BMC Research Notes

DATA NOTE

Open Access

First report of the chloroplast and mitochondrial genomes of the Indian pitcher plant, *Nepenthes khasiana* Hook.f.

Ruchishree Konhar[1,2], Debasis Dash[1,2] and Devendra Kumar Biswal[3*]

Abstract

Objectives *Nepenthes,* sometimes known as tropical pitcher plants or monkey cups, is a carnivorous plant genus that contains more than 160 species. *Nepenthes khasiana,* India's sole representative of the genus, is a rare and endangered dioecious plant endemic to North-east India. Despite the fact that it is a prominent insectivorous plant in the Nepenthaceae family, genomic resources for the species are limited, making genomic breeding and understanding the genetic basis of botanical carnivory difficult. Herein, we report the complete chloroplast (cp) and mitochondrial (mt) genomes of *N. khasiana* for the first time. These organelle genomes were assembled as part of a whole-genome sequencing project aimed at gaining deeper insights into their evolutionary relations with genomes of other carnivorous plants.

Data description The complete cp genome (156,914 bp) and mt genome (900,031 bp) of *N. khasiana* are presented here. The cp genome contains two repeat regions and 131 genes (112 unique genes): 86 protein coding genes, 8 rRNA coding genes and 37 tRNA coding genes. The mt genome contains 84 genes (55 unique genes): 50 protein coding genes, 7 rRNA coding genes and 27 tRNA coding genes. The cp and mt genomic data generated will be useful for future molecular characterization and evolutionary research related to botanical carnivory.

Keywords Nepenthes khasiana, Mitochondrial genome, Chloroplast genome, Botanical carnivory, De novo assembly

Objective

Because of their distinctive adaptations, carnivorous plants have become a valuable model system for studying a variety of ecological and evolutionary issues. The cone-shaped leaves of carnivorous pitcher plants (CPPs) facilitate in trapping prey and deriving nutrients. The Indian pitcher plant *Nepenthes khasiana,* is endemic to West and South Garo Hills, West and East Khasi Hills and Jaintia Hills of Meghalaya, India [1]. It is categorized as endangered on the IUCN Red List of Threatened Species and is included in the Government of India's Negative List of Exports [2]. It is also included in Appendix I of CITES (the Convention on International Trade in Endangered Species of Wild Fauna and Flora), which bans trade of this species [3]. The fluid from unopened pitcher has been in use by the local tribes for treating cataract and night blindness, skin infections, gastrointestinal and gynecological problems. The juice from leaf is used for treating diabetes and urinary disorders [4, 5]. The different adaptation strategies of pitcher plants serve as a source of inspiration for a wide range of applications, including therapeutic treatments, reagents for

*Correspondence:
Devendra Kumar Biswal
devbioinfo@gmail.com
[1] CSIR-Institute of Genomics and Integrative Biology (CSIR-IGIB), Mathura Road, Delhi 110025, India
[2] Academy of Scientific and Innovative Research (AcSIR), Ghaziabad 201002, India
[3] Bioinformatics Center, North-Eastern Hill University, Shillong 793022, India

proteomics experiments and as biocontrol solutions in agriculture [6–8].

The genus *Nepenthes* contains more than 160 species and five incompletely diagnosed taxa [9]. The physical environment and plant-animal interactions are important drivers of species diversity. The genus *Nepenthes*, owing to its high diversity in certain biogeographic regions presents a challenge to taxonomists, who are still searching for reliable methods of characterizing members of this genus by applying species concept. The CPPs have a unique combination of biochemistry, morphology and physiology to enable prey capture and nutrient assimilation from digested prey [10]. We still do not fully understand if these adaptations are also reflected at the genome level. Previous studies have indicated that the chloroplast genomes of carnivorous plants have undergone distinct evolutionary routes compared to that of its non-carnivorous relatives [11]. Nevertheless, specifics like the extent of divergence and general characteristics remain unclear [12]. In this data note, we report the first complete chloroplast (cp) and mitochondria (mt) genomes of *N. khasiana*, which will be crucial for understanding its genetic diversity, and unique evolutionary adaptations.

Data description

Nepenthes khasiana leaf sample was collected from its natural habitat in Jaintia hills, Meghalaya. Plants were identified by Dr. Devendra Kumar Biswal. The Government of Meghalaya granted the necessary authorization to gather samples for research. An herbarium specimen was prepared and submitted to the herbarium at Eastern Regional Centre of Botanical Survey of India located in Shillong, Meghalaya, India where it was assigned the accession number 86834. Using Qiagen Genomic-tip 100/G, high-quality gDNA was extracted from the *N. khasiana* leaf sample; Agarose gel electrophoresis, Nanodrop (Thermo Fisher Scientific, NanoDrop One) and Qubit Fluorometer (Thermo Fisher Scientific, Qubit 3.0) were used to confirm the quantity and quality of the resulting gDNA. The paired-end sequencing library was prepared using TruSeq Nano DNA Library Prep Kit for Illumina. The mean fragment size of shotgun libraries ranged from 586 to 612 bp. The libraries were sequenced on Illumina NextSeq 500 using 2×150 bp chemistry, which generates two reads of 150 base pairs per DNA fragment, one in the forward direction and the other in reverse. After preprocessing i.e., removing low quality reads (reads with N > 5%) and adapter sequences, the filtered reads were used for assembling the organelle genomes. The cp and mt genomes were assembled as part of a whole genome sequencing project for *N. khasiana*, aimed at enhancing our understanding of the evolutionary relations among the genomes of carnivorous plants.

For the chloroplast genome assembly, *rbcL* gene sequence from *N. khasiana* was used as seed sequence for the de novo assembly with NOVOPlasty [13]. When the assembler was run, the K-mer length, i.e., the amount of overlap between matching reads was set to 39. The assembled cp genome has a length of 156,914 bp, with 37% GC content. It has a quadripartite structure that consisted of two inverted repeat regions of 25,193 bp each, one large single-copy region of 87,237 bp and one small single-copy region of 19,291 bp. The assembled genome was annotated using CpGAVAS [14], and GeSeq [15]. A total of 86 protein coding genes of which 79 are unique, 37 tRNAs of which 29 are unique and 8 rRNAs of which four are unique were annotated in the assembled cp genome. Twelve genes had introns: *rps16, atpF, rpoc1, ycf3, clpP, ndhB, ccsA, ndhA* and both copies of *rps12* and *rpl2*. Repeat regions harbored 20 genes: *ndhB*, *rpl2*, *rpl23*, *rps19*, *rps7*, *rrn16*, *rrn23*, *rrn4.5*, *rrn5*, *trnA-UGC*, *trnH-GUG* , *trnI-CAU*, *trnI-GAU*, *trnL-CAA*, *trnN-GUU*, *trnR-ACG*, *trnV-GAC*, *ycf1*, *ycf2* and *rps12*. The annotated cp genome (Fig. 1A) has been submitted to GenBank with accession number: MH923233. Furthermore, NCBI RefSeq Accession number NC_051455 has been added to this submitted cp genome. (Data file 1, Table 1).

For the mitochondrial genome assembly, *matR* gene sequence of *N. mirabilis* was used as seed sequence for assembly using NOVOPlasty with K-mer length set to 39 and genome range of 20,000–2200000 bp to prevent premature circularization. The assembled mt genome was annotated using MITOFY [16] and BLAST (blastn, blastp and tblastx) [17]. The annotations generated by MITOFY were manually curated to check the genomic coordinates for start and stop codons. Manual curation is necessitated as automated annotation tools sometimes fail to identify the entire coding region leading to the annotation of a coding sequence as "misc_features". Homology tools like BLAST were used to verify if complete coding region had been annotated. This approach reduced instances of partial annotations. The assembled mt genome had a length of 900,001bp with 43.8% GC content. The mt genome features a total of 84 genes, with 55 unique genes. The annotated genes include 50 protein coding genes (36 unique), 27 tRNA genes (16 unique) and 7 rRNA genes (3 unique). It has two copies of twelve protein coding genes: *atp9, cob, cox2, nad2, nad3, nad5, nad6, nad9, rpl5, rps1, rps12, rps4*; two rRNA genes: *rrn18* and *rrn5*; and 9 tRNA genes: *trnC*, *trnD-GTC*, *trnE-TTC*, *trnF-GAA*, *trnM-CAT*, *trnP-TGG*, *trnS-TGA*, *trnW-CCA* and *trnY-GTA* . The assembled and annotated mt genome (Fig. 1B) of *N. khasiana* is deposited in GenBank with accession number MK330891 (Data file 2, Table 1).

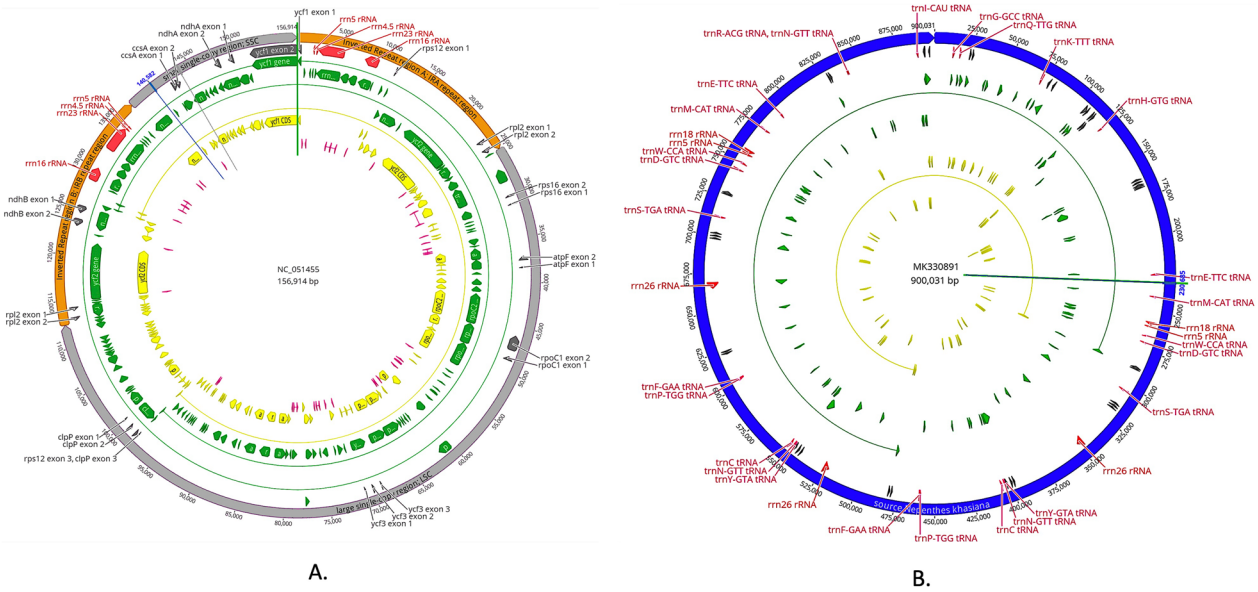

Fig. 1 Organelle genomes of *Nepenthes khasiana*. **A** Whole chloroplast genome annotated, with GenBank accession number MH923233, containing 87 protein-coding genes, 37 tRNAs,and 8 rRNAs. **B** Assembled mitochondrial genome, with GenBank accession number MK330891 depicts a total of 84 genes–50 protein coding genes, 27 tRNAs and 7 rRNAs. Thegraph shows two copies of twelve genes: atp9, cob, cox2, nad2, nad3, nad5, nad6, nad9, rpl5,rps1, rps12, rps4. Genes functioning in related processes are coded with colors.

Table 1 Overview of data files/data sets

Label	Name of data file/data set	File types (File extension)	Data repository and identifier (DOI or accession number)
Data file 1	*Nepenthes khasiana* Chloroplast Genome	Fasta file	GenBank NCBI (https://identifiers.org/nucleotide:MH923233) [18]
Data file 2	*Nepenthes khasiana* mitochondrial genome	Fasta file	GenBank NCBI (https://identifiers.org/nucleotide:MK330891) [19]
Data file 3	Figure 1. Organelle genomes of *Nepenthes khasiana*	Picture file (.jpg)	Figshare (https://doi.org/10.6084/m9.figshare.25984036.v2) [20]

The generated datasets will be used for comparative genomic analysis to identify variations in the cp and mt genomes of *N. khasiana*.

Limitations

The present study reports organelle genomes, which were assembled with short reads. Data from long read sequencer would have been an added advantage, as it would have enabled hybrid strategy in assembling the genomes. The mtDNA is not sufficient as a single source of information for species identification because of introgression, maternal inheritance, recombination, unequal mutation rates, and compounding evolutionary processes. Organelle genomes are less likely to undergo mutation than nuclear genomes, hence this information may not be enough to fully understand the population genetics of *N. khasiana*.

Acknowledgements
We thank the Bioinformatics Centre, NEHU, Shillong, which is financed by DBT, GoI for providing its facilities so that this research could be done. RK is grateful to the UGC, New Delhi, for their financial support.

Author contributions
RK conceived the project, performed bioinformatics analysis and drafted the manuscript. DKB performed the sampling. DKB and DD conceived the project, guided and supervised the data analysis, provided funding support and revised the first drafts of the manuscript. All authors have read and approved the final version of this manuscript.

Funding
This work was funded by the Department of Biotechnology, Govt. of India (Grant ID BT/BII04/035/98, DEPARTMENT OF BIOTECHNOLOGY, GOVT OF INDIA). The funders had no role in the design of this study, during its execution, analyses, interpretation of the data, or decision to publish the paper.

Availability of data and materials
The data sets are openly available in GenBank, RefSeq of NCBI and Figshare at https://identifiers.org/nucleotide:MH923233 (data file 1; Bio-project PRJNA473234) [18], https://identifiers.org/nucleotide:MK330891 (data file 2;

Bio-project PRJNA473234) [19], https://doi.org/https://doi.org/10.6084/m9.figshare.25984036.v2 (data file 3 [20])

Declarations

Ethics approval and consent to participate

Samples were collected from the wild in their natural habitat, complying with the local and national guidelines. A special permission was obtained from the Government of Meghalaya for Dr. Devendra Kumar Biswal to collect *Nepenthes khasiana* samples from Jarain vide Memo No.—FWC/Research/25/4301–04.

Consent for publication

Not applicable.

Competing interests

The authors declare no competing interests.

Received: 27 May 2024 Accepted: 3 January 2025
Published online: 30 January 2025

References

1. Mao AA, Kharbuli P. Distribution and status of Nepenthes khasiana Hook. F.—a rare endemic pitcher plant of Meghalaya, India. Phytotaxonomy. 2002;2:77–83.
2. Ved D, Saha D, Haridasan K, Ravikumar K. Nepenthes khasiana. The IUCN Red List of Threatened Species. 2015;2015.
3. Appendices I, II and III of CITES. Convention on international trade in endangered species of wild fauna and flora. 2023. https://cites.org/eng/app/appendices.php.
4. Singh JN, Mudgal V. Studies on habitat conditions of a few plants species of medicinal values of Nokrek biosphere reserve, Meghalaya. J Non-Timber Forest Products. 1999;6(3/4):192–8.
5. Ghosh D, Ghosh S. India's only pitcher plant in peril. Sci Report. 2012;49:53–6.
6. Dhamecha D, Jalalpure S, Jadhav K. Nepenthes khasiana mediated synthesis of stabilized gold nanoparticles: characterization and biocompatibility studies. J Photochem Photobiol, B. 2016;1(154):108–17.
7. Athauda SB, Matsumoto K, Rajapakshe S, Kuribayashi M, Kojima M, Kubomura-Yoshida N, Iwamatsu A, Shibata C, Inoue H, Takahashi K. Enzymic and structural characterization of nepenthesin, a unique member of a novel subfamily of aspartic proteinases. Biochem J. 2004;381(1):295–306.
8. Gaume L, Bazile V, Huguin M, Bonhomme V. Different pitcher shapes and trapping syndromes explain resource partitioning in Nepenthes species. Ecol Evol. 2016;6(5):1378–92.
9. Clarke C, Schlauer J, Moran J, Robinson AS. Systematics and evolution of Nepenthes. In: Ellison AM, Adamec L, editors. Carnivorous plants: physiology, ecology, and evolution. Oxford University Press; 2018. p. 58–69.
10. Brownlee C. Carnivorous plants: trapping, digesting and absorbing all in one. Curr Biol. 2013;23(17):R714–6.
11. Fu CN, Wicke S, Zhu AD, Li DZ, Gao LM. Distinctive plastome evolution in carnivorous angiosperms. BMC Plant Biol. 2023;23(1):660.
12. Biswal DK, Debnath M, Konhar R, Yanthan S, Tandon P. Phylogeny and biogeography of carnivorous plant family Nepenthaceae with reference to the Indian pitcher plant Nepenthes khasiana reveals an Indian subcontinent origin of Nepenthes colonization in South East Asia during the Miocene epoch. Front Ecol Evol. 2018;2(6):108.
13. Dierckxsens N, Mardulyn P, Smits G. NOVOPlasty: de novo assembly of organelle genomes from whole genome data. Nucleic Acids Res. 2017;45(4): e18.
14. Liu C, Shi L, Zhu Y, Chen H, Zhang J, Lin X, Guan X. CpGAVAS, an integrated web server for the annotation, visualization, analysis, and GenBank submission of completely sequenced chloroplast genome sequences. BMC Genom. 2012;13:715.
15. Tillich M, Lehwark P, Pellizzer T, Ulbricht-Jones ES, Fischer A, Bock R, Greiner S. GeSeq–versatile and accurate annotation of organelle genomes. Nucleic Acids Res. 2017;45(W1):W6–11.
16. Alverson AJ, Wei X, Rice DW, Stern DB, Barry K, Palmer JD. Insights into the evolution of mitochondrial genome size from complete sequences of *Citrullus lanatus* and *Cucurbita pepo* (Cucurbitaceae). Mol Biol Evol. 2010;27(6):1436–48.
17. Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. Basic local alignment search tool. J Mol Biol. 1990;215(3):403–10.
18. Konhar R, Dash D, Biswal D. Nepenthes khasiana GenBank chloroplast genome. GenBank of NCBI. 2020. https://identifiers.org/nucleotide:MH923233.
19. Konhar R, Dash D, Biswal D. Nepenthes khasiana GenBank mitochondrial genome. GenBank of NCBI. 2023. https://identifiers.org/nucleotide:MK330891.
20. Konhar R, Dash D, Biswal D. Fig.1. Organelle genomes of Nepenthes khasiana. figshare. 2024. Figure. https://doi.org/10.6084/m9.figshare.25984036.v2

Publisher's Note