MITOCHONDRIAL DNA PART B: RESOURCES
2024, VOL. 9, NO. 2, 237–241
https://doi.org/10.1080/23802359.2024.2310133

Taylor & Francis
Taylor & Francis Group

PLASTOME REPORT

OPEN ACCESS Check for updates

The complete plastome sequence of *Datisca cannabina* L. (Datiscaceae, Cucurbitales)

Sangjin Jo[a], Ziyoviddin Yusupov[b], Sangho Choi[a], Jin-Hyub Paik[a] and Komiljon Sh. Tojibaev[b]

[a]International Biological Material Research Center (IBMRC), Korea Research Institute of Bioscience and Biotechnology (KRIBB), Daejeon, South Korea; [b]Institute of Botany, Academy of Sciences, Tashkent, Uzbekistan

ABSTRACT

In this study, we report the first complete plastome sequence of *Datisca cannabina* (GenBank acc. no. OP432690). The plastome had a typical quadripartite structure. Its size was 162,914 bp, consisting of 90,890 bp large single-copy (LSC), 19,296 bp small single-copy (SSC), and 26,364 bp inverted repeat (IR) regions. It contained 112 genes, including 78 protein-coding, 30 tRNA, and four rRNA genes. The *inf*A gene was pseudogenized. Sixteen genes contain one intron and two genes (*clp*P and *ycf*3) had two introns. Our phylogenetic tree showed that *D. cannabina* formed a close relationship with Begoniaceae. However, further samples are required to determine the phylogenetic placement of Datiscaceae in Cucurbitales.

ARTICLE HISTORY
Received 8 September 2023
Accepted 21 January 2024

KEYWORDS
Cucurbitales; Datiscaceae; *Datisca cannabina*; phylogeny; plastome

Introduction

The genus *Datisca* belongs to the family Datiscaceae in the order Cucurbitales (APG IV 2016). It contains only two species of the genus *Datisca* (*D. cannabina* Linnaeus 1753 and *D. glomerata* (C.Presl) Baill 1870) with a disjunct distribution. *Datisca glomerata* is endemic to the western United States and Mexico, whereas *D. cannabina* is distributed from the eastern

Figure 1. *Datisca cannabina* L. *D. cannabina* is perennial or subshrub. The leaves of this species resemble hemp, so it is also called false hemp. The species photo was taken by the authors (KST and ZY) in Charvak, Uzbekistan, June 2021, without any copyright issues. A voucher specimen was mounted by KST and ZY and deposited in the Korea Research Institute of Bioscience & Biotechnology (KRIBB) Herbarium (KRIB, acc. no. KRIB 0088968, Jin-Hyub Paik, jpaik@kribb.re.kr).

CONTACT Komiljon Sh. Tojibaev ktojibaev@mail.ru Institute of Botany, Academy of Sciences, Durmon yuli str. 32, Tashkent 100125, Uzbekistan

Supplemental data for this article can be accessed online at https://doi.org/10.1080/23802359.2024.2310133.

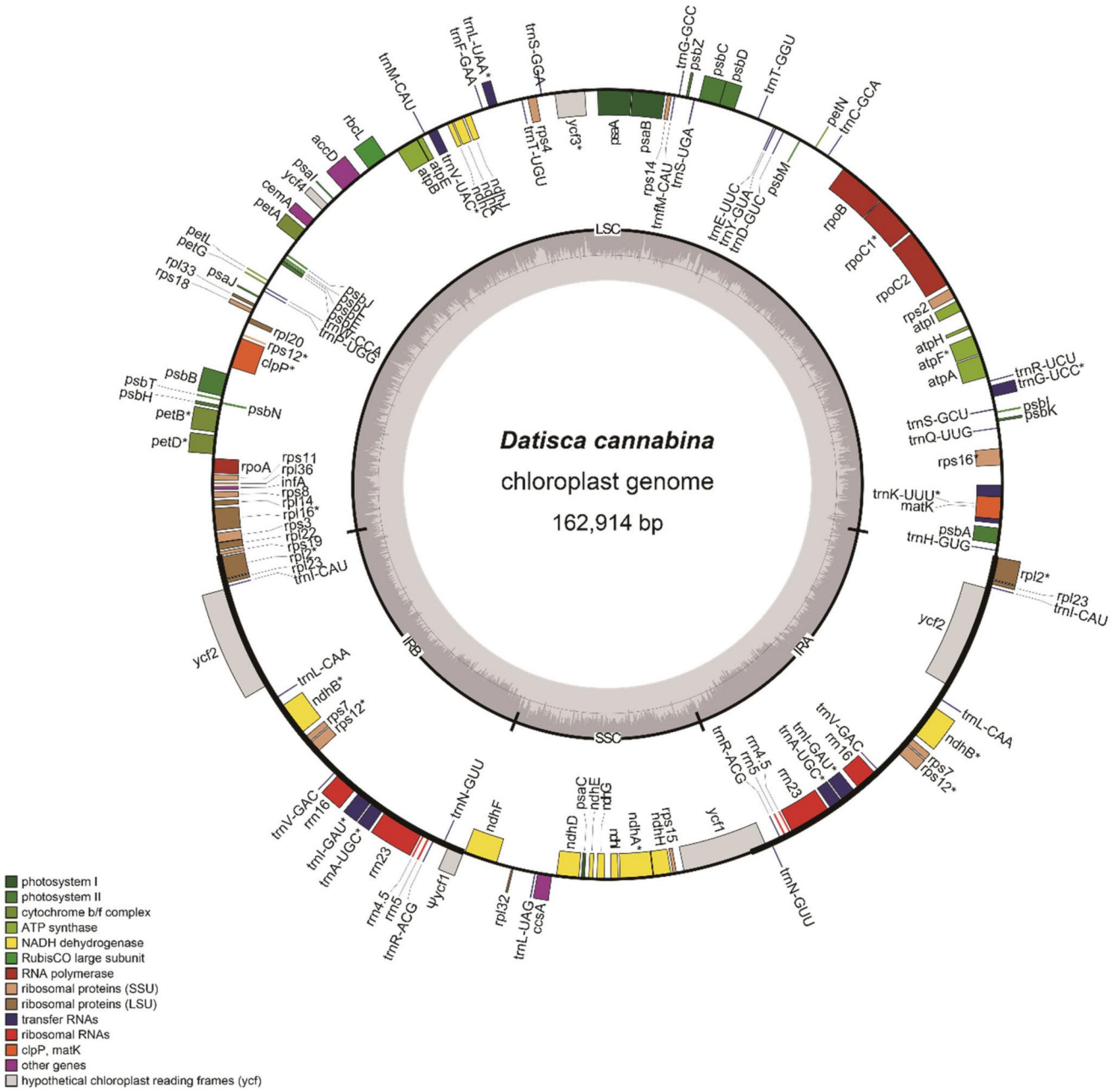

Figure 2. Complete plastome map of *Datisca cannabina* L. The circular plastome map was constructed by OGDRAW. Colors identify genes from the same functional category, following the figure legends. In the inner circle, the dark bars indicate the guanine + cytosine content. IRa and IRb: inverted repeat regions; LSC: large single-copy region; SSC: small single-copy; Ψ: pseudogene. Asterisks indicate genes containing introns.

Mediterranean to Central Asia. *D. cannabina* has been used to treat rheumatism and to obtain yellow dye (Singh and Kachroo 1976; Chopra et al. 1986). Previously, Datiscaceae formed a family with Tetramelaceae (Bentham and Hooker 1867), and these two families formed a close relationship with Begoniaceae. However, their exact relationships remain unclear due to scarce genetic data. In this study, we report for the first time the complete plastome of *D. cannabina* (Datiscaceae), a useful plant resource. This study was conducted to resolve the unclear phylogenetic relationships within Cucurbitales.

Materials and methods

The silica-dried leaves used in this study were collected from Chatkal mountain, near the Charvak reservoir in Uzbekistan (N 41° 36′ 48.3″ E 69° 58′ 24.6″) (Figure 1). *Datisca cannabina* is not an endangered or protected species. Ethics approval for this study was obtained from the Institutional Bioethics Committee of KRIBB. The total DNAs using the silica-dried leaves were extracted using the G-spin™ IIp for Plant Genomic DNA Extraction Kit (iNtRON Biotechnology, Seongnam, South Korea). The voucher specimen and genomic DNAs were deposited in the Korea Research Institute of Bioscience & Biotechnology (KRIBB) Herbarium (KRIB, acc. no. KRIB 0088968, Jin-Hyub Paik, jpaik@kribb.re.kr) and the International Biological Material Research Center (IBMRC acc. no. KRIB 0088968). Approximately, 100 ng of genomic DNAs were used for library construction. We performed the next-generation sequencing (NGS) using an Illumina MiSeq platform (Illumina Inc., San Diego, CA). The total raw reads were

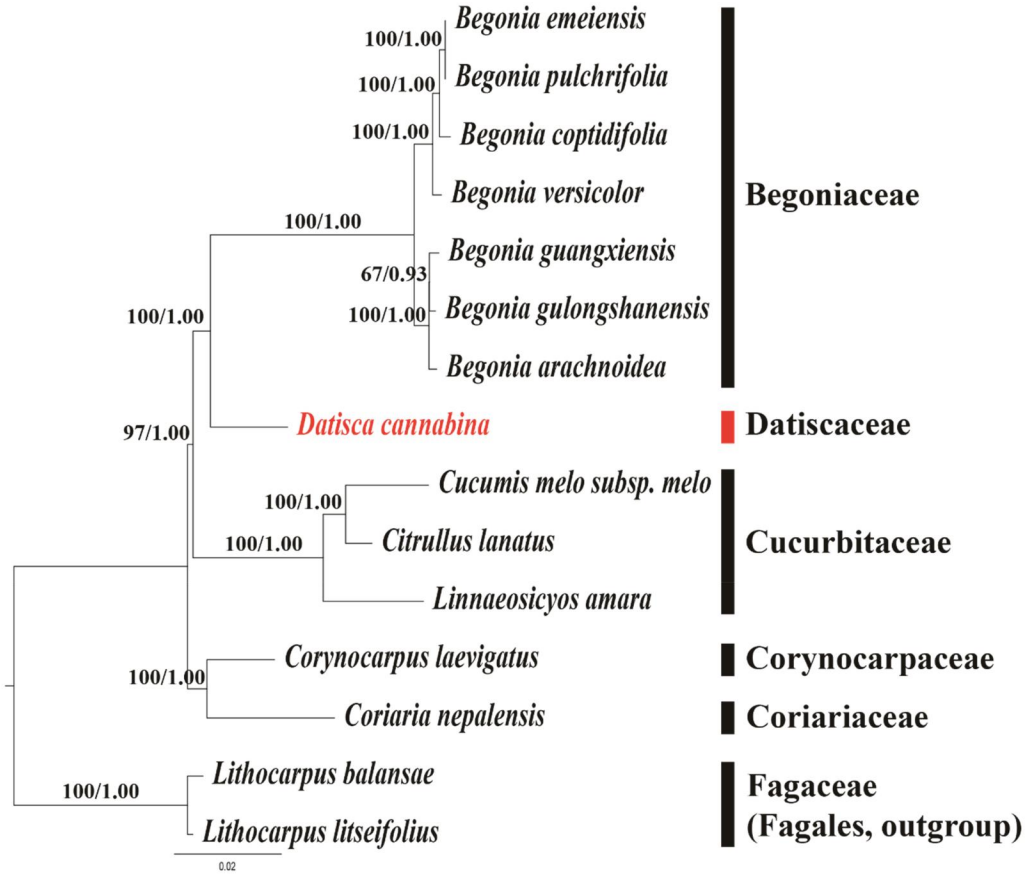

Figure 3. Maximum-likelihood (ML) tree based on 77 protein-coding genes from 15 plastomes as determined by RAxML (–ln L = 176502.749233). The numbers at each node indicate the ML bootstrap values and Bayesian posterior probabilities. Both ML and BI trees have the same topology. *Begonia arachnoidea* (MZ671994) (Tao et al. 2022); *B. coptidifolia* (NC_056110) (Wang et al. 2021); *B. emeiensis* (NC_061410) (unpublished); *B. guangxiensis* (NC_046385) (Dong et al. 2019); *B. gulongshanensis* (MZ671995) (Guan et al. 2022); *B. pulchrifolia* (NC_045096) (Fan et al. 2019); *B. versicolor* (NC_047450) (Zhou et al. 2020); *Citrullus lanatus* (NC_032008) (Zhu et al. 2016); *Coriaria nepalensis* (NC_059766) (Fang et al. 2020); *Corynocarpus laevigata* (NC_014807) (Atherton et al. 2010); *Cucumis melo* subsp. *melo* (NC_015983) (Rodríguez-Moreno et al. 2011); *Datisca cannabina* L. (OP432690) (this study); *Linnaeosicyos amara* (NC_046863) (Bellot et al. 2020); *Lithocarpus balansae* (NC_026577) (Unpublished); *Lithocarpus litseifolius* (NC_063927) (unpublished).

trimmed using BBDuk and then went through a normalization process. Normalized reads were used for *de novo* assembly in Geneious Prime v. 2022.2.2 (Kearse et al. 2012). The complete plastome was annotated using the Geneious Prime v. 2022.2.2, National Center for Biotechnology Information (NCBI) BLAST (blast.ncbi.nlm.nih.gov/Blast.cgi) and tRNAscan-SE programs (Lowe and Eddy 1997). The circular plastome map was constructed by OrganellarGenomeDraw (OGDRAW) (Greiner et al. 2019). The structures of intron-containing genes were visualized in CPGView (Liu et al. 2023). For the phylogenetic analysis, we selected and downloaded 14 complete plastome sequences (12 Cucurbitales and two Faglaes (outgroup)) based on the APG IV system (APG IV 2016) from the NCBI database. Phylogenetic analysis was performed on a data set that included 77 protein-coding genes (excl. *inf*A and *ndh*F) from the 15 selected taxa using RAxML v.8.2.12 in CIPRES webserver (phylo.org) (Stamatakis 2014; Miller et al. 2015). The 77 gene sequences (67,476 bp in length) were aligned with the MUSCLE program using Geneious Prime v. 2022.2.2. The data set was also used to construct a Bayesian inference (BI) tree. MrBayes in Geneious Prime v. 2022.2.2 was used to construct a BI tree with a Markov chain Monte Carlo (MCMC) chain length of 1,000,000 and GTR model (Huelsenbeck and Ronquist 2001).

Results and discussion

The *Datisca cannabina* plastome was 162,914 bp in length and exhibited a typical quadripartite structure (Figure 2). The large single-copy (LSC) region (90,890 bp) and a small single-copy (SSC) region (19,296 bp) were separated by a pair of inverted repeat (IR) regions (26,364 bp). Coverage was 2515.9× (Figure S1). The plastome comprised 112 unique genes (78 protein-coding genes, 30 tRNA genes, and four rRNA genes). Six protein-coding, seven tRNA, and four rRNA genes were duplicated in the IR regions. The *inf*A gene was predicted to be a pseudogene. The average GC content was 36.6%. Sixteen genes contain one intron. Among them, 10 genes are protein-coding genes (Figures S2 and S3) and the others are tRNA genes (*trn*A-UGC, *trn*G-UCC, *trn*I-UGC, *trn*K-UUU, *trn*L-UAA, and *trn*V-UAC). Additionally, two genes, *clp*P and *ycf*3, had two introns (Figure S2).

The topologies of the maximum-likelihood (ML) and BI phylogenetic trees constructed in this study were the same. The phylogenetic tree showed that *D. cannabina* is sister to the Begoniaceae clade with 100% bootstrap support (Figure 3). The clade formed a sister group with Cucurbitaceae (Figure 3). In previous studies, using partial organelle DNA

sequences such as *mat*K, *mat*R, and ITS (Zhang et al. 2006; Schaefer et al. 2009; Filipowicz and Renner 2010; Schaefer and Renner 2011), Datiscaceae formed a single clade with Begoniaceae and Tetramelaceae. Although their relationship was reported to be close, the exact relationship remained still unclear due to differing results for each marker and a lack of complete plastome studies of Tetramelaceae. Our updated results also show that Datiscaceae and Begoniaceae are sister, but further studies on Tetramelaceae samples are needed to clarify their relationship. In addition to the relationship between Begoniaceae-Datiscaceae-Tetramelaceae (BDT), the relationship with other families of Cucurbitales also needs to be addressed.

Conclusions

In this study, we reported the first complete plastome sequence of Datiscaceae (*Datisca cannabina* L.). The plastome shows the typical structure commonly observed in angiosperms. Our phylogenetic tree indicated a close relationship between Datiscaceae and Begoniaceae. In previous studies, Begoniaceae formed a closer relationship with Tetramelaceae than with Datiscaceae. Therefore, further studies of Tetramelaceae are needed to resolve the relationships of the BDT clade. Additionally, research into why the distribution of the two species included in Datiscaceae is significantly different is also an interesting topic. Therefore, this result will provide a useful resource for the phylogenetic studies of Cucurbitales.

Author contributions

SJ, ZY, SC, JHP, and KST contributed to the concept and design of the study. SJ, KST, and ZY collected samples and performed the experiments and analysis. SJ wrote the manuscript. ZY, SC, JHP, and KST revised the manuscript. All authors read and approved the manuscript.

Disclosure statement

No potential conflict of interest was reported by the author(s).

Funding

This research was supported by the KRIBB Initiative Program [KGM4582423] of the Republic of Korea.

ORCID

Ziyoviddin Yusupov http://orcid.org/0000-0003-2278-542X

Data availability statement

The data that support the finding of this study are openly available in GenBank of NCBI at https://www.ncbi.nlm.nih.gov, reference number OP432690 for *Datisca cannabina*. The associated BioProject, BioSample, and SRA numbers are PRJNA880439, SAMN30840781, and SRR21559895, respectively.

References

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):1–20.

Atherton R, McComish BJ, Shepherd LD, Berry LA, Albert NW, Lockhart PJ. 2010. Whole genome sequencing of enriched chloroplast DNA using the Illumina GAII platform. Plant Methods. 6(1):22. doi:10.1186/1746-4811-6-22.

Bellot S, Mitchell TC, Schaefer H. 2020. Phylogenetic informativeness analyses to clarify past diversification processes in Cucurbitaceae. Sci Rep. 10(1):488. doi:10.1038/s41598-019-57249-2.

Bentham G, Hooker JD. 1867. Genera Plantarum I—part 3. London: Reeve.

Chopra RN, Nayar SL, Chopra IC. 1986. Glossary of Indian medicinal plants (including the Supplement). New Delhi: Publisher Council of Scientific and Industrial Research.

Dong LN, Du XY, Zhou W. 2019. The complete plastid genome sequence of *Begonia guangxiensis*. Mitochondrial DNA B Resour. 4(2):3766–3767. doi:10.1080/23802359.2019.1681322.

Fan J, Li XJ, Li CH, Yang BN. 2019. The complete chloroplast genome and phylogenetic analysis of *Begonia pulchrifolia*, a near endangered Begoniaceae plant. Mitochondrial DNA B Resour. 4(2):2830–2831. doi:10.1080/23802359.2019.1660268.

Fang H, Shang F, Qian J, Duan B. 2020. Phylogenetic relationship and characterization of the complete chloroplast genome of the *Coriaria nepalensis* Wall. in China, a least concern folk medicine. Mitochondrial DNA Part B Resour. 5(2):1718–1719. doi:10.1080/23802359.2020.1749179.

Filipowicz N, Renner SS. 2010. The worldwide holoparasitic Apodanthaceae confidently placed in the Cucurbitales by nuclear and mitochondrial gene trees. BMC Evol Biol. 10(1):219. doi:10.1186/1471-2148-10-219.

Greiner S, Lehwark P, Bock R. 2019. OrganellarGenomeDRAW (OGDRAW) version 1.3.1: expanded toolkit for the graphical visualization of organellar genomes. Nucleic Acids Res. 47(W1):W59–W64. doi:10.1093/nar/gkz238.

Guan S, Song Q, Zhou J, Liu Y, Yan H, Tao D, Luo S. 2022. The complete chloroplast genome sequence of *Begonia gulongshanensis*. Mitochondrial DNA B Resour. 7(1):191–193. doi:10.1080/23802359.2021.2018949.

Huelsenbeck JP, Ronquist F. 2001. MRBAYES: Bayesian inference of phylogenetic trees. Bioinformatics. 17(8):754–755. doi:10.1093/bioinformatics/17.8.754.

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(12):1647–1649. doi:10.1093/bioinformatics/bts199.

Liu S, Ni Y, Li J, Zhang X, Yang H, Chen H, Liu C. 2023. CPGView: a package for visualizing detailed chloroplast genome structures. Mol Ecol Resour. 23(3):694–704. doi:10.1111/1755-0998.13729.

Lowe TM, Eddy SR. 1997. tRNAscan-SE: a program for improved detection of transfer RNA genes in genomic sequence. Nucleic Acids Res. 25(5):955–964. doi:10.1093/nar/25.5.955.

Miller MA, Schwartz T, Pickett BE, He S, Klem EB, Scheuermann RH, Passarotti M, Kaufman S, O'Leary MA. 2015. A RESTful API for access to phylogenetic tools via the CIPRES science gateway. Evol Bioinform Online. 11:43–48.

Rodríguez-Moreno L, González VM, Benjak A, Martí MC, Puigdomènech P, Aranda MA, Garcia-Mas J. 2011. Determination of the melon chloroplast and mitochondrial genome sequences reveals that the largest reported mitochondrial genome in plants contains a significant amount of DNA having a nuclear origin. BMC Genomics. 12(1):424. doi:10.1186/1471-2164-12-424.

Schaefer H, Heibl C, Renner SS. 2009. Gourds afloat: a dated phylogeny reveals an Asian origin of the gourd family (Cucurbitaceae) and numerous overseas dispersal events. Proc Biol Sci. 276(1658):843–851. doi:10.1098/rspb.2008.1447.

Schaefer H, Renner SS. 2011. Phylogenetic relationships in the order Cucurbitales and a new classification of the gourd family (Cucurbitaceae). Taxon. 60(1):122–138. doi:10.1002/tax.601011.

Singh G, Kachroo P. 1976. Forest flora of Srinagar and plants of neighbourhood. Dehradun, India: Bishen Singh Mahendra Pal Singh.

Stamatakis A. 2014. RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. Bioinformatics. 30(9):1312–1313. doi:10.1093/bioinformatics/btu033.

Tao DY, Yan HX, Zhou JY, Rong JB, Song Q, Guan SK, Luo S. 2022. The complete chloroplast genome sequence of *Begonia arachnoidea* (Begoniaceae). Mitochondrial DNA B Resour. 7(6):1106–1108. doi:10.1080/23802359.2022.2087548.

Wang ZF, Liu TH, Cao HL. 2021. The complete chloroplast genome sequence of *Begonia coptidifolia*. Mitochondrial DNA B Resour. 6(2):548–549. doi:10.1080/23802359.2021.1872434.

Zhang L-B, Simmons MP, Kocyan A, Renner SS. 2006. Phylogeny of the Cucurbitales based on DNA sequences of nine loci from three genomes: implications for morphological and sexual system evolution. Mol Phylogenet Evol. 39(2):305–322. doi:10.1016/j.ympev.2005.10.002.

Zhou S, Wen M, Wang L, Wang X, Guo W, Xu Y. 2020. The complete chloroplast genome sequence of *Begonia versicolor* Irmsch. (Begoniaceae). Mitochondrial DNA B Resour. 5(3):2113–2114. doi:10.1080/23802359.2020.1765706.

Zhu Q, Cui H, Zhao Y, Gao P, Liu S, Wang P, Luan F. 2016. The complete chloroplast genome sequence of the *Citrullus lanatus* L. subsp. *vulgaris* (Cucurbitaceae). Mitochondrial DNA B Resour. 1(1):943–944. doi:10.1080/23802359.2016.1261611.