

## The complete mitochondrial genome of Antarctic *Phaeodactylum tricorutum* ICE-H

Lina Liu<sup>a,b</sup>, Yingying He<sup>a,b</sup>, Kai Wang<sup>b</sup>, Jinlai Miao<sup>a,b,c</sup> and Zhou Zheng<sup>a,b,c</sup>

<sup>a</sup>Department of Specialty Medicine, School of Basic Medicine, Qingdao University, Qingdao, China; <sup>b</sup>Key Laboratory of Marine Bioactive Substances, First Institute of Oceanography, Ministry of Natural Resource, Qingdao, China; <sup>c</sup>Laboratory for Marine Drugs and Bioproducts of Qingdao National Laboratory for Marine Science and Technology, Qingdao, China

### ABSTRACT

In this study, the complete mitochondrial genome (mtDNA) of the Antarctic *Phaeodactylum tricorutum* ICE-H was sequenced using Illumina NovaSeq PE150. The circular mtDNA was 77055 bp in size and encodes 60 genes, contains 24 tRNA genes, 34 protein-coding genes, and 2 rRNA genes. The composition of A + T in ICE-H mtDNA was 65.34%. The phylogenetic relationship of 17 species of plant mitochondria were analyzed using the maximum likelihood method by the MEGA-X. *Phaeodactylum tricorutum* ICE-H was most closely related to *Phaeodactylum tricorutum*.

### ARTICLE HISTORY

Received 11 June 2020  
Accepted 20 June 2020

### KEYWORDS

Mitochondrial genome;  
*Phaeodactylum tricorutum*  
ICE-H; Antarctica

Antarctic sea ice microalgae have crucial effects in maintaining the stability of polar ecosystems (Thomas and Dieckmann 2002). They contain many active substances and have great potential to be used as biotechnology products (Zuliani et al. 2016). *Phaeodactylum tricorutum* ICE-H belongs to Phaeodactylaceae, a eukaryotic alga that can grow in extreme Antarctic environment. Therefore, there is an urgent need to understand the genome information of ICE-H. Here, we first assemble and annotate the complete mitochondrial genome (mtDNA) of the Antarctic ICE-H.

Samples used for sequencing were isolated from floating ice near the Zhongshan Research Station of Antarctica (S 69°48', E 77°48'). The sample (Accession no. FIO2008697701) is stored in the Key Laboratory of First Institute of Oceanography, Ministry of Natural Resources. We isolated the complete mtDNA of ICE-H and sequenced it based on the NovaSeq PE150 at the Beijing Novogene Bioinformatics Technology Co., Ltd. The mtDNA was assembled by SOAP denovo (version 2.04) (Li et al. 2008, 2010), integrated using CISA (Lin and Liao 2013) and finally forecasted by GeneMarkS (Version4.17) (Besemer et al. 2001).

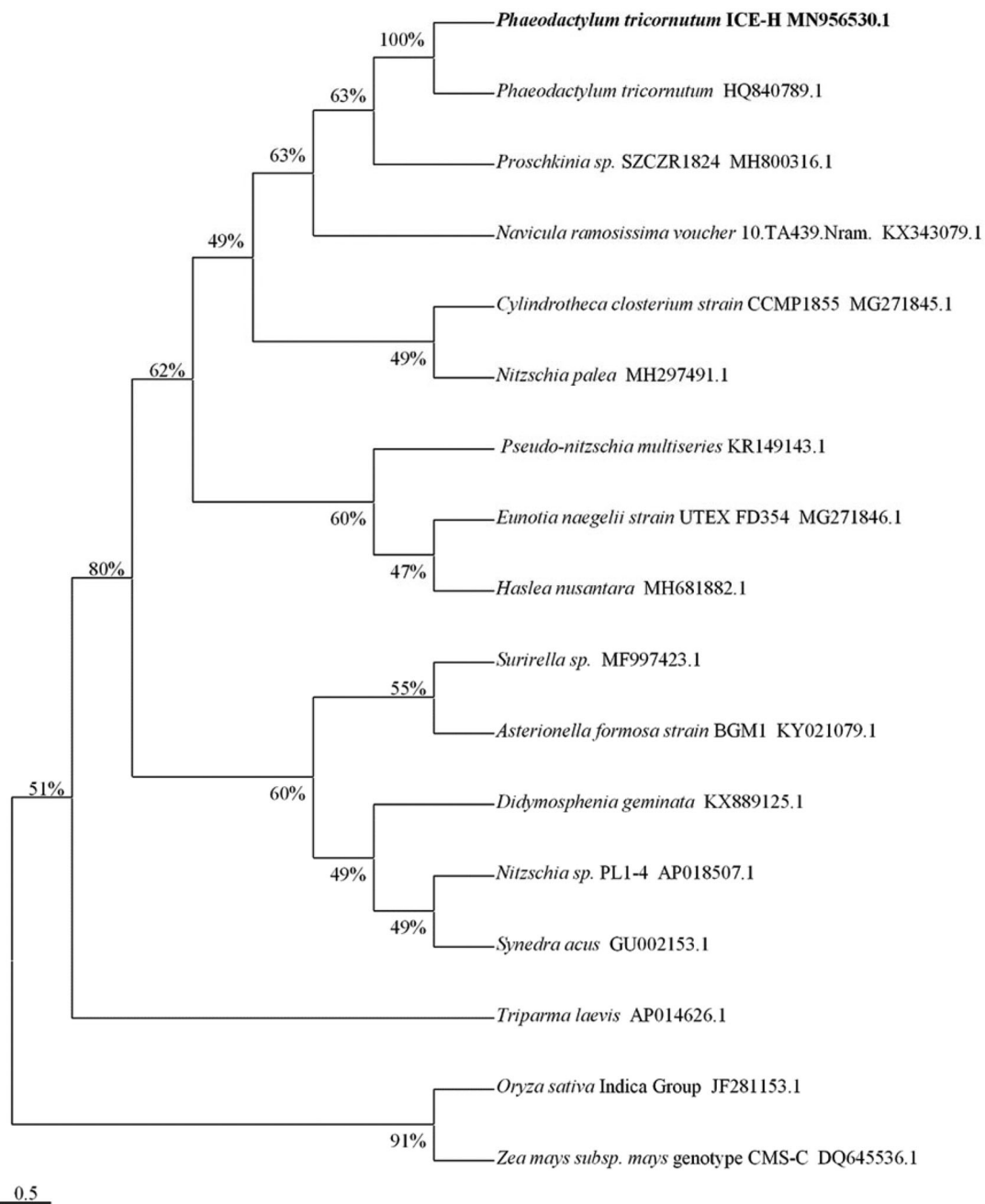
The complete mtDNA of ICE-H is 77055 bp (GenBank registration number: MN956530.1), including 34 protein-coding genes, 24 tRNA genes, and 2 rRNA genes. The mtDNA of ICE-

H composition is 33.40% for A, 16.71% for C, 17.95% for G, and 31.94% for T. The percentage of A + T in ICE-H mtDNA was 65.34%. The 34 PCGS include COX2-3, NAD1-7, NAD4I, NAD11-a, NAD11-b, RPL6, RPL2, RPL5, RPL14, RPL16, RPS2-4, RPS8, RPS10, RPS12-13, RPS19, ATP6, ATP8, ATP9, COB, TATC. Among the 34 PCGS, except NAD7, COB all protein initiation codons are ATG. The termination codon of most protein-coding genes was TAA (5 of 34 genes) or TAG (3 of 34 genes). The 24 tRNA-coding genes ranged in size from 72 bp to 93 bp.

Phylogenetic relationship of 17 complete mtDNAs was analyzed using the maximum likelihood method with 1000 bootstraps (Kumar et al. 2016), which revealed that the ICE-H were most affinitive to the *Phaeodactylum tricorutum* (HQ840789.1). And the distance between *Oryza Sativa Indica* Group and *Zea mays* subsp. was the farthest (Figure 1). The results of the study help to construct the molecular identification system of the Phaeodactylaceae and contribute to the phylogenetic research of the Bacillariophyta family.

### Disclosure statement

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



**Figure 1.** Maximum likelihood phylogenetic tree based on 17 complete mtDNAs.

## Data availability

The data that support the findings of this study are openly available in NCBI (<https://www.ncbi.nlm.nih.gov/>) (GeneBank Number: MN956530.1).

## Funding

This work was supported by the National Key Research and Development Program of China [2018YFD0901105], National Natural Science Foundation of China [31971060] and Key Research and Development Program of Shandong Province [2018YYSP024].

## References

Besemer J, Lomsadze A, Borodovsky M. 2001. GeneMarkS: a self-training method for prediction of gene starts in microbial genomes.

- Implications for finding sequence motifs in regulatory regions. *Nucleic Acids Res.* 29(12):2607–2618.
- Kumar S, Stecher G, Tamura K. 2016. MEGA7: Molecular Evolutionary Genetics Analysis Version 7.0 for Bigger Datasets. *Mol Biol Evol.* 33(7): 1870–1874.
- Li R, Li Y, Kristiansen K, Wang J. 2008. SOAP: short oligonucleotide alignment program. *Bioinformatics.* 24(5):713–714.
- Li R, Zhu H, Ruan J, Qian W, Fang X, Shi Z, Li Y, Li S, Shan G, Kristiansen K, et al. 2010. De novo assembly of human genomes with massively parallel short read sequencing. *Genome Res.* 20(2):265–272.
- Lin SH, Liao YC. 2013. CISA: contig integrator for sequence assembly of bacterial genomes. *PLOS One.* 8(3):e60843.
- Thomas DN, Dieckmann GS. 2002. Antarctic sea ice—a habitat for extremophiles. *Science.* 295(5555):641–644.
- Zuliani L, Frison N, Jelic A, Fatone F, Bolzonella D, Ballottari M. 2016. Microalgae cultivation on anaerobic digestate of municipal wastewater, sewage sludge and agro-waste. *IJMS.* 17(10): 1692.