

## Complete chloroplast genome of the tiny marine diatom *Nanofrustulum shiloi* (Bacillariophyta) from the Adriatic Sea

Chunlian Li<sup>a,b</sup>, Romain Gastineau<sup>b</sup>, Monique Turmel<sup>c</sup>, Andrzej Witkowski<sup>b</sup>, Christian Otis<sup>c</sup>, Ana Car<sup>d</sup> and Claude Lemieux<sup>c</sup>

<sup>a</sup>Ecological Institute, South China Normal University, Guangzhou, Guangdong, China; <sup>b</sup>Institute of Marine and Environmental Sciences, University of Szczecin, Szczecin, Poland; <sup>c</sup>Département de biochimie, de microbiologie et de bio-informatique, Institut de Biologie Intégrative et des Systèmes, Université Laval, Québec, Canada; <sup>d</sup>Institute for Marine and Coastal Research, University of Dubrovnik, Dubrovnik, Croatia

### ABSTRACT

We report the chloroplast genome sequence of *Nanofrustulum shiloi*, a tiny araphid pennate diatom collected from the Adriatic Sea. The 160,994-bp *N. shiloi* genome displays a quadripartite structure and its gene repertoire resembles those of other diatom chloroplast genomes. Besides the genes located in the inverted repeat, *psbY* is duplicated. A gene-poor region in the large single-copy region contains multiple ORFs sharing sequence similarities with plasmids and chloroplast ORFs found in other diatom species. The genome features a single intron, a group II intron in *petB*. Phylogenomic analysis identified *N. shiloi* at a basal position within the araphid 2 clade.

### ARTICLE HISTORY

Received 19 August 2019  
Accepted 23 September 2019

### KEYWORDS

Plastid genome; group II intron; pennate araphid diatoms; phylogenomics; Staurosiraceae



*Nanofrustulum shiloi* is a small benthic araphid diatom widely distributed in marine littorals (Li et al. 2018). In the Mediterranean Sea, this pennate diatom is well known for its toxicity against marine invertebrates (Ruocco et al. 2018). With a diameter ranging from 2.0 to 6.0 µm, *N. shiloi* has a potential for diatom-based nanotechnologies (Losic et al. 2010). No complete chloroplast genome has yet been published for the family Staurosiraceae to which this species belongs.

The *N. shiloi* strain (clone SZCZM404) we examined for DNA sequencing was collected in October 2013 from Lumbarda Beach on the Adriatic Sea (Croatia). The culture is available at the Culture Collection of the University of Szczecin (Poland), where permanent slides with cleaned frustules are also deposited. Total DNA was extracted following Doyle and Doyle (1990). Paired-end sequencing was conducted on the BGISEQ-500 platform by the Beijing Genomics Institute (Beijing, China). A total of 60 million 100-bp reads were obtained and assembled using SPAdes v. 3.10.1 (Bankevich et al. 2012). Contigs of chloroplast origin were identified by BLASTN and BLASTX searches. Gene annotations were performed using a custom-built suite of bioinformatics tools (Turmel et al. 2017).

The *N. shiloi* chloroplast genome maps as a circular molecule of 160,994 bp (GenBank: MN276191), with two copies of a large inverted repeat (IR) sequence (11,617 bp) separating the small (SSC, 44,989 bp) and large (LSC, 92,774 bp)

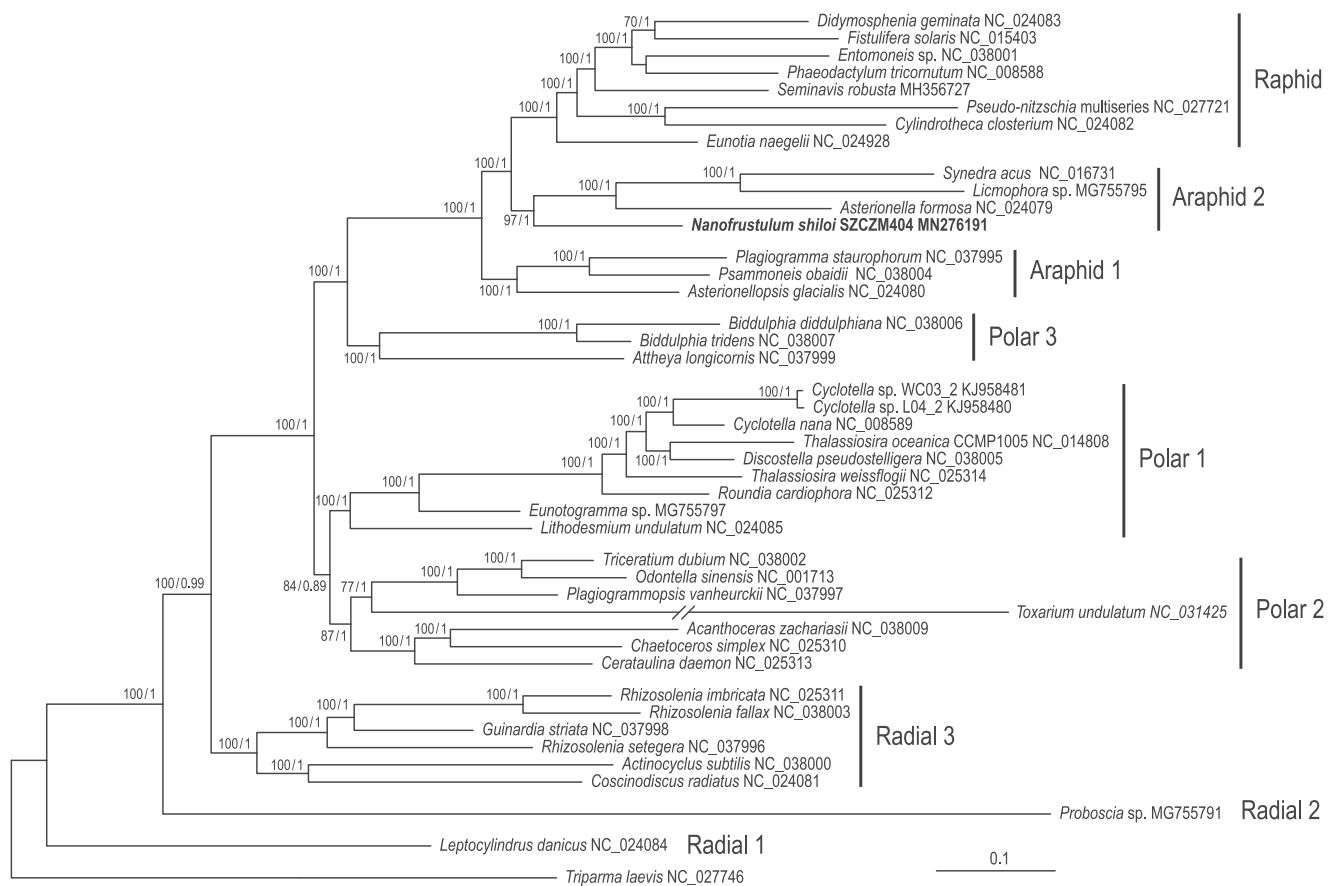
single-copy regions. It encodes 128 proteins of known function, 3 rRNAs, 27 tRNAs, tmRNA (ssrA), and the signal recognition particle RNA (ffs), for a total of 160 gene products. These genes are also usually encoded in the chloroplast in other diatoms (Yu et al. 2018). Two identical copies of the *N. shiloi psbY* gene lie side by side but on different strands in the SSC region, a characteristic apparently specific to this species. Note that *psbY* is located in the IR in several diatoms (Galachyants et al. 2012; Ruck et al. 2014; Sabir et al. 2014). The large intergenic region between *psaJ* and *psaA* in the LSC region harbors two putative serine recombinase genes (*serC*), annotated as *orf100* and *orf221*, as well as other ORFs showing similarities with coding sequences present in chloroplast genomes and plasmids of a number of raphid pennates (Hildebrand et al. 1992; Brembu et al. 2014; Ruck et al. 2014) and dinotoms (Imanian et al. 2010). The *N. shiloi* chloroplast genome features a single intron, a group II intron encoding a putative reverse transcriptase/maturase. This intron resides within *petB* at the same location as a similar intron found in the raphid *Halamphora calidilacuna* (Hamsher et al. 2019).

Maximum-likelihood and Bayesian inference trees were inferred from 129 concatenated chloroplast-encoded proteins of 42 diatom taxa using RAXML v.8.2.3 (Stamatakis 2014) and PhyloBayes v4.1 (Lartillot et al. 2009), respectively, as described by Lemieux et al. (2014). The results

**CONTACT** Claude Lemieux  [claud.lemieux@bcm.ulaval.ca](mailto:claud.lemieux@bcm.ulaval.ca)  Département de biochimie, de microbiologie et de bio-informatique, Institut de Biologie Intégrative et des Systèmes, Université Laval, Pavillon Charles-Eugène-Marchand 1030, avenue de la Médecine, Québec G1V 0A6, Canada

© 2019 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 License (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.



**Figure 1.** Phylogenetic analysis of 129 concatenated chloroplast-encoded proteins from 42 diatoms. The tree represents the best-scoring RAxML tree inferred under the GTR +  $\Gamma$ 4 model. Support values are reported on the nodes, with bootstrap values for the RAxML analysis and the posterior probability values for the PhyloBayes CATGTR +  $\Gamma$ 4 analysis shown from left to right, respectively. *Triparma laevis* (Bolidophyceae) was used to root the tree. The proteins selected for analysis are those conserved in most taxa. Clade labelling is identical to that in Yu et al. (2018). GenBank accession numbers for the chloroplast genomes of all taxa are provided. The scale bar denotes the estimated number of amino acid substitutions per site.

unambiguously positioned *N. shiloi* as the first divergence within the araphid 2 clade, which is sister to all raphid pennates (Figure 1).

## Disclosure statement

The authors report no conflict of interest. The authors alone are responsible for the content and writing of this article.

## Funding

This work was supported by the European Commission's Horizon 2020 Research and Innovation Programme GHANA [734708/GHANA/H2020-MSCA-RISE-2016]; Ministerstwo Nauki i Szkolnictwa Wyższego, Topical Subsidy to the University of Szczecin; Project Maestro from Narodowe Centrum Nauki, Cracow; National Institute of Biological Resources (NIBR) [2012/04/A/ST10/00544]; and the Natural Sciences and Engineering Research Council of Canada [RGPIN-2017-04506].

## References

- 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.
- Brembu T, Winge P, Tooming-Klunderud A, Nederbragt AJ, Jakobsen KS, Bones AM. 2014. The chloroplast genome of the diatom *Seminavis robusta*: new features introduced through multiple mechanisms of horizontal gene transfer. *Mar Genomics.* 16:17–27.
- Doyle JJ, Doyle JL. 1990. Isolation of plant DNA from fresh tissue. *Focus.* 12:13–15.
- Galachyants YP, Morozov AA, Mardanov AV, Beletsky AV, Ravin NV, Petrova DP, Likhoshway YV. 2012. Complete chloroplast genome sequence of freshwater araphid pennate diatom alga *Synedra acus* from lake Baikal. *Int J Biol.* 4:27–35.
- Hamsher SE, Keepers KG, Pogoda CS, Stepanek JG, Kane NC, Kociolek JP. 2019. Extensive chloroplast genome rearrangement amongst three closely related *Halamphora* spp. (Bacillariophyceae), and evidence for rapid evolution as compared to land plants. *PLoS One.* 14(7): e0217824.
- Hildebrand M, Hasegawa P, Ord RW, Thorpe VS, Glass CA, Volcani BE. 1992. Nucleotide sequence of diatom plasmids: identification of open reading frames with similarity to site-specific recombinases. *Plant Mol Biol.* 19(5):759–770.
- Imanian B, Pombert JF, Keeling PJ. 2010. The complete plastid genomes of the two 'dinotoms' *Durinskia baltica* and *Kryptoperidinium foliaceum*. *PLoS One.* 5(5):e10711.
- Lartillot N, Lepage T, Blanquart S. 2009. PhyloBayes 3: a Bayesian software package for phylogenetic reconstruction and molecular dating. *Bioinformatics.* 25(17):2286–2288.
- Lemieux C, Otis C, Turmel M. 2014. Chloroplast phylogenomic analysis resolves deep-level relationships within the green algal class Trebouxiophyceae. *BMC Evol Biol.* 14:211.
- Li CL, Witkowski A, Ashworth MP, Dąbek P, Sato S, Zgłobicka I, Witak M, Khim JS, Kwon C-J. 2018. The morphology and molecular phylogenetics of some marine diatom taxa within the Fragilariaceae, including twenty undescribed species and their relationship to *Nanofrustulum*, *Opephora* and *Pseudostaurosira*. *Phytotaxa.* 355(1):1.

- Losic D, Yu Y, Aw MS, Simovic S, Thierry B, Addai-Mensah J. 2010. Surface functionalisation of diatoms with dopamine modified iron-oxide nanoparticles: toward magnetically guided drug microcarriers with biologically derived morphologies. *Chem Commun.* 46(34): 6323–6325.
- Ruck EC, Nakov T, Jansen RK, Theriot EC, Alverson AJ. 2014. Serial gene losses and foreign DNA underlie size and sequence variation in the plastid genomes of diatoms. *Genome Biol Evol.* 6(3):644–654.
- Ruocco N, Costantini S, Zupo V, Lauritano C, Caramiello D, Ianora A, Budillon A, Romano G, Nuzzo G, D'ippolito G, et al. 2018. Toxigenic effects of two benthic diatoms upon grazing activity of the sea urchin: morphological, metabolomic and de novo transcriptomic analysis. *Sci Rep.* 8(1):5622.
- Sabir JS, Yu M, Ashworth MP, Baeshen NA, Baeshen MN, Bahieldin A, Theriot EC, Jansen RK. 2014. Conserved gene order and expanded inverted repeats characterize plastid genomes of Thalassiosirales. *PLoS One.* 9(9):e107854.
- Stamatakis A. 2014. RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics.* 30(9): 1312–1313.
- Turmel M, Otis C, Lemieux C. 2017. Divergent copies of the large inverted repeat in the chloroplast genomes of ulvophyceean green algae. *Sci Rep.* 7(1):994.
- Yu M, Ashworth MP, Hajrah NH, Khiyami MA, Sabir MJ, Alhebshi AM, Al-Malki AL, Sabir JSM, Theriot EC, Jansen RK. 2018. Evolution of the plastid genomes in diatoms. In: Chaw S-M, Jansen RK, editors. *Adv Bot Res.* London (UK): Academic Press; p. 129–155.