MITOCHONDRIAL DNA PART B: RESOURCES
2024, VOL. 9, NO. 1, 143–147
https://doi.org/10.1080/23802359.2024.2305403

Taylor & Francis
Taylor & Francis Group

PLASTOME REPORT

OPEN ACCESS Check for updates

# The complete chloroplast genome of *Brassica rapa* var. *purpuraria* (L.H.Bailey) Kitam 1950 and its phylogenetic analysis

Xianyu Zhou*, Hailong Ren*, Jing Zhang*, Donglin Xu, Wanyu Xiao*, Hongdi Huang, Guangguang Li, Hua Zhang* and Yansong Zheng*

Guangzhou Academy of Agricultural Sciences, Guangzhou, PR China

**ABSTRACT**

Zicaitai (*Brassica rapa* var. *purpuraria* (L.H.Bailey) Kitam 1950) is a vegetable crop that boasts a high nutritional value and unique flavor. It originated from Central China and was formed after long-term cultivation and domestication. In this study, we obtained the complete sequence of the chloroplast genome of zicaitai, a circular molecule of 153,483 bp in length. This chloroplast genome consists of a large single-copy (LSC) region (83,282 bp), a small single-copy (SSC) region (17,775 bp), and a pair of inverted repeats (IRs) (26,213 bp). By sequence annotation, 132 genes, including 87 protein-coding genes, 37 tRNA genes, and eight rRNA genes were identified in the zicaitai chloroplast. A total of 315 simple sequence repeats (SSRs) were found located in LSC (197), SSC (72), and IR (46), respectively. Phylogenetic analysis based on chloroplast genomes indicated the relationship of zicaitai and the *Brassicaceae* family, which supports zicaitai as a variety of *B. rapa* in taxonomy. The results obtained in this study provide insight into further research on *Brassica* chloroplasts and their phylogeny.

**ARTICLE HISTORY**
Received 9 June 2023
Accepted 9 January 2024

**KEYWORDS**
*Brassica rapa* var. *purpuraria*; chloroplast genome; phylogenetic analysis; SSR

## Introduction

*Brassica rapa* var. *purpuraria* (L.H.Bailey) Kitam, also called zicaitai (Figure 1), is a vegetable variety of *B. rapa* that has been formed after long-term cultivation and domestication in Central China (Li et al. 2019). The variety is known for its unique flavor and high nutritional value, especially the high levels of anthocyanin content in its leaves and stalks (Guo et al. 2015). The comparison between the variety and other *Brassica rapa* plants is important, so several phylogenetic analyses have been carried out based on the chloroplast genomes of *Brassicaceae* species in recent years (Prabhudas et al. 2015; Seol et al. 2017; Kim et al. 2018; Du et al. 2020; Han et al. 2020; Wu et al. 2021; Liu et al. 2022). In this research, we first reported the complete chloroplast genome sequence of zicaitai and relevant phylogenetic analyses were conducted to provide a scientific basis for understanding its taxonomic status.

## Materials and methods

In this research, fresh leaf samples and specimens of zicaitai were collected from Guangzhou, Guangdong, China (N 23°06′, E 113°15′). The specimen was deposited at the Guangzhou Academy of Agricultural Sciences under voucher number V02D0643 (Guangzhou, China; Hailong Ren, renhailong_2006@163.com). Total genomic DNA was extracted from leaf samples using a modified CTAB method (Sahu et al. 2012), and qualified DNA samples were submitted for sequencing using Oxford Illumina NovaSeq 6000 platforms at Genepioneer Biotechnologies Company (Nanjing, China). The DNA insert fragments, about 350 bp in length, were used for library construction. Based on the 2 × 150 bp paired-end sequencing mode, a total of 3.29 G of sequencing data was generated. After removing the adapter sequences and low-quality reads, 19,543,592 bp of clean reads were obtained. The de novo assembly process was carried out using SPAdes v3.10.1 (http://cab.spbu.ru/software/spades/) (Bankevich et al. 2012), with k-mer sizes of 55, 87, and 121. Subsequently, the assembled sequence was annotated using Prodigal v2.6.3 (https://www.github.com/hyattpd/Prodigal) for CDS (coding sequence), HMMER v3.1b2 (http://www.hmmer.org) for rRNA, and Aragorn v1.2.38 (http://130.235.244.92/ARAGORN) for tRNA (Laslett and Canback 2004; Hyatt et al. 2010; Potter et al. 2018), with the chloroplast genome of *B. rapa* subsp. *ra*pa (NCBI accession number: NC_049891) as a reference. MISA (Beier et al. 2017) was employed to identify simple sequence repeats (SSRs) of one to six nucleotide units (mono-, di-, tri-, tetra-, penta-, and hexa-) in the chloroplast genome, with a minimal threshold of 8, 5, 3, 3, 3, and 3 for each unit size, respectively. For phylogenetic tree analysis on the entire genome, the circular sequences were set to start

CONTACT Hailong Ren renhailong_2006@163.com Guangzhou Academy of Agricultural Sciences, Guangzhou, PR China
*These authors contributed equally to this work.

Supplemental data for this article can be accessed online at https://doi.org/10.1080/23802359.2024.2305403.

**Figure 1.** Photograph of *Brassica rapa* var. *purpuraria* (L.H.Bailey) Kitam. (This photograph was taken by Prof. Guangguang Li in Guangzhou City, Guangdong Province). The stem, leaf vein, petiole, and inflorescence axis of *Brassica rapa* var. *purpuraria* are purple. The basal leaves are pinnately divided. The lower stem leaves are triangular-ovate or needle-shaped oblong, and the upper leaves slightly cling to the stem.

from the same point. Multiple sequence alignment between different species was done using the MAFFT software (v7.427) (Katoh and Standley 2013) in auto mode. The aligned data were then analyzed using RAxML v8.2.10 (https://cme.h-its.org/exelixis/software.html) (Stamatakis 2014) with the GTRGAMMA model and rapid bootstrap analysis with a bootstrap of 1000. The maximum-likelihood (ML) evolutionary tree was constructed using the following sequences: *Carica papaya* NC_010323.1 (outgroup), *Camelina anomala* NC_061686.1 (Brock et al. 2022), *Boechera stricta* NC_049599.1, *Capsella bursa-pastoris* NC_009270.1, *Crucihimalaya himalaica* NC_061290.1 (Chen et al. 2022), *Turritis glabra Ohmi-Shirahama* LC325490.1, *Cardamine bipinnata* NC_060865.1 (Ru et al. 2020), *Schrenkiella parvula* NC_028726.1 (He et al. 2016), *Sinapis arvensis* NC_035303.1, *Raphanus sativus* NC_024469.1, *Arabidopsis thaliana* MN603471.1 (Zhang et al. 2020), *B. carinata* NC_059807.1, *B. juncea* NC_028272.1, *B. napus* NC_016734.1 (Hu et al. 2011), *B. oleracea* MN396847.1 (Perumal et al. 2021), *B. rapa* subsp. *rapa* NC_049891.1, *B. nigra* NC_030450.1 (Seol et al. 2017), *B. rapa* subsp. *pekinensis* NC_015139.1, *B. rapa* subsp. *chinensis* KX681648.1, and *B. rapa* var. *purpuraria* OP729397.1.

## Results

The average and minimum read mapping depths of the assembled genome were 5141× and 911×, respectively (Figure S1). A circular map of the chloroplast genome and a schematic map of the cis- and trans-splicing genes (Figure 2 and Figure S2) were visualized by CPGView (Liu et al. 2023). The *zicaitai* chloroplast genome map showed that the chloroplast has a genome of 153,483 bp in length (NCBI accession OP729397), including a large single-copy (LSC) region (83,282 bp), a small single-copy (SSC) region (17,775 bp), and a pair of inverted repeats (IRs) (26,213 bp). The overall GC content of the chloroplast genome is 36.36%. In total, 132 chloroplast genes, including 87 protein-coding genes, 37 tRNA genes and eight rRNA genes, were annotated, and 315 SSRs, including 229 mononucleotides, 17 dinucleotides, 63 trinucleotides, and six tetranucleotides, were identified. These SSRs were located in LSC (197), SSC (72), and IR (46), respectively (Table S1). To investigate the phylogenetic status of *zicaitai* within the genus *Brassica*, the chloroplast genomes of 19 *Brassicaceae* species and one outgroup were subject to phylogenetic tree construction, indicating that *zicaitai* clustered with the clade of *B. rapa* ssp. with full support, which

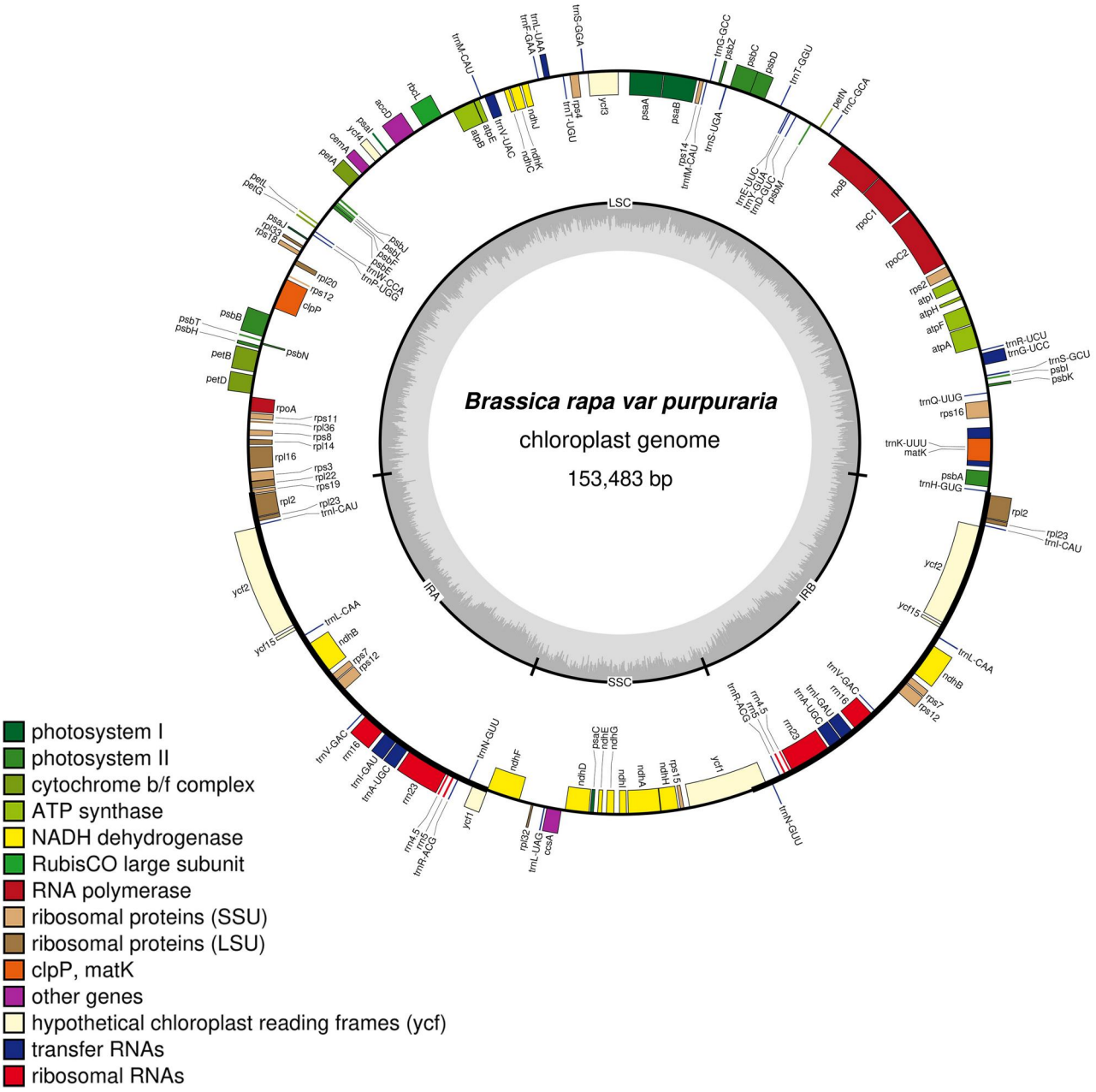


**Figure 2.** The circle map of *Brassica rapa* var. *purpuraria* chloroplast genome. Distinctive colored boxes encircling the outer circle depict genes, represented within the circle are genes transcribed in a clockwise direction, and counter-clockwise transcribed genes are featured outside the circle. The inner circle features a gray region indicating the GC content, while the quadripartite structure (LSC, SSC, IRA, and IRB) is illustrated on the inner circle accordingly. The chloroplast genome was drawn using OGDRAW (https://chlorobox.mpimp-golm.mpg.de/OGDraw.html).

further supports zicaitai as a variety of *B. rapa* in taxonomy (Figure 3).

## Discussion and conclusions

The complete chloroplast genome of *zicaitai* was sequenced and assembled to investigate the phylogenetic relationships between it and 19 other *Brassicaceae* species. The ML analysis revealed that *zicaitai* (OP729397) was closely related to *B. rapa* ssp. and *B. juncea* (NC_028272.1). These findings are consistent with Du et al. (2020), Han et al. (2020), and mitochondrial genome studies by Xue et al. (2020), which suggest that *B. juncea* was maternally derived from *B. rapa*, thus confirming *zicaitai* as a variety of *B. rapa*. This may serve as a valuable genetic resource for taxonomic and related studies of *Brassica*.

## Ethical approval

The authors declare no ethical or legal violations when obtaining the study materials and performing the research. The species used in this study is not listed on the IUCN Red List, and the sample was legally collected by guidelines stipulated in national and international regulations. The materials were collected in a location not designated as a protected area in China.

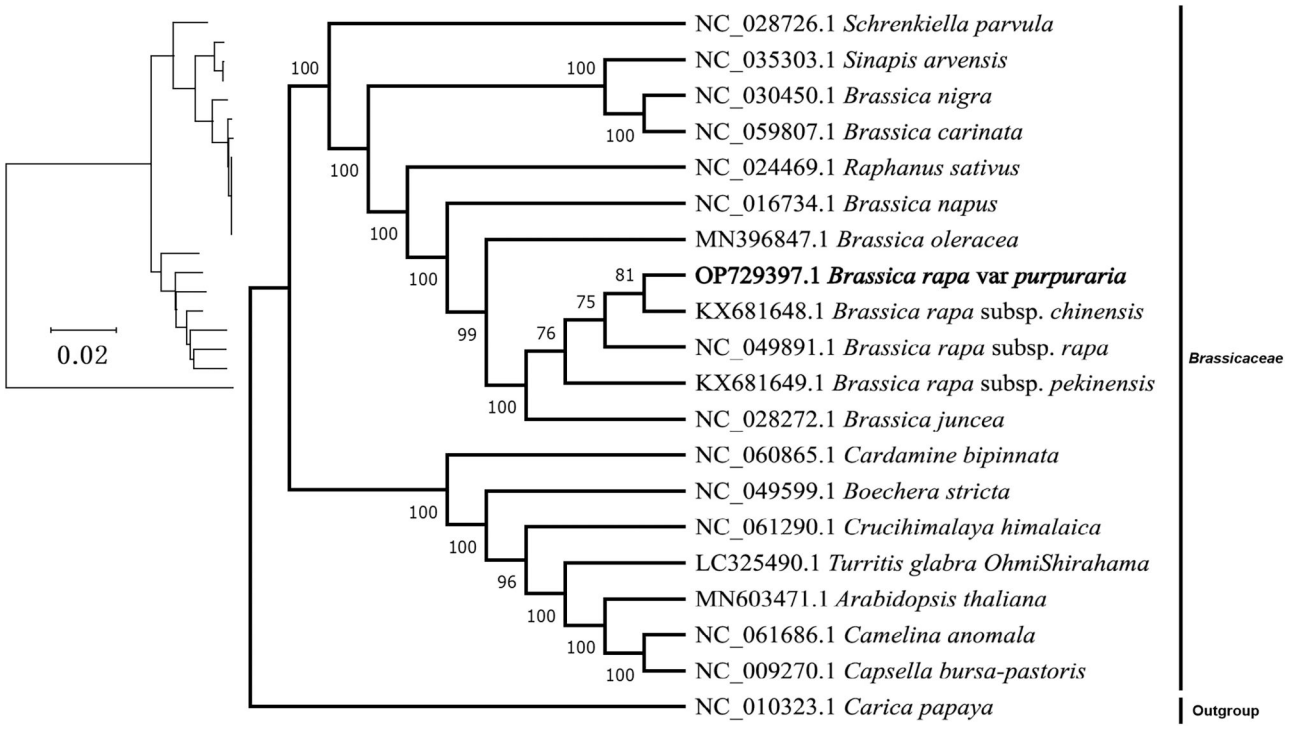


**Figure 3.** The maximum-likelihood phylogenetic tree of 19 *Brassicaceae* species was constructed based on genome sequences, with *Carica papaya* added as an outgroup. The phylogenetic tree was constructed using the maximum-likelihood (ML) method and bootstrap was performed 1000 times. The number on each branch indicates the boot support value. The following sequences were used: *Carica papaya* NC_010323.1 (outgroup), *Camelina anomala* NC_061686.1 (Brock et al. 2022), *Boechera stricta* NC_049599.1, *Capsella bursa-pastoris* NC_009270.1, *Crucihimalaya himalaica* NC_061290.1 (Chen et al. 2022), *Turritis glabra Ohmi-Shirahama* LC325490.1, *Cardamine bipinnata* NC_060865.1 (Ru et al. 2020), *Schrenkiella parvula* NC_028726.1 (He et al. 2016), *Sinapis arvensis* NC_035303.1, *Raphanus sativus* NC_024469.1, *Arabidopsis thaliana* MN603471.1 (Zhang et al. 2020), *Brassica carinata* NC_059807.1, *Brassica juncea* NC_028272.1, *Brassica napus* NC_016734.1 (Hu et al. 2011), *Brassica oleracea* MN396847.1 (Perumal et al. 2021), *Brassica rapa* subsp. *ra*pa NC_049891.1, *Brassica nigra* NC_030450.1 (Seol et al. 2017), *Brassica rapa* subsp. *pekinensis* NC_015139.1, *Brassica rapa* subsp. *chinensis* KX681648.1, and *B. rapa* var. *purpuraria* OP729397.1.

## Author contributions

Xianyu Zhou and Hailong Ren are mainly responsible for the data analysis and writing of the manuscript; Hailong Ren, Jing Zhang, Hua Zhang, and Yansong Zheng are responsible for providing guidance and participating in the related data analysis; Donglin Xu, Wanyu Xiao, Hongdi Huang, and Guangguang Li have participated in the analysis and drafting of the paper, as well as revised it critically for intellectual content. All authors approved the final manuscript.

## Disclosure statement

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

## Funding

The research was supported by the Seed Industry Revitalization Project of the Provincial Rural Revitalization Strategy Special Fund in 2022 [nos. 2022-NPY-03-001, 2022-NJS-03-001]

## Data availability statement

The genome sequence data that support the findings of this study are openly available in GenBank of NCBI at https://www.ncbi.nlm.nih.gov/ under the accession no. OP729397. The associated BioProject, SRA, and Bio-Sample numbers are PRJNA971340, SRR24517193, and SAMN35028576, respectively.

## References

Bankevich A, Nurk S, Antipov D, Gurevich AA, Dvorkin M, Kulikov AS, Lesin VM, Nikolenko SI, Pham S, Prjibelski 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.

Beier S, Thiel T, Münch T, Scholz U, Mascher M. 2017. MISA-web: a web server for microsatellite prediction. Bioinformatics. 33(16):2583–2585. doi:10.1093/bioinformatics/btx198.

Brock JR, Mandáková T, McKain M, Lysak MA, Olsen KM. 2022. Chloroplast phylogenomics in Camelina (*Brassicaceae*) reveals multiple origins of polyploid species and the maternal lineage of *C. sativa*. Hortic Res. 9:uhab050. doi:10.1093/hr/uhab050.

Chen HL, Al-Shehbaz IA, Qian LS, Zhang JW, Xu B, Zhang TC, Yue JP, Sun H. 2022. *Pulvinatusia* (Brassicaceae), a new cushion genus from China and its systematic position. PhytoKeys. 189:9–28. doi:10.3897/phytokeys.189.77926.

Du X, Zeng T, Feng Q, Hu L, Luo X, Weng Q, He J, Zhu B. 2020. The complete chloroplast genome sequence of yellow mustard (*Sinapis alba* L.) and its phylogenetic relationship to other *Brassicaceae* species. Gene. 731:144340. doi:10.1016/j.gene.2020.144340.

Guo N, Wu J, Zheng S, Cheng F, Liu B, Liang J, Cui Y, Wang X. 2015. Anthocyanin profile characterization and quantitative trait locus mapping in zicaitai (*Brassica rapa* L. ssp. *chinensis* var. *purpurea*). Mol Breed. 35(5):113. doi:10.1007/s11032-015-0237-1.

Han R, Tian M, Zhan G, Shao D, Ren Y. 2020. Complete chloroplast genome sequence of turnip (*Brassica rapa*. ssp. *rapa*): genome structure and phylogenetic analysis. Mitochondrial DNA B Resour. 5(3):3555–3557. doi:10.1080/23802359.2020.1829124.

He Q, Hao G, Wang X, Bi H, Li Y, Guo X, Ma T. 2016. The complete chloroplast genome of *Schrenkiella parvula* (Brassicaceae). Mitochondrial DNA A DNA Mapp Seq Anal. 27(5):3527–3528. doi:10.3109/19401736.2015.1074202.

Hu ZY, Hua W, Huang SM, Wang HZ. 2011. Complete chloroplast genome sequence of rapeseed (*Brassica napus* L.) and its evolutionary implications. Genet Resour Crop Evol. 58(6):875–887. doi:10.1007/s10722-010-9626-9.

Hyatt D, Chen GL, Locascio PF, Land ML, Larimer FW, Hauser LJ. 2010. Prodigal: prokaryotic gene recognition and translation initiation site identification. BMC Bioinformatics. 11(1):119. doi:10.1186/1471-2105-11-119.

Katoh K, Standley DM. 2013. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. Mol Biol Evol. 30(4):772–780. doi:10.1093/molbev/mst010.

Kim CK, Seol YJ, Perumal S, Lee J, Waminal NE, Jayakodi M, Lee SC, Jin S, Choi BS, Yu Y, et al. 2018. Re-exploration of U's triangle *Brassica* species based on chloroplast genomes and 45S nrDNA sequences. Sci Rep. 8(1):7353. doi:10.1038/s41598-018-25585-4.

Laslett D, Canback B. 2004. ARAGORN, a program to detect tRNA genes and tmRNA genes in nucleotide sequences. Nucleic Acids Res. 32(1):11–16. doi:10.1093/nar/gkh152.

Li GH, Chen HC, Liu JL, Luo WL, Xie DS, Luo SB, Wu TQ, Akram W, Zhong YJ. 2019. A high-density genetic map developed by specific-locus amplified fragment (SLAF) sequencing and identification of a locus controlling anthocyanin pigmentation in stalk of Zicaitai (*Brassica rapa* L. ssp. *chinensis* var. *purpurea*). BMC Genomics. 20(1):343. doi:10.1186/s12864-019-5693-2.

Liu H, Zhao W, Hua W, Liu J. 2022. A large-scale population based organelle pan-genomes construction and phylogeny analysis reveal the genetic diversity and the evolutionary origins of chloroplast and mitochondrion in *Brassica napus* L. BMC Genomics. 23(1):339. doi:10.1186/s12864-022-08573-x.

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.

Perumal S, Waminal NE, Lee J, Koo HJ, Choi BS, Park JY, Ahn K, Yang TJ. 2021. Nuclear and chloroplast genome diversity revealed by low-coverage whole-genome shotgun sequence in 44 *Brassica oleracea* breeding lines. Hortic Plant J. 7(6):539–551. doi:10.1016/j.hpj.2021.02.004.

Potter SC, Luciani A, Eddy SR, Park Y, Lopez R, Finn RD. 2018. HMMER web server: 2018 update. Nucleic Acids Res. 46(W1):W200–W204. doi:10.1093/nar/gky448.

Prabhudas SK, Raju B, Thodi SK, Parani M, Natarajan P. 2015. The complete chloroplast genome sequence of Indian mustard (*Brassica juncea* L.). Mitochondrial DNA A DNA Mapp Seq Anal. 27(6):4622–4623. doi:10.3109/19401736.2015.1101586.

Ru Y, Schulz R, Koch MA. 2020. Successful without sex – the enigmatic biology and evolutionary origin of coralroot bittercress (*Cardamine bulbifera*, Brassicaceae). Perspect Plant Ecol Evol Syst. 46:125557. doi:10.1016/j.ppees.2020.125557.

Sahu SK, Thangaraj M, Kathiresan K. 2012. DNA extraction protocol for plants with high levels of secondary metabolites and polysaccharides without using liquid nitrogen and phenol. ISRN Mol Biol. 2012:205049. doi:10.5402/2012/205049.

Seol YJ, Kim K, Kang SH, Perumal S, Lee J, Kim CK. 2017. The complete chloroplast genome of two *Brassica* species, *Brassica nigra* and *B. Oleracea*. Mitochondrial DNA A DNA Mapp Seq Anal. 28(2):167–168. doi:10.3109/19401736.2015.1115493.

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.

Wu JY, Ma XC, Ma L, Fang Y, Zhang YH, Liu LJ, Li XC, Zeng R, Sun WC. 2021. Complete chloroplast genome sequence and phylogenetic analysis of winter oil rapeseed (*Brassica rapa* L.). Mitochondrial DNA B Resour. 6(3):723–731. doi:10.1080/23802359.2020.1860697.

Xue JY, Wang Y, Chen M, Dong S, Shao ZQ, Liu Y. 2020. Maternal inheritance of U's triangle and evolutionary process of *Brassica* mitochondrial genomes. Front Plant Sci. 11:805. doi:10.3389/fpls.2020.00805.

Zhang M, Zhao J, Li W, Wen S, Huang H, Dong J, Liu B, Zhang G, Wang H-B, Shen Y, et al. 2020. Increased photosystem ii translation efficiency as an important photoprotective mechanism in an *Arabidopsis thaliana* ecotype (tibet-0) adapted to high light environments. Environ Exp Bot. 183:104350. doi:10.1016/j.envexpbot.2020.104350.