflict of Interest

The authors declare no conflict of interest.

Author Contributions

RJT designed the study and collected the data. RJT and SRS conceived and conducted the analyses. RJT, SRS and MCEA discussed the results. RJT wrote the first version of the manuscript with input from all authors. All authors read and approved the final version.

References

- Alves-Araújo A, Moraes QS, Miranda VS, Nichio-Amaral R and Ribeiro M (2020) Novelities in Sapotaceae from the Espírito Santo state, Brazil: Updated checklist and new records with emended descriptions for two species. *Feddes Repert* 131:251-259.
- Amiryousefi A, Hyvönen J and Poczai P (2018) IRscope: An online program to visualize the junction sites of chloroplast genomes. *Bioinformatics* 34:3030-3031.
- Aubreville A (1964) Les Sapotacées: Taxonomie et phytogéographie. *Adansonia* 1:1-157.
- Baehni C (1965) Mémoires sur les Sapotacées. 3. Inventaire des genres. *Boissiera* 11:1-262.
- Beier S, Thiel T, Münch T, Scholz U and Mascher M (2017) MISA-web: A web server for microsatellite prediction. *Bioinformatics* 33:2583-2585.
- Braukmann TWA, Broe MB, Stefanovic S and Freudenstein JV (2017) On the brink: The highly reduced plastomes of nonphotosynthetic Ericaceae. *New Phytol* 216:254-266.
- Bustamante K, Santos-Ordóñez E, Miranda M, Pacheco R, Gutiérrez Y and Scull R (2019) Morphological and molecular barcode analysis of the medicinal tree *Mimusops coriacea* (A.DC.) Miq. collected in Ecuador. *PeerJ* 7:e7789.
- Bustamante Pesantes KE, Gutiérrez Gaitén YI, Chóez Guaranda IA and Miranda Martínez M (2021) Composición química, capacidad antioxidante y actividad antiinflamatoria de los frutos de *Mimusops coriacea* (A.DC) Mig (Sapotaceae) que crece en Ecuador. *J Pharm Pharmacogn Res* 9:33-48.
- Cai X-L, Landis JB, Wang H-X, Wang J-H, Zhu Z-X and Wang H-F (2021) Plastome structure and phylogenetic relationships of Styracaceae (Ericales). *BMC Ecol Evol* 21:103.

- Darling ACE, Mau B, Blattner FR and Perna NT (2004) Mauve: Multiple alignment of conserved genomic sequence with rearrangements. *Genome Res* 14:1394-1403.
- Dierckxsens N, Mardulyn P and Smits G (2017) NOVOPlasty: De novo assembly of organelle genomes from whole genome data. *Nucleic Acids Res* 45:e18.
- Doyle JJ and Doyle JL (1987) A rapid DNA isolation procedure from small quantities of fresh leaf tissue. *Phytochemistry* 19:11-15.
- Gautier L, Nusbaumer L, Garratt R, Randrianaivo R and Phillipson PB (2012) *Mimusops coriacea* (A. DC.) Miq. (Sapotaceae): Nomenclature, distribution and ecology. *Candollea* 67:148-151.
- Jo S, Kim HW, Kim YK, Cheon SH and Kim KJ (2016) The first complete plastome sequence from the family Sapotaceae, *Pouteria campechiana* (Kunth) Baehni. *Mitochondrial DNA B Resour* 1:734-736.
- Katoh K and Standley DM (2013) MAFFT multiple sequence alignment software version 7: Improvements in performance and usability. *Mol Biol Evol* 30:772-780.
- Kearse M, Moir R, Wilson A, Stones-Havas S, Cheung M, Sturrock S, Buxton S, Cooper A, Markowitz S, Duran C *et al.* (2012) Geneious Basic: An integrated and extendable desktop software platform for the organization and analysis of sequence data. *Bioinformatics* 28:1647-1649.
- Khayri S, Gaboun F, Pirro S, Tatusova T, Mousadik AE, Ghazal H and Mentag R (2020) Complete chloroplast genome of *Argania spinosa*: Structural organization and phylogenetic relationships in Sapotaceae. *Plants (Basel)* 9:1354.
- Kurtz S, Choudhuri JV, Ohlebusch E, Schleiermacher C, Stoye J and Giegerich R (2001) REPuter: The manifold applications of repeat analysis on a genomic scale. *Nucleic Acids Res* 29:4633-4642.
- Langmead B and Salzberg S (2012) Fast gapped-read alignment with Bowtie 2. *Nat Methods* 9:357-359.
- Larson DA, Walker JF, Vargas OM and Smith SA (2020) A consensus phylogenomic approach highlights paleopolyploid and rapid radiation in the history of Ericales. *Am J Bot* 107:1-17.
- Laslett D and Canback B (2004) ARAGORN, a program to detect tRNA genes and tmRNA genes in nucleotide sequences. *Nucleic Acids Res* 32:11-16.
- Liu J, Ren SN, Li K-X, Xu T, Niu Y-F and Shi C (2019) The complete chloroplast genome sequence of *Manilkara zapota* (Linn.) van Royen. *Mitochondrial DNA B Resour* 4:2127-2128.
- Liu T-J, Zhang C-Y, Yan H-F, Zhang L, Ge X-J and Hao G (2016) Complete plastid genome sequence of *Primula sinensis* (Primulaceae): Structure comparison, sequence variation and evidence for *accD* transfer to nucleus. *PeerJ* 4: e2101.
- Logacheva MD, Schelkunov MI, Shtratnikova VY, Matveeva MV and Penin AA (2016) Comparative analysis of plastid genomes of non-photosynthetic Ericaceae and their photosynthetic relatives. *Sci Rep* 6:30042.
- Lohse M, Drechsel O, Kahlau S and 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:W575-W581.
- Minh BQ, Schmidt HA, Chernomor O, Schrempf D, Woodhams MD, von Haeseler A and Lanfear R (2020) IQ-TREE 2: New models and efficient methods for phylogenetic inference in the genomic era. *Mol Biol Evol* 37:1530-1534.
- Miquel FAW (1863) Sapotae. In: von Martius CFP (ed) *Flora Brasiliensis*. Monachii & Lipsiae, F. Fleischer, vol. 7B, pp 37-118.
- Niu Y-F, Ni S-B, Liu Z-Y, Zheng C, Mao C-L, Shi C and Liu J (2018) The complete chloroplast genome of tropical and sub-tropical fruit tree *Lucuma nervosa* (Sapotaceae). *Mitochondrial DNA B Resour* 3:440-441.
- Niu Y-F., Ni S.-B. and Liu J (2020) Complete chloroplast genome of *Synsepalum dulcificum* D.: A magical plant that modifies sour flavors to sweet. *Mitochondrial DNA B Resour* 5:3052-3053.
- Pennington TD (1990) Sapotaceae. *Flora Neotropica Monograph* 52:1-770.
- Rose JP, Kleistb TJ, Löfstrand SD, Drewd BT, Schönenbergere J and Sytsma KJ (2018) Phylogeny, historical biogeography, and diversification of angiosperm order Ericales suggest ancient Neotropical and East Asian connections. *Mol Phylogenet Evol* 122:59-79.
- Silva SR, Michael TP, Meer EJ, Pinheiro DG, Varani AM and Miranda VFO (2018) Comparative genomic analysis of *Genlisea* (corkscrew plants – Lentibulariaceae) chloroplast genomes reveals an increasing loss of the *ndh* genes. *PLoS One* 13:e0190321.
- Tao L, Shi ZG and Long QY (2021) Complete chloroplast genome sequence and phylogenetic analysis of a wild species of abiu fruit, *Pouteria caimito* (Ruiz & Pavon.) Radlk. *Mitochondrial DNA B Resour* 6:138-139.
- Thiel T, Michalek W, Varshney R and Graner A (2003) Exploiting EST databases for the development and characterization of gene-derived SSR-markers in barley (*Hordeum vulgare* L.). *Theor Appl Genet* 106:411-422.
- Tillich M, Lehwark P, Pellizzer T, Ulbricht-Jones ES, Fischer A, Bock R and Greiner S (2017) GeSeq - versatile and accurate annotation of organelle genomes. *Nucleic Acids Res* 45:W6-W11.
- Wang W-C, Chen S-Y and Zhang X-Z (2016) Chloroplast genome evolution in Actinidiaceae: *clpP* loss, heterogeneous divergence and phylogenomic practice. *PLoS One* 11:e0162324.
- Wang Y-H, Wicke S, Wang H, Jin J-J, Chen S-Y, Zhang S-D, Li D-Z and Yi T-S (2018) Plastid genome evolution in the early-diverging legume subfamily Cercidoideae (Fabaceae). *Front Plant Sci* (9):138.
- Wang Y, Wang H-T, Chen YK, Yu J and Yang Y (2021) The complete chloroplast genome of *Madhuca hainanensis* (Sapotaceae), an endemic and endangered timber species in Hainan Island, China. *Mitochondrial DNA B Resour* 6:755-757.

Internet Resources

- Alves-Araújo A (2020) *Mimusops* in Flora do Brasil 2020. Jardim Botânico do Rio de Janeiro, <http://floradobrasil.jbrj.gov.br/reflora/floradobrasil/FB75884> (accessed 26 April 2021).
- Chlorobox, <https://chlorobox.mpimp-golm.mpg.de/> (accessed 15 April 2021).
- Stevens PF (2001) Angiosperm Phylogeny Website. Version 14 [and more or less continuously updated since]. <http://www.mobot.org/MOBOT/research/APweb/>.

Supplementary material

- The following online material is available for this article:
- Table S1 – Species names, family, and GenBank accession number for the sequences included in this study.
- Table S2 – Number of simple sequence repeats (SSRs) and long repeats present in six Sapotaceae species.
- Figure S1 – Comparison of the IR and single-copy regions junctions from six Sapotaceae plastomes.

Associate Editor: Rogério Margis

License information: This is an open-access article distributed under the terms of the Creative Commons Attribution License (type CC-BY), which permits unrestricted use, distribution and reproduction in any medium, provided the original article is properly cited.