

The complete chloroplast genome dataset of *Acca sellowiana* and its basic characterization

Bohan Yang

Shanghai Jiao Tong University

Yaru Luo

Shanghai Jiao Tong University

Xiaohui Wang

Yuansheng Jinhua Tea Development Co., Ltd

Xiaojuan Xie

Yuansheng Jinhua Tea Development Co., Ltd

Shiren Song

sr.song@sjtu.edu.cn

Shanghai Jiao Tong University

Data Note

Keywords: *Acca sellowiana*, plastome, chloroplast genome, genomic resource, Myrtaceae

Posted Date: May 8th, 2026

DOI: <https://doi.org/10.21203/rs.3.rs-9493341/v1>

License: © ⓘ This work is licensed under a Creative Commons Attribution 4.0 International License. [Read Full License](#)

Additional Declarations: No competing interests reported.

Version of Record: A version of this preprint was published at BMC Genomic Data on September 17th, 2026. See the published version at <https://doi.org/10.1186/s12863-026-01473-z>.

Abstract

Objective: *Acca sellowiana* (feijoa) is an economically important fruit tree species in the family Myrtaceae. However, genomic resources, particularly complete chloroplast genome datasets, remain limited for this species. The generation of plastome data is essential for phylogenetic analysis, molecular marker development, and germplasm research.

Data description: In this study, we generated and characterized the complete chloroplast genome of *Acca. sellowiana* using Illumina sequencing. The plastome is 159,374 bp in length and exhibits a typical quadripartite structure, including a large single-copy (LSC) region (87,932 bp), a small single-copy (SSC) region (18,598 bp), and a pair of inverted repeats (IRs) (26,422 bp each). A total of 129 unique genes were annotated, including 84 protein-coding genes, 37 tRNA genes, and 8 rRNA genes. Phylogenetic analysis based on complete chloroplast genomes confirmed that *Acca sellowiana* is nested within the Myrteae clade. This dataset provides a valuable genomic resource for plastome evolution, phylogenetic inference, and molecular breeding in Myrtaceae.

Objective

Acca sellowiana (feijoa) is an economically important fruit tree species in the family Myrtaceae, valued for its nutritional composition and increasing cultivation. Despite the growing interest in its breeding and utilization, genomic resources for this species, particularly complete chloroplast genome datasets, remain limited. Chloroplast genomes are widely used for phylogenetic analysis, molecular marker development, and studies of genome structure evolution in plants. However, plastome data for the genus *Acca* are still insufficient to support comprehensive comparative and evolutionary analyses. Therefore, the aim of this study was to generate and provide a complete chloroplast genome dataset of *Acca. sellowiana*, together with its basic structural and genomic characterization, as a publicly available resource to support future research in phylogenetics, germplasm characterization, and molecular breeding within Myrtaceae.

Data description

Sample collection and sequencing Young leaves of *Acca. sellowiana* were collected from Sanming City, Fujian Province, China, and stored at -80°C prior to DNA extraction. Genomic DNA was extracted using a CTAB-based method. A paired-end library (~350 bp insert size) was constructed and sequenced on the Illumina NovaSeq 6000 platform, generating 150 bp paired-end reads (Table 1, Data file 2).

Genome assembly and annotation: After quality filtering, the chloroplast genome was assembled using GetOrganelle and validated through read mapping. Gene annotation was performed using GeSeq and manually curated. The circular genome map was visualized using OGDRAW (Figure 1). The complete chloroplast genome of *Acca. sellowiana* is 159,374 bp in length, with a typical quadripartite structure consisting of LSC (87,932 bp), SSC (18,598 bp), and two IR regions (26,422 bp each). The overall GC content is 37.0%. A total of 129 unique genes were identified, including 84 protein-coding genes, 37 tRNA genes, and 8 rRNA genes.

Phylogenetic analysis: To validate the assembled genome, phylogenetic relationships were inferred based on complete chloroplast genomes of representative Myrtaceae species using the maximum-likelihood method. *A. sellowiana* clustered within the Myrteae clade, consistent with previous studies (Figure 2, Data file 4).

Data validation: The datasets generated and analyzed in this study, including the complete chloroplast genome sequence, raw sequencing reads, genome annotation files, and phylogenetic tree data, are publicly available as described in Table 1.

Table 1 Overview of data files/data sets

Data file	Name of data file/data set	File types	Data repository and identifier
Data file 1	Complete chloroplast genome sequence	.fasta	NCBI GenBank (https://www.ncbi.nlm.nih.gov/nuccore/PX562857.1) >https://www.ncbi.nlm.nih.gov/nuccore/PX562857.1[1]
Data file 2	Raw sequencing reads	.fastq.gz	NCBI SRA (https://www.ncbi.nlm.nih.gov/sra/?term=PRJNA1359583) >https://www.ncbi.nlm.nih.gov/sra/?term=PRJNA1359583[2]
Data file 3	Genome annotation files	.gff3	Figshare (https://doi.org/10.6084/m9.figshare.32073273) >https://doi.org/10.6084/m9.figshare.32073273[3]
Data file 4	Phylogenetic tree data	.treefile	Figshare (https://doi.org/10.6084/m9.figshare.32073378) >https://doi.org/10.6084/m9.figshare.32073378[4]

Limitations

This dataset has several limitations. First, the chloroplast genome was derived from a single individual, which may not capture intraspecific variation. Second, only plastome data were generated, and no nuclear genome or transcriptomic data were included. Third, functional validation and comparative analyses across broader populations were not performed. Future studies incorporating multi-omics data and population-level sampling would provide deeper insights.

Abbreviations

CTAB: cetyltrimethylammonium bromide

IR: inverted repeat

LSC: large single-copy

ML: maximum-likelihood

NGS: next-generation sequencing

SSC: small single-copy

GetOrganelle: organelle genome assembly pipeline

GeSeq: organelle genome annotation tool

OGDRAW: Organelle Genome DRAW tool for genome map visualization

Declarations

Acknowledgments

This work was supported by Sanming Yuansheng Jinhua Tea Development Co., Ltd., which provided the feijoa materials for this study. The chloroplast genome sequencing, assembly, and annotation were carried out with support from Shanghai Personalbio Biotechnology Co., Ltd.

Author contributions

BY: Conceptualisation, Methodology, Software, Validation, Resources, Writing–Original Draft. YL: Data Curation, Formal Analysis, Writing–Review & Editing. XW: Investigation, Data Curation. XX: investigation, Data Curation. SS: Supervision, Writing - Review & Editing, Project Administration.

Funding

This study was supported by Shanghai Agricultural Science and Technology Innovation Program (Grant No.K2023017, Grant No.T2023208) and Sanming Sci-tech Commissioner Special Project (No.2023-N-11).

Availability of data and materials

The genome sequence data that support the findings of this study are openly available in the GenBank of NCBI (<https://www.ncbi.nlm.nih.gov/>) under the accession number PX562857.1. The raw sequencing reads used for chloroplast genome assembly have been deposited in the NCBI Sequence Read Archive (SRA) under the BioProject accession PRJNA1359583.

Ethics approval and consent to participate

Not applicable.

Consent to publish

Not applicable.

Competing interests

The authors declare no competing interests.

Publisher's note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

References

1. Feijoa sellowiana isolate SJTU-AS-001 cultivar andre chloroplast, complete genome. 2026. <http://www.ncbi.nlm.nih.gov/nuccore/PX562857.1>
2. Raw reads for chloroplast genome assembly: Feijoa young leaf - SRA - NCBI. 2026. <https://www.ncbi.nlm.nih.gov/sra/?term=PRJNA1359583>
3. Yang B, Luo Y. Genome annotation of feijoa's chloroplast. Figshare. 2026. <https://doi.org/10.6084/m9.figshare.32073273>
4. Yang B, Luo Y. Phylogenetic relationships of feijoa. Figshare. 2026. <https://doi.org/10.6084/m9.figshare.32073378>
5. Zhang XF, Landis JB, Wang HX, Zhu ZX, Wang HF. Comparative analysis of chloroplast genome structure and molecular dating in myrtales. *BMC Plant Biol.* 2021;21(1):219. <https://doi.org/s12870-021-02985-9>
6. Reginato M, Neubig KM, Majure LC, Michelangeli FA. The first complete plastid genomes of melastomataceae are highly structurally conserved. *PeerJ.* 2016;4:e2715. <https://doi.org/10.7717/peerj.2715>
7. Jo S, Kim HW, Kim YK, Cheon SH, Kim KJ. Complete plastome sequence of psidium guajava L. (myrtaceae). *Mitochondrial DNA B Resour.* 1(1):612-614. <https://doi.org/10.1080/23802359.2016.1209096>
8. Rodrigues NF, Balbinott N, Paim I, Guzman F, Margis R. Comparative analysis of the complete chloroplast genomes from six neotropical species of myrteae (myrtaceae). *Genet Mol Biol.* 2020;43(2):e20190302. <https://doi.org/10.1590/1678-4685-GMB-2019-0302>
9. Eguiluz M, Yuyama PM, Guzman F, Rodrigues NF, Margis R. Complete sequence and comparative analysis of the chloroplast genome of plinia trunciflora. *Genet Mol Biol.* 2017;40(4):871-876. <https://doi.org/10.1590/1678-4685-GMB-2017-0096>
10. Machado L de O, Stefenon VM, Vieira L do N, Nodari RO. Structural and evolutive features of the plinia phitrantha and P. cauliflora plastid genomes and evolutionary relationships within tribe myrteae (myrtaceae). *Genet Mol Biol.* 45(1):e20210193. <https://doi.org/10.1590/1678-4685-GMB-2021-0193>
11. Smissen RD, Heenan PB, Maurin KJL. New zealand endemic neomyrtus is sister to new caledonian endemic myrtastrum (myrtaceae, myrteae). Accessed April 15, 2026. <https://rsnz.onlinelibrary.wiley.com/doi/10.1080/0028825X.2021.1965629>
12. Gu C, Tembrock LR, Johnson NG, Simmons MP, Wu Z. The complete plastid genome of lagerstroemia fauriei and loss of rpl2 intron from lagerstroemia (lythraceae). *PLoS One.* 2016;11(3):e0150752. <https://doi.org/10.1371/journal.pone.0150752>
13. Chakraborty A, Mahajan S, Bisht MS, Sharma VK. Genome sequencing of Syzygium cumini (jamun) reveals adaptive evolution in secondary metabolism pathways associated with its medicinal properties. *Front Plant Sci.* 2023;14:1260414. <https://doi.org/10.3389/fpls.2023.1260414>
14. Machado L de O, Vieira L do N, Stefenon VM, et al. Molecular relationships of campomanesia xanthocarpa within myrtaceae based on the complete plastome sequence and on the plastid ycf2 gene. *Genet Mol Biol.* 43(2):e20180377. <https://doi.org/10.1590/1678-4685-GMB-2018-0377>
15. Liang D, Zhu L, He Y, Xiong X. Characterization of the complete chloroplast genome of melaleuca cajuputi subsp. cumingiana (myrtaceae). *Mitochondrial DNA B Resour.* 6(2):462-464. <https://doi.org/10.1080/23802359.2020.1871438>
16. Jo S, Kim HW, Kim YK, Cheon SH, Kim KJ. The complete plastome sequence of diospyros blancoi a. DC. (ebenaceae). *Mitochondrial DNA B Resour.* 2016;1(1):690-692. <https://doi.org/10.1080/23802359.2016.1219647>

Figures

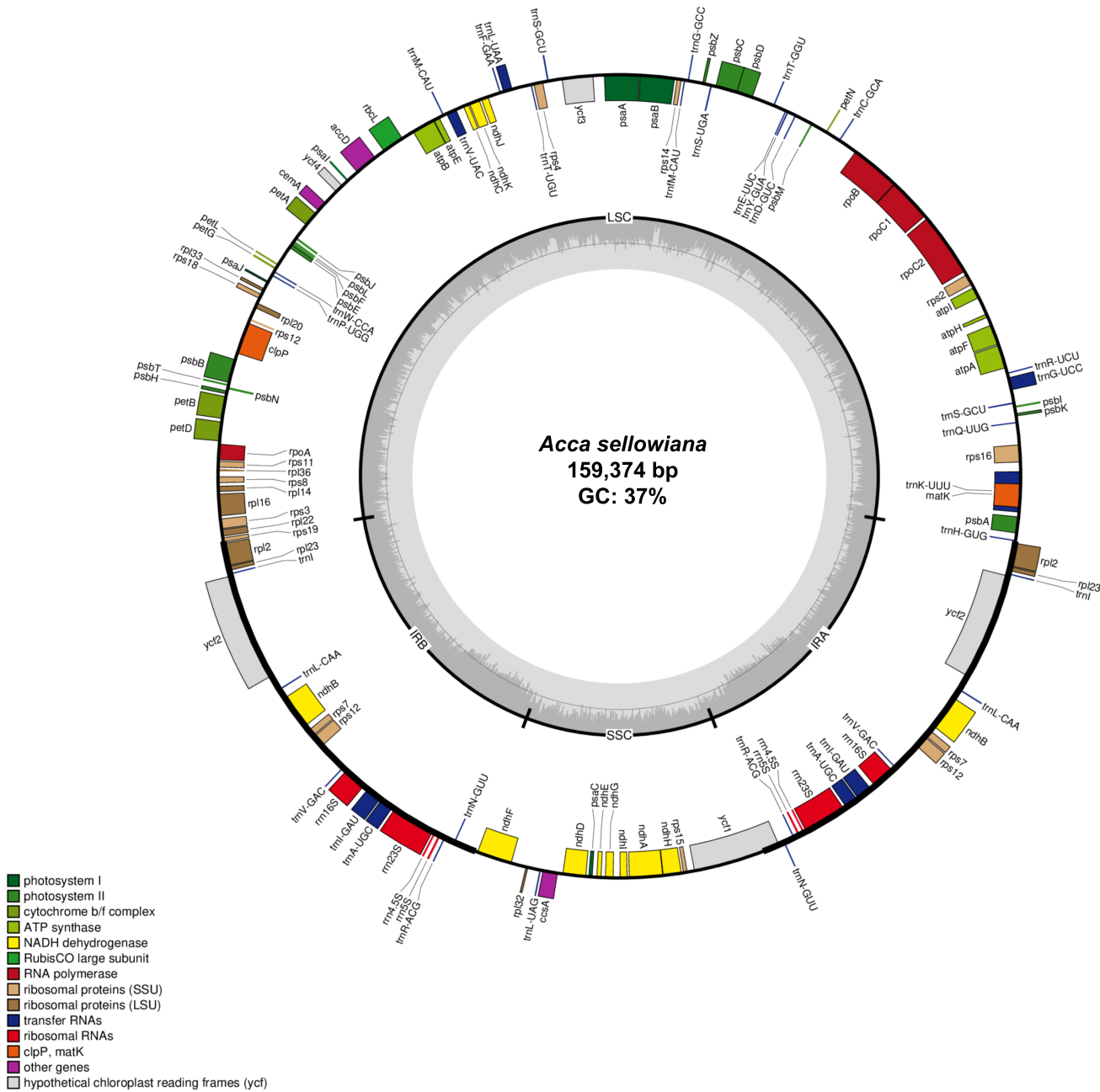


Figure 1

Circular map of the *Acca sellowiana* chloroplast genome. Genes shown on the outside and inside of the circle indicate transcription in clockwise and counterclockwise directions, respectively. The inner rings represent GC content variation and the boundaries of the LSC, SSC, and IR regions.

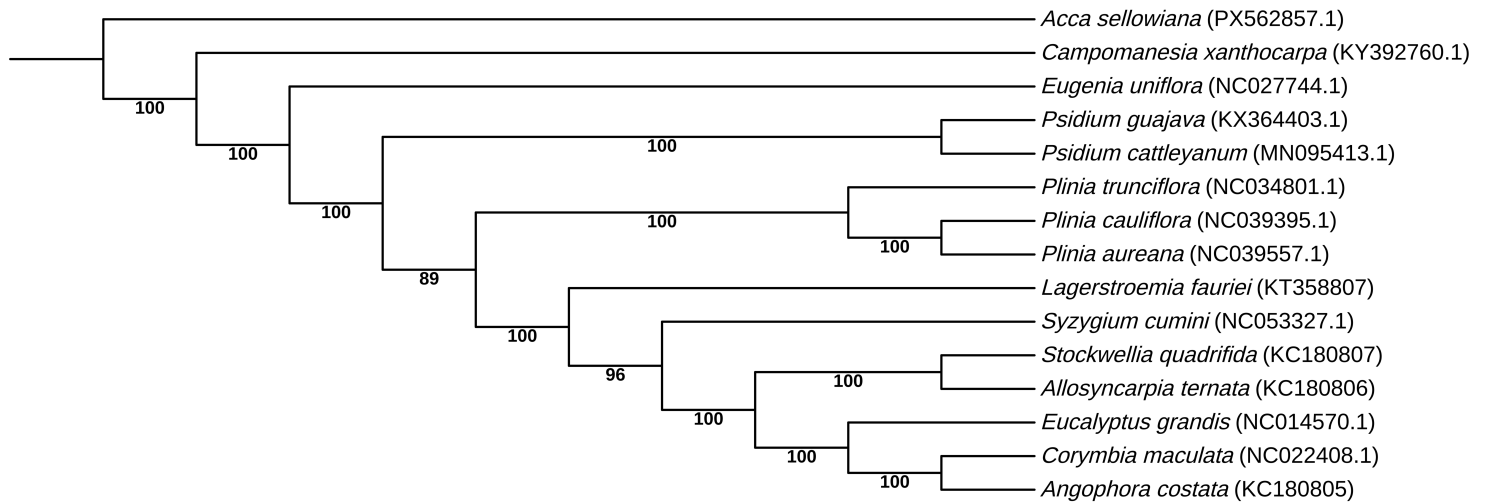


Figure 2

Phylogenetic relationships of *Acca. sellowiana* within Myrtaceae inferred from complete chloroplast genome sequences, with the outgroup species indicated in the tree. Numbers at the nodes represent support values for the corresponding clades (maximum-likelihood analysis). All chloroplast genome sequences used to construct the tree were obtained from GenBank, and the species names together with their accession numbers are shown at the branch tips. The following details the sequence used to build the phylogenetic tree: the sequences of *Campomanesia xanthocarpa* O.Berg 1857, KY392760.1[5], *Eugenia uniflora* L. 1753, NC027744.1[6], *Psidium guajava* L. 1753, KX364403.1[7], *Psidium cattleyanum* Sabine 1821, MN095413.1[8], *Plinia trunciflora* (O.Berg) Kausel 1956, NC034801.1[9], *Plinia cauliflora* (Mart.) Kausel 1956, NC039395.1[10], *Plinia aureana* (Mattos) Mattos 1998, NC039557.1[11], *Lagerstroemia fauriei* Koehne 1907, KT358807[12], *Syzygium cumini* (L.) Skeels 1912, NC053327.1[13], *Stockwellia quadrifida* D.J.Carr, S.G.M.Carr & B.Hyland 2002, KC180807, *Allosyncarpia ternata* S.T.Blake 1977, KC180806[14], *Eucalyptus grandis* W.Hill ex Maiden 1919, NC014570.1[15], *Corymbia maculata* (Hook.) K.D.Hill & L.A.S.Johnson 1995, NC022408.1[16], *Angophora costata* (Gaertn.) Britten 1916, KC180805[14].