

## The complete mitochondrial genome of *Cuminum cyminum* (Apiales: Apiaceae) and phylogenetic analysis

Ge Chen<sup>a\*</sup>, Ruiping Kang<sup>b\*</sup>, Zihao Wang<sup>a</sup>, Yu Jiang<sup>a</sup>, Huiying Zhou<sup>b</sup>, Aifeire Abuduaini<sup>b</sup>, Feiya Suo<sup>b</sup> and Luodong Huang<sup>a</sup>

<sup>a</sup>College of Life Science and Technology, Guangxi University, Nanning, Guangxi, China; <sup>b</sup>College of Life Science and Technology, Xinjiang University, Urumchi, Xinjiang, China

### ABSTRACT

Cumin (*Cuminum cyminum* L.) belongs to the family Apiaceae and the order Apiales, which is a widely grown spice and medicinal plant in Xinjiang province, China. In the current study, whole genome sequencing of *C. cyminum* was performed using the Illumina HiSeq 4000 platform, and the complete mitogenome sequence was assembled and annotated. We found that the single circular mitogenome of *C. cyminum* was 246,721 bp in length, and has about 45.5% GC content. It comprised 73 genes in the coding region (35 protein-coding genes, 18 tRNA genes, 3 rRNA genes, and 15 open-reading frames) and a non-coding region. Phylogenetic analysis indicated that *C. cyminum* is closely related to *Daucus carota* and the subtribes *Daucinae*. The mitogenome of *C. cyminum* revealed its phylogenetic relationships with other species in the Apiaceae family, which would further help in understanding its evolution.

### ARTICLE HISTORY

Received 18 November 2022  
Accepted 13 July 2023

### KEYWORDS

*Cuminum cyminum*;  
mitogenome; apiaceae;  
phylogenetic analysis

### Introduction

*Cuminum cyminum* L. (1753), commonly known as cumin, is an annual or biennial herbaceous plant belonging to the genus *Cuminum* of the Apiaceae family. Cumin fruit is native to the Mediterranean, Egypt, and Ethiopia. It is widely used as a spice and also for medicinal purposes (Sultanbawa 2016; Singh 2017; Zhou et al. 2023). It is extensively cultivated as a spice plant in several regions of the Former Soviet Union, Iran, India, Mexico, North America, Northern Africa, and Xinjiang province in China.

The cultivation of *C. cyminum* in Xinjiang province began thousands of years ago, and it was the main cumin-producing area in China. The least 15 landraces of *C. cyminum* have been domesticated for adaptation to the unique climatic conditions of the southern regions of Xinjiang province. Our investigation study revealed that the morphological diversity of *C. cyminum* is generated due to its morphological appearance based on the climatic conditions and the region where it is grown which ultimately affects its production as well as the stress resistance traits. Therefore, the evolutionary relationship of genetic diversity is worth further study. However, studies presenting the complete mitogenome of *C. cyminum* are lacking. The current study is focused on the characterization of complete mitogenome (mitochondrial genome)

features of *C. cyminum* and its phylogenetic relationship with other related species belonging to the clade Campanulids. The complete mitogenome sequence of *C. cyminum* will serve as a valuable tool for elucidating the phylogenetic relationship and evolutionary history of *C. cyminum* in the family Apiaceae.

### Materials and methods

#### Plant materials and DNA extraction

The seeds of *C. cyminum* L. (1753) were collected from Tulufan in Xinjiang Province, China (89°1857'N, 42°6793'W) (Figure 1(A,B)). The voucher specimen (Figure 1(C)) was deposited in the Herbarium of the College of Life Science and Technology at Guangxi University, Nanning, Guangxi, China (Dr. Luodong Huang, [yh Huangld@gxu.edu.com](mailto:yh Huangld@gxu.edu.com)) with the voucher number GXU-Z2020018.

The seeds of *C. cyminum* were sterilized by washing with 75% alcohol for 1 min followed by treatment with 1% sodium hypochlorite solution. After treatment, the seeds were washed with deionized water. Then, the sterilized seeds were germinated on MS medium under a 16 h-light/8 h-dark photoperiod at 28 °C for 15 d (Figure 1(D)). The samples were

**CONTACT** Feiya Suo  [sophi3106@sina.com](mailto:sophi3106@sina.com)  College of Life Science and Technology, Xinjiang University, No. 666 Shengli Road, Urumchi, Xinjiang 830046, China; Luodong Huang  [yh Huangld@gxu.edu.cn](mailto:yh Huangld@gxu.edu.cn)  College of Life Science and Technology, Guangxi University, No.100 Daxue East Road, Nanning, Guangxi 530004, China

\*These authors have contributed equally to this work

 Supplemental data for this article can be accessed online at <https://doi.org/10.1080/23802359.2023.2238357>.

© 2023 The Author(s). Published by Informa UK Limited, trading as Taylor & Francis Group.  
This is an Open Access article distributed under the terms of the Creative Commons Attribution-NonCommercial License (<http://creativecommons.org/licenses/by-nc/4.0/>), which permits unrestricted non-commercial use, distribution, and reproduction in any medium, provided the original work is properly cited. The terms on which this article has been published allow the posting of the Accepted Manuscript in a repository by the author(s) or with their consent.



**Figure 1.** Plant characteristics image of *Cuminum cyminum*. (A) the *C. cyminum* cultivated in Tulufan city, Xinjiang province, China. It is between 50 and 120 cm high and racemes termina; stem is erect, upright part is quadrangular, the edge is densely covered with white hairs, the lower and Middle branches are often purple; leaves are rounded kidney-shaped, leaves green glabrous or pubescent, grey-green below pubescent. (B) the *C. cyminum* matures, dries and turns yellow in the plantation at June. (C) the *C. cyminum* seeds are ellipsoidal or quasi-circular in shape, 5–8 mm in length, grayish white to dark grey in surface, with irregular longitudinal projections and striations, shiny. (D) the germination of seed tissue culture to form aseptic seedling was used to extract DNA in this study. The photograph A, and B were taken by the Feiya Suo in the Tulufan, Xinjiang province, China (89°1857'N, 42°6793'W). the photograph C, and D was taken by the Luodong Huang on 10 October 2022 at the biological resources and environmental microbiology laboratory of the College of Life Science and Technology, Guangxi University.

collected from 15-day-old sterile seedlings for DNA isolation using the MiniBEST Plant Genomic DNA Extraction Kit (TaKaRa, China).

### Genome sequencing, assembly, and annotation of the mitogenome

Total genomic DNA was deposited at Guangxi University, and purified DNA was used to construct a short 270 bp genomic library and sequenced on the Illumina HiSeq4000 platform to produce 150 bp paired-end reads. The sequencing was performed by BGI Biotechnology Co. Ltd (Shenzhen, China). After quality control and removal of low-quality reads, the high-quality clean reads were assembled by the SPAdes-3.14.0 tool (Bankevich et al. 2012). After assembly, one contig of the mitochondrial DNA was visualized by Bandage software (Wick et al. 2015) and identified aligning the sequences to the mitogenome of *Daucus carota* (NC\_017855.1) (Iorizzo

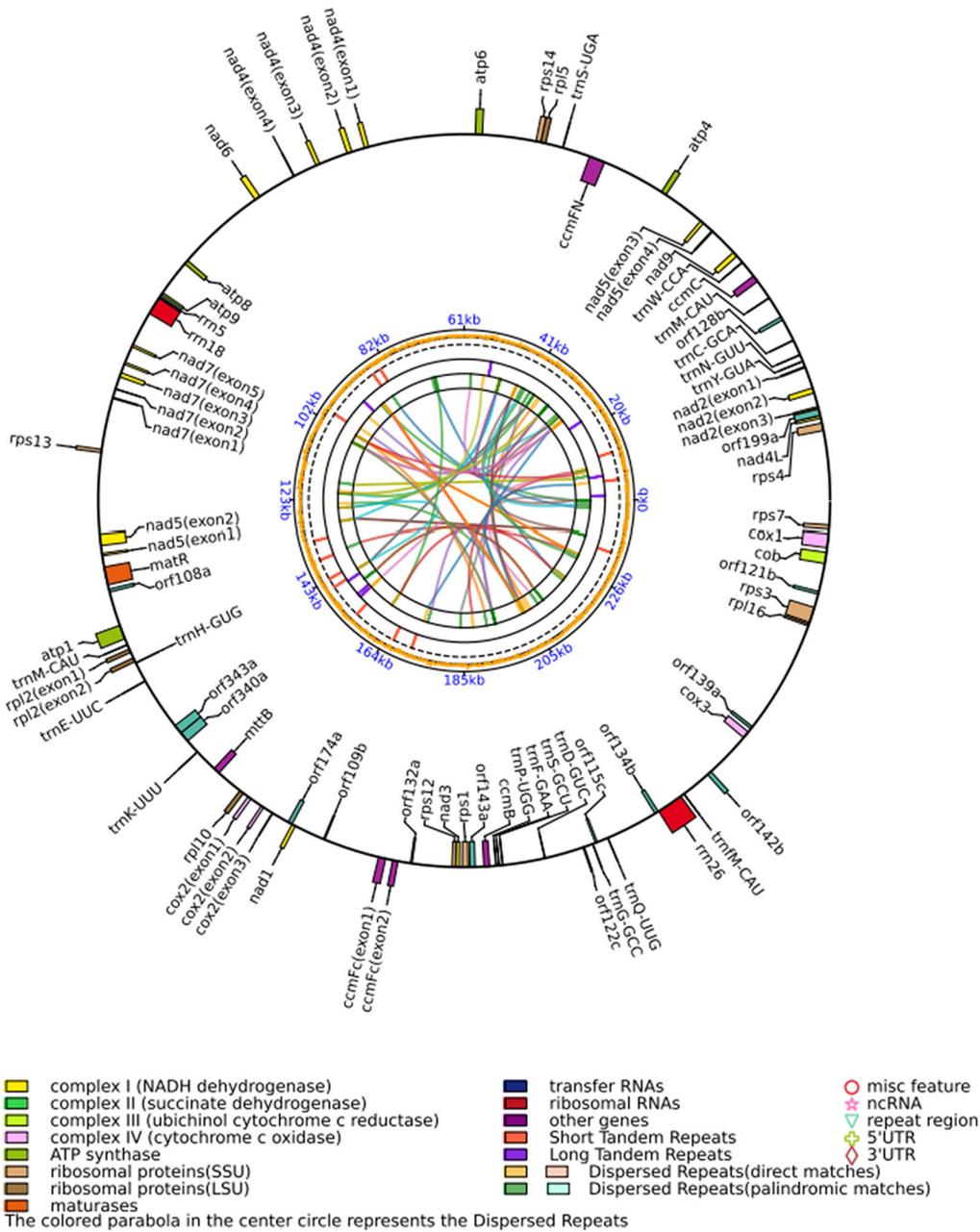
et al. 2012) and *Bupleurum falcatum* (KX887330.1) (Kim et al. 2020) using blastn software. Finally, the mitogenome annotation was performed using the MITOFY Web server, GeSeq web server, UGENE ORFs finder, and tRNAscan-SE followed by manual corrections. After multiple rounds of modification and manual corrections by Appollo software, the complete mitogenome of *C. cyminum* was submitted to NCBI (GenBank accession no. OP800912). The ogview (<http://www.1kmpg.cn/ogview>) used to draw circular genome images, and cis-splicing gene images.

### Phylogenetic analysis

Twenty-seven mitogenomes from different species belonging to the clade Campanulids were chosen for phylogenetic analysis. The sequences of common 30 genes (*atp1*, *atp4*, *atp6*, *atp8*, *atp9*, *ccmB*, *ccmC*, *ccmFn*, *ccmFc*, *cytb*, *cox1*, *cox2*, *cox3*, *matR*, *mttB*, *nad1*, *nad2*, *nad3*, *nad4*, *nad4L*, *nad5*, *nad6*, *nad7*,

# Cuminum cyminum

Mitochondrial Genome  
246,721 bp GC: 45.51%

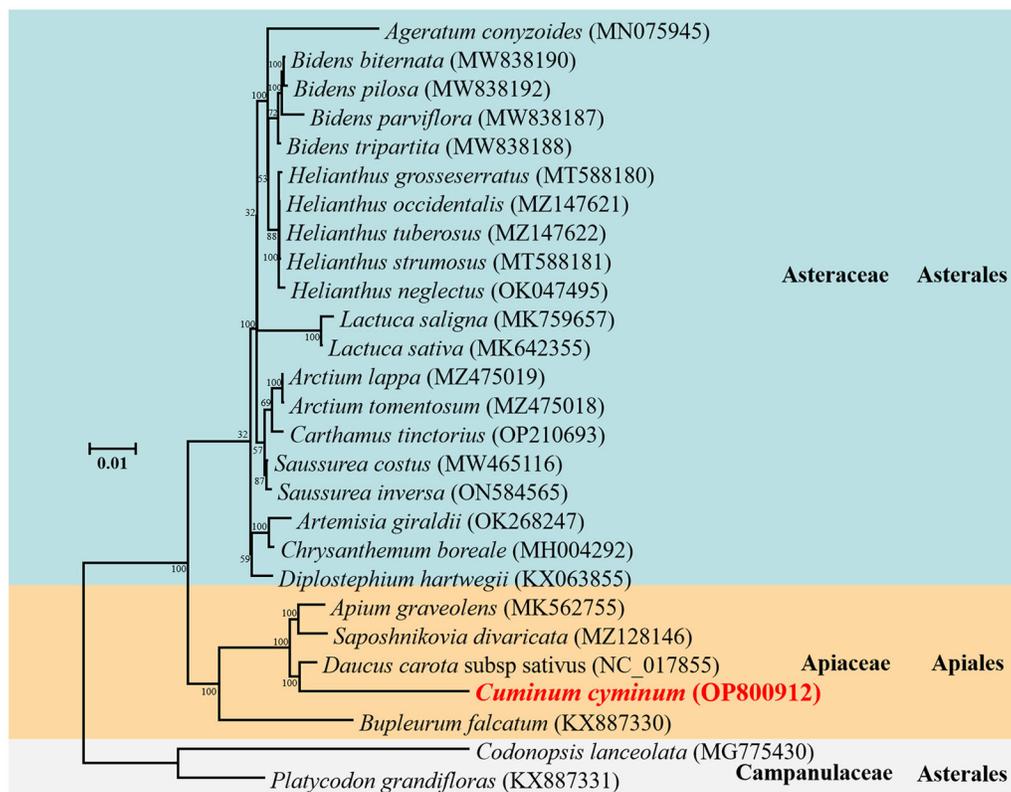


**Figure 2.** Mitogenome circular map of *Cuminum cyminum*. The genes name on the outside line side indicated that these genes were located on the H-strand, whereas the others were located on the L-strand. Genes are shown in different color blocks on the map.

*nad9*, *rpl5*, *rpl10*, *rps3*, *rps4*, *rps12*, and *rps13*) within the mitogenomes were extracted, concatenated, and calculated with Phylosuite (Zhang et al. 2020). The sequences were aligned by MACSE (Ranwez et al. 2018) and Homblocks (Bi et al. 2018), resulting in the alignments of each species with 26,781 bp. The phylogenetic tree was constructed with IQ-TREE (Nguyen et al. 2015) using the maximum-likelihood method based on the GTR+I+G4 model with 1000 bootstrap replicates.

## Results and discussion

The length of the circular *C. cyminum* mitogenome was 246,721 bp. The read coverage depth map was shown in [Supplementary Figure S1](#), the average and minimum read mapping depths of the assembled genome were 1944 $\times$  and 421 $\times$ , respectively. The mitogenome genome contained 35 protein-coding genes (PCGs), 3 ribosomal RNA genes (5S rRNA, 18S rRNA, 26S rRNA), 18 transfer RNA, 17 open reading



**Figure 3.** Phylogenetic relationships for *Cuminum cyminum* according to the maximum-likelihood phylogenetic tree constructed based on 27 mitogenomes of campanulids species. Node labels indicate bootstrap values. GenBank accession numbers of the following sequences were used: *Ageratum conyzoides* (MN075945) (Luo and Pan 2019), *Bidens biternata* (MW838190) (Yao et al. 2022<sup>#</sup>), *Bidens pilosa* (MW838192) (Yao et al. 2022<sup>#</sup>), *Bidens parviflora* (MW838187) (Yao et al. 2022<sup>#</sup>), *Bidens tripartita* (MW838188) (Yao et al. 2022<sup>#</sup>), *Helianthus grosseserratus* (MT588180) (Makarenko et al. 2020), *Helianthus occidentalis* (MZ147621) (Makarenko et al. 2021), *Helianthus tuberosus* (MZ147622) (Makarenko et al. 2021), *Helianthus strumosus* (MT588181) (Makarenko et al. 2020), *Helianthus neglectus* (OK047495) (Khachumov et al. 2021<sup>#</sup>), *Lactuca saligna* (MK759657) (Kozik et al. 2019), *Lactuca sativa* (MK642355) (Kozik et al. 2019), *Arctium lappa* (MZ475019) (Xu and Kang 2021<sup>#</sup>), *Arctium tomentosum* (MZ475018) (Xu and Kang 2021<sup>#</sup>), *Carthamus tinctorius* (OP210693) (Yang 2019<sup>#</sup>), *Saussurea costus* (MW465116) (Song et al. 2021), *Saussurea inversa* (ON584565) (Dai et al. 2022<sup>#</sup>), *Artemisia giraldii* (OK268247) (Yue 2022<sup>#</sup>), *Chrysanthemum boreale* (MH004292) (Won et al. 2018), *Diplostephium hartwegii* (KX063855) (Vargas et al. 2017), *Apium graveolens* (MK562755) (Cheng et al. 2021), *Saposhnikovia divaricata* (MZ128146) (Ni 2021<sup>#</sup>), *Daucus carota* subsp. *sativus* (NC\_017855) (Iorizzo et al. 2012), *Bupleurum falcatum* (KX887330) (Kim et al. 2020), *Codonopsis lanceolata* (MG775430) (Lee et al. 2018), *Platycodon grandifloras* (KX887331) (Lee et al. 2018). <sup>#</sup>direct submission to NCBI, unpublished.

frames, and non-coding region (Figure 2). This cis-splicing gene images was shown in Supplementary Figure S2.

The overall nucleotide composition of the mitogenome is as follows: A: 27.1%, T: 27.4%, G:22.6%, and C: 22.9%, with a slight A + T bias (54.5%). Furthermore, the 35 PCGs were *atp1*, *atp4*, *atp6*, *atp8*, *atp9*, *ccmB*, *ccmC*, *ccmFN*, *ccmFc*, *cytb*, *cox1*, *cox2*, *cox3*, *matR*, *mttB*, *nad1*, *nad2*, *nad3*, *nad4*, *nad4L*, *nad5*, *nad6*, *nad7*, *nad9*, *rpl2*, *rpl5*, *rpl10*, *rpl16*, *rps1*, *rps3*, *rps4*, *rps7*, *rps12*, *rps13*, and *rps14*. In the current study, we found that some of ribosomal protein-encoding genes, such as *rps10*, were missing from the *C. cyminum* mitogenome. The mitogenome of *D. carota* had lost *rps10* and *rps14*. The mitogenome of *B. falcatum* had lost *rpl2*, and *rps7*. which is consistent with the previous studies suggesting that a high-frequency loss is observed in the ribosomal protein-encoding in the angiosperm mitogenome (Adams and Palmer 2003). The possible reason could be the loss of non-essential genes due to the course of evolution.

During the evolution of plant mitochondrial genomes, the phenomena of intron loss and transfer are highly evident (Palmer et al. 2000; In-Su Choi et al. 2021). Compared to the mitogenomes of other species belonging to the Apiaceae family, the *C. cyminum* mitogenome was found to have more introns and intergenic regions, with seven genes (*nad2*, *nad4*,

*nad5*, *nad7*, *rpl2*, *cox2*, and *ccmFc*) containing introns. Of these, *nad7* contain four introns, *nad4* and *nad5* contain three introns, *nad2* and *cox2* contain two introns, and the *rpl2* and *ccmFc* contain one intron (Figure 2). The *nad4* gene has the longest intron of 3163 bp, while the *rpl2* gene has the shortest intron of 650 bp. Using the RNAweasel search approach (Lang et al. 2007), we found that all introns belong to the type of Group II in *C. cyminum* mitogenome.

According to the current classification, Campanulids, as a large evolutionary branch, comprises many species, including seven orders, Apiales, Aquifoliales, Asterales, Bruniales, Dipsacales, Escalloniales, and Paracryales (Bremer et al. 2002). However, there are only 29 mitogenomes have been sequenced from the orders Apiales, Aquifoliales, and Asterales. A comprehensive phylogenetic analysis was conducted via the construction of a phylogenetic tree including 27 campanulids species, which unraveled the evolutionary relationship of *C. cyminum*. The phylogenetic analysis showed that *C. cyminum* belongs to the family Apiaceae and order Apiales (Figure 3). All five species from Apiaceae were clustered in one clade. The topology of the phylogenetic tree constructed in this study suggested that *C. cyminum* is positioned in the sister branch of *D. carota* (NC 017855) with a supporting bootstrap value of 100 and was phylogenetically

closer to other species in the family Apiaceae. Furthermore, the Apiaceae formed a clade sister to the clade consisting of other 20 Asteraceae species.

## Conclusion

The current study presents the complete mitogenome of *C. cyminum* and its phylogenetic relationship within the evolutionary branch encompassing other Campanulids species. These results would provide a basis to study the genetics, systematics, and evolution relationship of *C. cyminum* within the family Apiaceae. The mitogenome study within the family Apiaceae must be expanded to better understand the phylogenetic relationship with other species in the order Apiales species.

## Acknowledgement

We thank Bullet Edits Limited for the linguistic editing and proofreading of the manuscript.

## Ethical approval

Since the Project of the National Natural Science Foundation of China [31860077], Ms. Feiya Suo has obtained permission by Xinjiang University to collect *C. cyminum* samples. Dr. Luodong Huang collected the specimen of *C. cyminum* for further molecular study. The plant material collection and experimental research were conducted according to the Plant Protection and Regulation of Xinjiang University and Guangxi University.

## Authors' contributions

All authors contributed to the concept and design of the study. Sample collection and preparation of materials were done by Huiying Zhou, Aifeire Abuduaini, and Feiya Suo. Ge Chen, and Ruiping Kang performed the experiments, analyzed the data, and wrote the first draft of the manuscript. Zihao Wang, and Yu Jiang were involved in tools for the analysis, sorting out the results, and preparing figures. Luodong Huang read and revised the manuscript. Feiya Suo and Luodong Huang were involved in the conception and design of the work, funding acquisition, revising it critically for intellectual content and the final approval of the version to be published. And all authors agree to be accountable for all aspects of the work.

## Disclosure statement

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

## Funding

This work was supported by the National Natural Science Foundation of China [31860077], and the High-level introduction of talents research startup fund of Guangxi University [A3130051004].

## Data availability statement

The genome sequence data that support the findings of this study are openly available in GenBank of NCBI under accession no. OP800912. The associated BioProject, BioSample, and SRA numbers are PRJNA899008, SAMN31636213, and SRR222214714, respectively.

## References

- Adams KL, Palmer JD. 2003. Evolution of mitochondrial gene content: gene loss and transfer to the nucleus. *Mol Phylogenet Evol.* 29(3):380–395. doi: [10.1016/s1055-7903\(03\)00194-5](https://doi.org/10.1016/s1055-7903(03)00194-5).
- Bankevich A, Nurk S, Antipov D, Gurevich AA, Dvorkin M, Kulikov AS, Lesin VM, Nikolenko SI, Pham S, Pribelski AD, et al. 2012. SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing. *J Comput Biol.* 19(5):455–477. doi: [10.1089/cmb.2012.0021](https://doi.org/10.1089/cmb.2012.0021).
- Bremer B, Bremer K, Heidari N, Erixon P, Olmstead RG, Anderberg AA, Kallersjö M, Barkhordarian E. 2002. Phylogenetics of asterids based on 3 coding and 3 non-coding chloroplast DNA markers and the utility of non-coding DNA at higher taxonomic levels. *Mol Phylogenet Evol.* 24(2):274–301. doi: [10.1016/s1055-7903\(02\)00240-3](https://doi.org/10.1016/s1055-7903(02)00240-3).
- Bi G, Mao Y, Xing Q, Cao M. 2018. Homblocks: a multiple-alignment construction pipeline for organelle phylogenomics based on locally collinear block searching. *Genomics.* 110(1):18–22. doi: [10.1016/j.ygeno.2017.08.001](https://doi.org/10.1016/j.ygeno.2017.08.001).
- Cheng Q, Wang P, Li T, Liu J, Zhang Y, Wang Y, Sun L, Shen H. 2021. Complete mitochondrial genome sequence and identification of a candidate gene responsible for cytoplasmic male sterility in celery (*Apium graveolens* L.). *Int J Mol Sci.* 22(16):8584. doi: [10.3390/ijms22168584](https://doi.org/10.3390/ijms22168584).
- Kim CK, Jin MW, Kim YK. 2020. The complete mitochondrial genome sequences of *Bupleurum falcatum* (Apiales: apiaceae)[J]. *Mitochondrial DNA B Resour.* 5(3):2576–2577. doi: [10.1080/23802359.2020.1781566](https://doi.org/10.1080/23802359.2020.1781566).
- Kozik A, Rowan BA, Lavelle D, Berke L, Schranz ME, Michelmore RW, Christensen AC. 2019. The alternative reality of plant mitochondrial DNA: one ring does not rule them all. *PLoS Genet.* 15(8):e1008373. doi: [10.1371/journal.pgen.1008373](https://doi.org/10.1371/journal.pgen.1008373).
- Lang BF, Marie-Josée L, Burger G. 2007. Mitochondrial introns: a critical view. *Trends Genet.* 23(3):119–125. doi: [10.1016/j.tig.2007.01.006](https://doi.org/10.1016/j.tig.2007.01.006).
- Lee HO, Choi JW, Baek JH, Oh JH, Lee SC, Kim CK. 2018. Assembly of the mitochondrial genome in the campanulaceae family using illumina low-coverage sequencing. *Genes.* 30(9):383.
- Luo ZP, Pan LW. 2019. Characterization of the complete mitochondrial genome of *Ageratum conyzoides*. *Mitochondrial DNA B Resour.* 4(2):3540–3541. doi: [10.1080/23802359.2019.1676176](https://doi.org/10.1080/23802359.2019.1676176).
- Makarenko M, Usatov A, Tatarinova T, Azarin K, Kovalevich A, GavriloVA V, Horn R. 2020. The investigation of perennial sunflower species (*Helianthus* L.) mitochondrial genomes. *Genes.* 11(9):982. doi: [10.3390/genes11090982](https://doi.org/10.3390/genes11090982).
- Makarenko MS, Omelchenko DO, Usatov AV, GavriloVA VA. 2021. The insights into mitochondrial genomes of sunflowers. *Plants.* 10(9):1774. doi: [10.3390/plants10091774](https://doi.org/10.3390/plants10091774).
- Nguyen LT, Schmidt HA, von Haeseler A, Minh BQ. 2015. IQ-TREE: a fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. *Mol Biol Evol.* 32(1):268–274. doi: [10.1093/molbev/msu300](https://doi.org/10.1093/molbev/msu300).
- Palmer JD, Adams KL, Cho Y, Parkinson CL, Qiu YL, Song K. 2000. Dynamic evolution of plant mitochondrial genomes: mobile genes and introns and highly variable mutation rates. *Proc Natl Acad Sci USA.* 97(13):6960–6999.
- Sultanbawa Y. 2016. Cumin (*Cuminum cyminum* L.) oils. In: Preedy VR, editors. *Essential oils in food preservation, flavor and safety*. Cambridge: Academic Press. p. 377–383.
- Song YY, Zhang TT, Tang H, Xu L, Xing YP, Zhao R, Bao GH, Li SN, Zhang DC, Qiao Y, et al. 2021. The complete mitochondrial genome of *Aucklandia lappa* Decne. (Asteraceae, Aucklandia Falc.). *Mitochondrial DNA B Resour.* 6(6):1691–1693. doi: [10.1080/23802359.2021.1914524](https://doi.org/10.1080/23802359.2021.1914524).
- Choi I-S, Wojciechowski MF, Ruhlman TA, Jansen RK. 2021. In and out: evolution of viral sequences in the mitochondrial genomes of legumes (Fabaceae). *Mol Phylogenet Evol.* 163:107236. doi: [10.1016/j.ympev.2021.107236](https://doi.org/10.1016/j.ympev.2021.107236).
- Iorizzo M, Senalik D, Szklarczyk M, Grzebelus D, Spooner D, Simon P. 2012. De novo assembly of the carrot mitochondrial genome using next generation sequencing of whole genomic DNA provides first evidence of DNA transfer into an angiosperm plastid genome. *BMC Plant Biol.* 12:61. doi: [10.1186/1471-2229-12-61](https://doi.org/10.1186/1471-2229-12-61).
- Ranwez V, Douzery EJP, Cambon C, Chantret N, Delsuc F. 2018. MACSE v2: toolkit for the alignment of coding sequences accounting for

- frameshifts and stop codons. *Mol Biol Evol.* 35(10):2582–2584. doi: [10.1093/molbev/msy159](https://doi.org/10.1093/molbev/msy159).
- Singh RP, Gangadharappa HV, Kenganora M. 2017. *Cuminum cyminum*-a popular spice: an updated review. *Pharmacogn J.* 9(3):292–301. doi: [10.5530/pj.2017.3.51](https://doi.org/10.5530/pj.2017.3.51).
- Vargas OM, Ortiz EM, Simpson BB. 2017. Conflicting phylogenomic signals reveal a pattern of reticulate evolution in a recent high-Andean diversification (Asteraceae: astereae: diplostephium). *New Phytol.* 214(4):1736–1750. doi: [10.1111/nph.14530](https://doi.org/10.1111/nph.14530).
- Wick RR, Schultz MB, Zobel J, Holt KE. 2015. Bandage: interactive visualization of de novo genome assemblies. *Bioinformatics.* 31(20):3350–3352. doi: [10.1093/bioinformatics/btv383](https://doi.org/10.1093/bioinformatics/btv383).
- Won SY, Jung JA, Kim JS. 2018. The complete mitochondrial genome sequence of *Chrysanthemum boreale* (Asteraceae). *Mitochondrial DNA Part B Resour.* 3:529–530.
- Zhang D, Gao F, Jakovlić I, Zou H, Zhang J, Li WX, Wang GT. 2020. PhyloSuite: an integrated and scalable desktop platform for streamlined molecular sequence data management and evolutionary phylogenetics studies. *Mol Ecol Resour.* 20(1):348–355. doi: [10.1111/1755-0998.13096](https://doi.org/10.1111/1755-0998.13096).
- Zhou HY, Deng HY, Kang RP, Wang Z, Aifeire A, Huang LD, Suo FY. 2023. The protective effect of cuminaldehyde on gastric mucosain *Rattus norregicus* of experimental gastric ulcer. *Chin Pharmacol Bull.* 39(5): 946–952.