CHARACTERIZATION OF THE COMPLETE CHLOROPLAST GENOME OF RORIPPA DUBIA (PERSOON) H. HARA（BRASSICACEAE）

Yang Bingzhi[1]*, Yang Wang[1], Siying Li[2]

[1]Traditional Chinese Medicine, Lijiang Culture and Tourism College, Lijiang, Yunnan, China
[2]Traditional Chinese Medicine, Dali University, Dali, Yunnan, China
*Corresponding Author: Yang Bingzhi

ABSTRACT

We employed high-throughput sequencing to decode the entire chloroplast (cp) genome of the medicinal annual herb *Rorippa dubia*. The genome is 154,735 bp in total length and exhibits a typical quadripartite structure, including a large single-copy (LSC) region of 83,736 bp, a small single-copy (SSC) region of 18,005 bp, and a pair of inverted repeats (IRs) of 26,497 bp. The overall guanine-cytosine (GC) content is 34.60%. The genome contains 130 genes, including 85 protein-coding genes, 37 tRNA genes, and 8 rRNA genes. Phylogenetic analysis based on 27 complete chloroplast (cp) genomes from Brassicaceae, using the maximum likelihood (ML) method, showed that *R. dubia* formed a monophyletic group with other *Rorippa* species, strongly supporting its taxonomic position within the genus. As the inaugural report of a complete plastome for R. *dubia*, our findings establish an essential genomic framework that will facilitate subsequent investigations into the phylogeny and population genetics of the *Rorippa* genus.

KEYWORDS: *Rorippa dubia*; Chloroplast genome; Brassicaceae; Phylogenetic relationship

1. INTRODUCTION

Rorippa dubia (Persoon) H. Hara, 1955, an annual herb of the Brassicaceae family, is widely distributed across moist habitats such as mountain roadsides, wetlands, and riverbanks, spanning altitudes from 500 m to 3700 m, which reflects its remarkable environmental adaptability (Yan BY. 2023). As a medicinal and edible plant, *R. dubia* has long been used in traditional Chinese medicine. Its major bioactive compounds, including rorifone and roriamide, have been reported to exhibit antitussive, expectorant, and antibacterial activities, as well as antihypertensive, diuretic, and hemostatic effects (Liu GM, et al. 2024). Clinically, preparations derived from *R. dubia* are used to treat cough, asthma, fever, hypertension, and inflammation. In addition, its tender shoots are commonly consumed as a nutritious wild vegetable rich in vitamins and minerals.

Despite its diverse applications, research on *R. dubia* has been predominantly focused on phytochemical extraction and pharmacological validation. However, molecular-level investigations, such as a comprehensive analysis of its chloroplast genome, remain underdeveloped. The chloroplast genome, characterized by its conserved structure, moderate mutation rate, and predominantly maternal inheritance, serves as a powerful tool for elucidating plant phylogeny, species identification, and genetic diversity (Li et al., 2024).

Here, we reported the complete chloroplast genome of *R. dubia* using high-throughput sequencing. We characterize its structural organization, gene content, and GC composition, and perform phylogenetic analysis with related Brassicaceae species. This work provides the first genomic resource for *R. dubia*, offering molecular insights for taxonomy, phylogenetic reconstruction, and future utilization of *Rorippa* genetic resources.

2. MATERIAL AND METHODS

Plant samples of *Rorippa dubia* were collected from Shimen Town, Yunlong County, Dali Prefecture, Yunnan Province, China (99°46′ E, 26°23′ N). The species was authenticated by Dr. Fan Min, an expert in medicinal plant resources. A voucher specimen (accession number HC2024072801) has been deposited in the Herbarium of Lijiang Cultural Tourism College (Contact: Bingzhi Yang, yangbinghzi@163.com) and is represented in Figure 1.

Figure 1. *R. dubia* (Persoon) H. Hara. (A) Whole plant. (B) Fruit. (C) Flower. (D) Flower, fruit, and leaf.The plant pictures were taken by the author.

We extracted genomic DNA from fresh leaves using a modified cetyltrimethylammonium bromide (CTAB) method, after immediate snap-freezing in liquid nitrogen. We then sent the high-quality DNA to Novogene Co., Ltd. (Beijing, China) for library construction and high-throughput sequencing.

To ensure data quality, raw Illumina reads underwent an initial filtering process with Trimmomatic v0.39 (Bolger et al., 2014) to remove adapter sequences, low-quality bases (Phred score < 20), and reads containing more than 5% ambiguous nucleotides. Clean reads were then assessed using FastQC v0.11.9 (Andrews, 2010) to confirm data reliability. The chloroplast genome was assembled *de novo* using GetOrganelle with k-mer values set to 21, 77, and 127 (Jin et al., 2020). To evaluate assembly completeness, clean reads were realigned to the assembled genome using Bowtie2, and coverage depth was calculated with SAMtools v1.15 (Danecek et al., 2021). Uniform and high-depth coverage confirmed the integrity and quality of the chloroplast genome assembly.

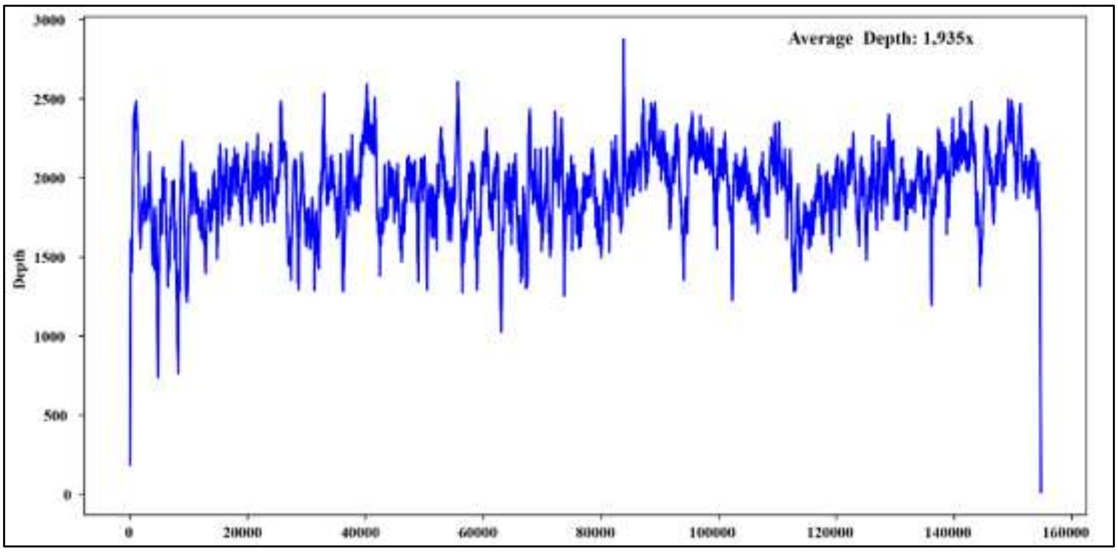
Genome annotation was performed using CPGAVAS2 (Shi et al., 2019; Wang et al., 2024), with the chloroplast genome of *R. palustris* (GenBank accession OR124737.1) as a reference. Annotation accuracy was further refined using CPStools (Huang et al., 2024) by adjusting start/stop codons and exon-intron boundaries. A circular genome map illustrating the quadripartite structure, gene organization, and GC content was generated using CPGView (Liu et al., 2023).

To determine the phylogenetic position of *R. dubia*, we analyzed its complete chloroplast genome alongside 26 Brassicaceae chloroplast genomes retrieved from genetic sequence database (GenBank). The 27 sequences were aligned using multiple alignment using fast fourier transform (MAFFT) v7.149 (Katoh et al., 2017), and the optimal substitution model was selected with ModelFinder (Kalyaanamoorthy et al., 2017). Phylogenetic inference was conducted under the maximum-likelihood criterion using randomized axelerated maximum likelihood (RAxML) (Stamatakis, 2014), with branch support evaluated from 1000 bootstrap replicates.

3. RESULTS

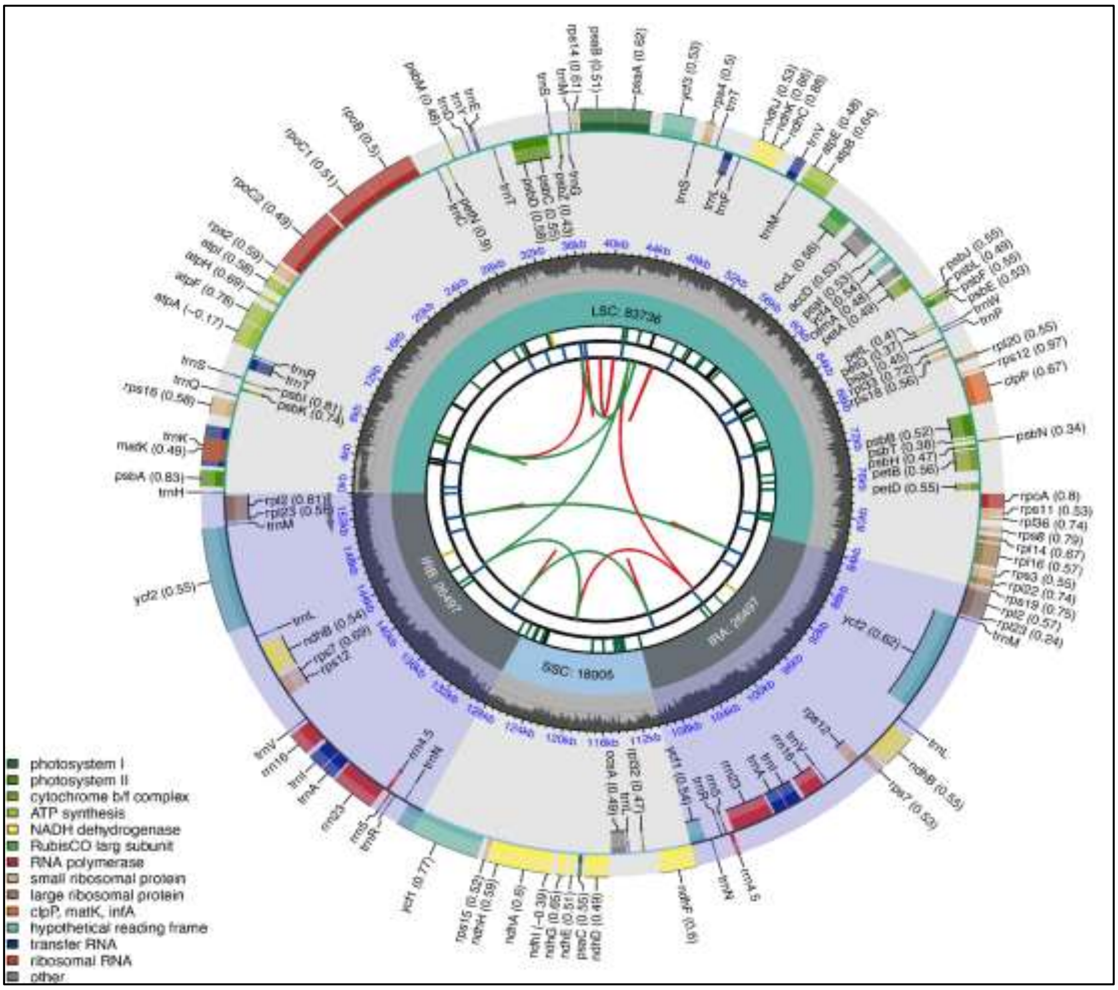
Coverage assessment showed that the chloroplast genome of *R. dubia* was assembled with uniform and high depth (ranging from ~270× to >500×, with an average depth of ~350×) and without gaps, indicating a reliable assembly (Figure 2).

Figure 2. Coverage depth distribution across the complete chloroplast genome of *Rorippa dubia*. The uniform and high-depth coverage indicates a high-quality and complete genome assembly.



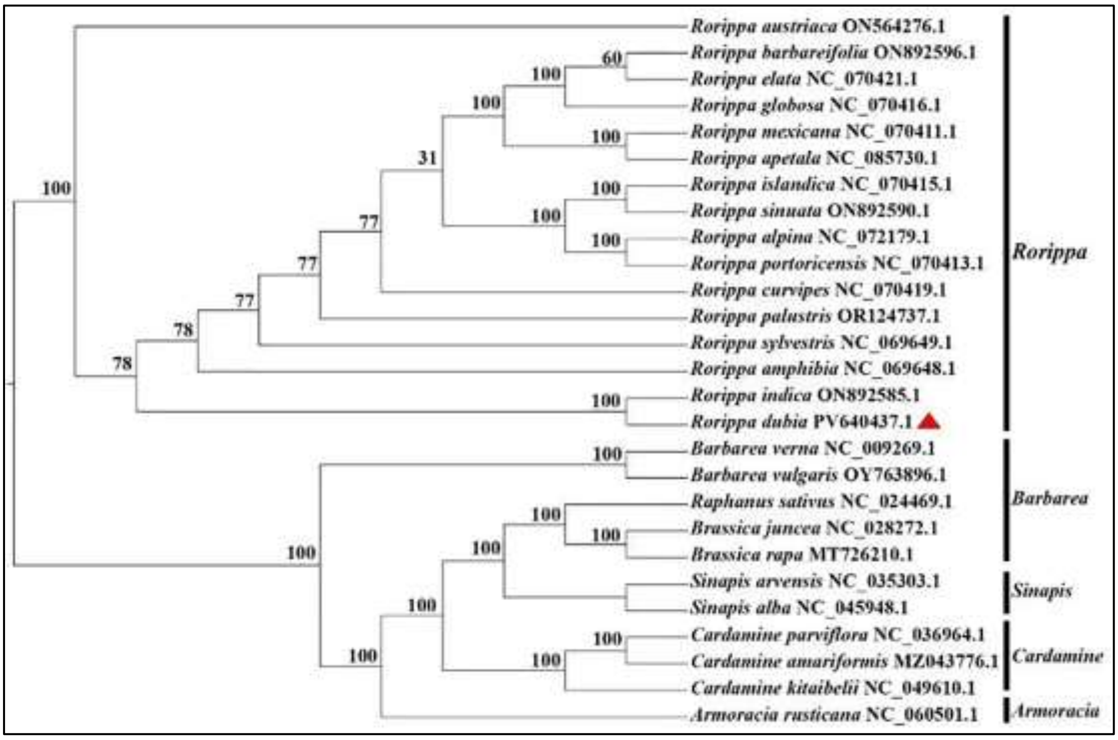
The assembled plastome of *R. dubia*has a total length of 154,735 bp (Figure 3) and displays the characteristic circular quadripartite architecture common to most angiosperms. This structure comprises a large single-copy (LSC) region of 83,736 bp and a small single-copy (SSC) region of 18,005 bp,separated by a pair of inverted repeat (IR) regions, each 26,497 bp in length.

Figure 3. Circular map of the *Rorippa dubia* chloroplast genome. Genes are color-coded by functional category, with inward- and outward-facing transcripts indicating clockwise and counter-clockwise transcription, respectively. The inner gray-scale track represents the GC content, and the boundaries of the large single-copy (LSC), small single-copy (SSC), and inverted repeat (IR) regions are indicated.



In total, 130 genes were annotated, including 85 protein-coding genes, 37 tRNA genes and 8 rRNA genes, with an overall GC content of 34.60%; GC was highest in the IRs (42.10%), followed by the LSC (32.20%) and SSC (28.50%), reflecting rRNA enrichment in the IRs. Boundary analysis further showed that the LSC/IR junction is located between *rps19* and *trnH*, and the SSC/IR junction between *ndhF* and *ycf1*, consistent with the typical Brassicaceae cp-genome organization. The ML tree based on 27 Brassicaceae cp genomes (Figure 4) recovered well-supported genus-level clades (mostly >90% bootstrap), with *R. dubia* forming a strongly supported monophyletic *Rorippa* clade (bootstrap support, BS = 100%) sister to other genera (e.g. *Cardamine, Lepidium*), and showing closer affinity to aquatic/wetland *Rorippa* species, suggesting possible habitat-driven diversification.

Figure 4. Maximum-likelihood tree inferred from complete chloroplast genome sequences, depicting the phylogenetic position of *Rorippa dubia* among 26 representative Brassicaceae species. *Armoracia rusticana* was used as the outgroup. Bootstrap support values from 1000 replicates are displayed on the branches, and corresponding GenBank accession numbers are listed for each species.



The following sequences were used: *R. dubia* (PV640437, this study), *R. palustris* (OR124737), *R. indica* (ON892585), *R. amphibia* (NC_069648), *R. sylvestris* (NC_069649), *R. curvipes* (NC_070419), *R. islandica* (NC_070415), *R. alpina* (NC_072179), *R. sinuate* (ON892590), *R. austriaca* (ON564276), *R. barbareifolia* (ON892596), *R. elata* (NC_070421), *R. globosa* (NC_070416), *R. mexicana* (NC_070411), *R. apetala* (NC_085730), *R. portoricensis* (NC_070413), Barbarea verna (NC_009269), *B. vulgaris* (OY763896), *R. sativus* (NC_024469), *B. juncea* (NC_028272), *B. rapa* (MT726210), *S. arvensis* (NC_035303), *S. alba* (NC_045948), *C. parviflora* (NC_036964), *C. amaraeformis* (MZ043776), *C. kitaibelii* (NC_049610), and *A. rusticana* (NC_060501).

4. DISCUSSION

Our work presents the initial characterization of the complete chloroplast genome for *Rorippa dubia*, thereby elucidating its structural organization and evolutionary relationships within the genus.. Analysis of the *R. dubia*plastome confirms a conserved quadripartite structure. Notably, both the total gene count (130) and the overall GC content (34.60%) align with the typical values observed across other Brassicaceae species, underscoring the genomic conservation within this plant family. And its gene content, intron patterns, and IR/SC junctions are highly conserved, reflecting the evolutionary stability of angiosperm plastomes. At the same time, *R. dubia* is slightly larger than the published cp genome of *R. palustris*, mainly due to an extended IR region, and it shows a marginally higher GC content in the IRs, which may serve as a potential molecular marker for species discrimination within *Rorippa.*

Phylogenetic analysis based on 27 Brassicaceae cp genomes recovered *R. dubia* in a fully supported monophyletic *Rorippa* clade, in agreement with its morphology-based taxonomy, and helps resolve taxa previously ambiguous in nuclear-gene-based studies. Its closer relationship with aquatic/wetland *Rorippa* species further suggests that habitat adaptation may have contributed to lineage divergence and deserves functional verification.

The plastome reported here provides an important genomic resource for *Rorippa* and can be used to develop chloroplast simple sequence repeat (cpSSR) / single nucleotide polymorphism (SNP) markers for population genetics and conservation, and to inform future breeding or germplasm assessment of this medicinal-edible species. Overall, this work enriches plastome data for Brassicaceae and provides a robust molecular basis for the taxonomy, evolutionary inference, and utilization of *Rorippa*.

Abbreviations

cp, chloroplast
LSC, large single-copy region
SSC, small single-copy region
IRs, inverted repeats
GC, guanine-cytosine
ML, maximum likelihood
CTAB, cetyltrimethylammonium bromide
MAFFT, multiple alignment using fast fourier transform
RAxML, randomized axelerated maximum likelihood
BS, bootstrap support
cpSSR, chloroplast simple sequence repeat
SNP, single nucleotide polymorphism

Ethical Approval

This study was conducted in compliance with the Regulations of the People's Republic of China on the Protection of Wild Plants and relevant institutional guidelines. The collection of *Rorippa dubia* plant material, a non-protected species, was performed without causing harm to the local ecosystem.

Disclosure Statement

The authors declare no competing financial or personal interests that could influence the work reported in this paper.

Funding

This work was supported by the Scientific Research Fund Project of the Yunnan Provincial Department of Education (Project No. 2024J1285).

Data Availability Statement

We have deposited the complete chloroplast genome sequence of *Rorippa dubia* in the NCBI GenBank (accession: PV640437). The corresponding raw sequence data are available under the SRA BioProject PRJNA1267994.

Author Contributions

Yang Bingzhi: Conceptualization, Investigation, Formal analysis, Writing – original draft.
Wang Yang: Data curation, Visualization.
Li Siying: Methodology, Supervision, Validation, Writing – review & editing.
All authors read and approved the final manuscript.

Acknowledgments

We sincerely thank Dr. Fan Min for his expertise in the identification of the plant material. We are grateful to Novogene Co., Ltd. (Beijing, China) for their assistance with high-throughput sequencing. We also acknowledge the technical support provided by the bioinformatics platform at our institutions. This research was funded by the Scientific Research Fund Project of the Yunnan Provincial Department of Education.

REFERENCES

1. Andrews, S. (2010). FastQC: A Quality Control tool for High Throughput Sequence Data. Available at: https://www.bioinformatics.babraham.ac.uk/projects/fastqc/. Accessed March 17, 2014.
2. Bolger, A.M., Lohse, M., and Usadel, B. (2014). Trimmomatic: a flexible trimmer for Illumina sequence data. Bioinformatics 30(15), 2114-2120.
3. Danecek, P., Bonfield, J.K., Liddle, J., Marshall, J., Ohan, V., Pollard, M.O., et al. (2021). Twelve years of SAMtools and BCFtools. GigaScience 10(2), giab008.

4. Huang, L., Yu, H., Wang, Z., and Xu, W. (2024). CPStools: A package for analyzing chloroplast genome sequences. iMetaOmics 1, e25.
5. Jin, J.J., Yu, W.B., Yang, J.B., Song, Y., dePamphilis, C.W., Yi, T.S., et al. (2020). GetOrganelle: a fast and versatile toolkit for accurate de novo assembly of organelle genomes. Genome Biol. 21(1), 241.
6. Kalyaanamoorthy, S., Minh, B.Q., Wong, T.K.F., von Haeseler, A., and Jermiin, L.S. (2017). ModelFinder: fast model selection for accurate phylogenetic estimates. Nat. Methods 14(6), 587-589.
7. Katoh, K., Rozewicki, J., and Yamada, K.D. (2017). MAFFT online service: multiple sequence alignment, interactive sequence choice and visualization. Brief. Bioinform. 20(4), 1160-1166.
8. Li, S., Wang, S., and Fan, M. (2024). The complete chloroplast genome of Fuchsia standishii J. Harrison, 1840 (Onagraceae) from Yunnan, China. Mitochondrial DNA B Resour. 9(10), 1370-1373.
9. Liu, G.M., Guo, S.H., Sun, X., et al. (2024). [Study on modified CTAB method for total RNA extraction from different crops]. Chin. Agric. Sci. Bull. 40(11), 28-35.
10. Liu, S., Ni, Y., Li, J., Zhang, X., Yang, H., Chen, H., et al. (2023). CPGView: A package for visualizing detailed chloroplast genome structures. Mol. Ecol. Resour. 23(3), 694-704.
11. Shi, L., Chen, H., Jiang, M., Wang, L., Wu, X., Huang, L., et al. (2019). CPGAVAS2, an integrated plastome sequence annotator and analyzer. Nucleic Acids Res. 47(W1), W65-W73.
12. Stamatakis, A. (2014). RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. Bioinformatics 30, 1312-1313.
13. Wang, X., Wang, L., Kong, J., Li, H., and Kong, N. (2024). The complete plastome of Rorippa palustris Besser 1821 and its phylogenetic analysis. Mitochondrial DNA B Resour. 9(9), 1273-1277.
14. Yan, B.Y. (2023). [Effects of community species diversity and different strata on the invasiveness of two Rorippa species with root tillers]. Doctoral dissertation, Shenyang Agricultural University.