Complete mitochondrial genome and draft chloroplastic genome of Northern Atlantic *Haslea ostrearia* (Gaillon/Bory) Simonsen, 1974 (Naviculaceae, Bacillariophyceae)

Aurélie Peticca[a], Mostefa Fodil[a], Hélène Gateau[a], Jean-Luc Mouget[a], François Sabot[b], Benoît Chénais[a], Nathalie Casse[a]

[a] BiOSSE (Biology of Organisms: Stress, Health, Environment), UFR Sciences et Techniques, Le Mans Université, F-72085 Le Mans, France
[b] DIADE, University of Montpellier, CIRAD, IRD, F-34394 Montpellier, France

ABSTRACT

The first completed, circular mitochondrial genome and the first draft chloroplastic genome of Northern Atlantic blue diatom *Haslea ostrearia* (Naviculaceae, Bacillariophyceae) are described. The mitochondrial genome is composed of 38,696 bases and contains 64 genes, including 31 protein-coding genes (CDS), 2 ribosomal RNA (rRNA) genes and 23 transfer RNA (tRNA) genes. For the chloroplast, the genome is composed of 130,200 bases with 169 genes (131 CDS, 6 rRNA genes, 31 tRNA genes, and 1 transfermessenger RNA (tmRNA) gene). Phylogenetic trees suggest the proximity of all *H. ostrearia* strains yet available and the possibility to use these genomes as future references.

KEYWORDS

Haslea ostrearia; Chloroplast genome; France ; diatom ; blue microalgae

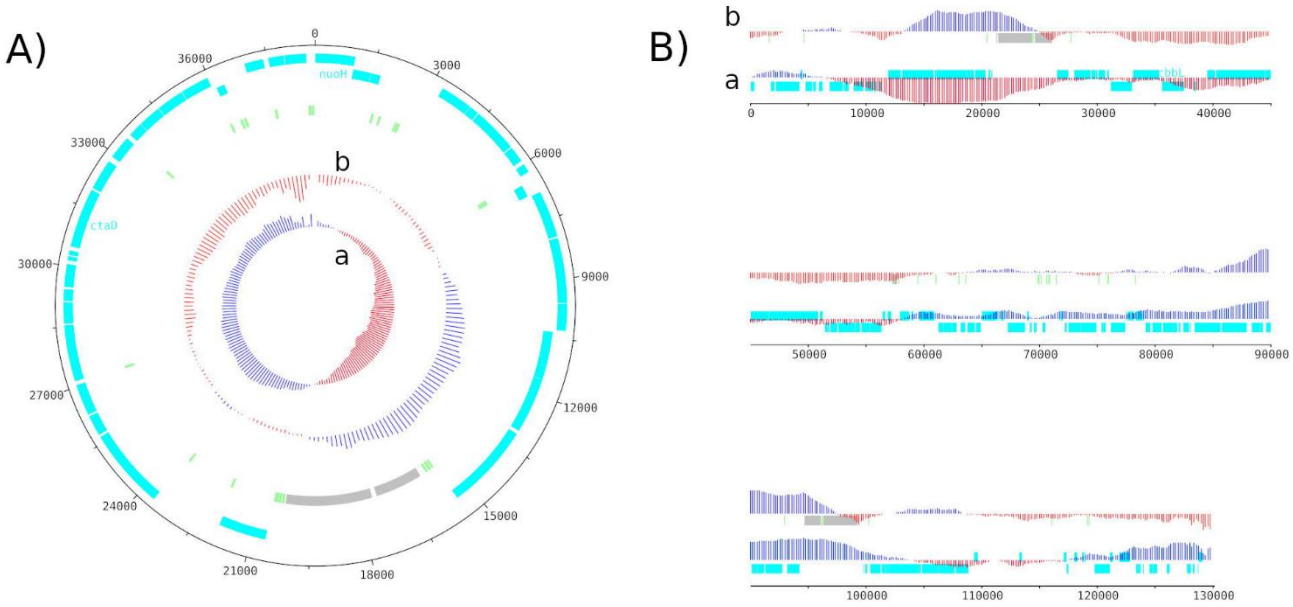
Haslea ostrearia is a blue microalga from the Naviculaceae family, which lives freely in benthic marine environments or as an epiphyte on brown macroalgae (Simonsen 1974). The blue colour comes from a pigment called marrenine that *H. ostrearia* produces and accumulates at cell apices (Gastineau et al. 2014A). This specific pigment is responsible for the greening of oyster gills in farming ponds in Western France. Furthermore, it has been shown that marrenine displays antibacterial, antiviral and antifungal effects (see Gastineau et al. 2014b). The whole DNA of a *H. ostrearia* originating from the Northern Atlantic Ocean was sequenced, allowing the reconstruction of its mitochondrial and chloroplast genomes. This study clears the way for future research about *H. ostrearia* and will help characterise the largely unknown genetics of this species.

The *H. ostrearia* used in this study was isolated in 2018 from an oyster pond in Bouin, France (46°57′12.4″N, 2°02′46.1″W) and deposited in the Nantes Culture Collection (Vona Meleder, vona.meleder@univ-nantes.fr ; Nantes, France) under the name NCC 532. During 21 days, the culture was grown in enriched artificial sea water (*Instant Ocean*, Aquarium systems O ; Harrison et al. 1980 modified by De Brouwer et al. 2002) at 14°C under 300 µm photons/m2/s with a 14h/10h light/dark cycle. On the 20th day of growth, a 1:100 dose of Sigma's antibiotic antimycotic solution (Sigma-Aldrich, Saint-Quentin Fallavier, France; catalogue#A5955) was added to the culture mix. After 24h, the biomass was collected by filtration (Whatman™ Binder-Free Glass Microfiber Filters, Grade GF/C, pore size of 1.2 µm) and the whole DNA was extracted using the method of Puppo et al. (2017). Extracted DNA was sequenced using the PacBio continuous long reads (PB CLR SEQUEL2) and Illumina MiSeq platforms (TrueSeqv3, 250pe; Genotoul, Toulouse, France), and 85 Gb and 2.36 Gb total read lengths were generated, respectively. Illumina reads were filtered to remove low-quality reads (<Q30), short reads (<75b) and adapter sequences were searched and trimmed by *trimmomatic* 0.39 (Bolger et al. 2014). The *de novo* genome assembly was performed using *Flye* 2.9 (Kolmogorov et al. 2020) and the long PB CLR reads with the following options: -g 100m --meta. Polishing was performed with filtered Illumina reads through three loops of *racon* 1.4.20 (Vaser et al. 2017) and *bwa-mem* v2 2.2.1 (Vasimuddin et al. 2019). The annotation and gene prediction were performed by *Prokka* v1.14.6 (Seeman 2014). Since the mitochondrial genome is circular, *samtools faidx* v1.14 (Danecek et al. 2021) was used to relocate the annotated origin of H-strand replication (OH) as the starting gene (position +1). Genome maps were generated with *Artemis* 18.2.0 (Rutherford et al. 2000).

Phylogenetic trees were created thanks to *NGPhylogeny.fr* pipeline (options trimAl and PhyML+SMS, Lemoine et al. 2019) with the partial *COX1* gene for the mitochondrial and the partial *rbcL* gene chloroplast genomes. The mitochondrial dataset grouped sequences from different diatom

CONTACT Aurélie Peticca aurelie.peticca@univ-lemans.fr; Benoît Chénais bchenais@univ-lemans.fr

Figure 1 Genome Map of the circular mitochondrial genome (A) and the linear chloroplastic genome (B). The red and blue colors in GC shiew (a) and GC plot (b) show if the value is below or above average, respectively. CDS are shown in light blue, tRNA in light green and the rRNA in light grey.

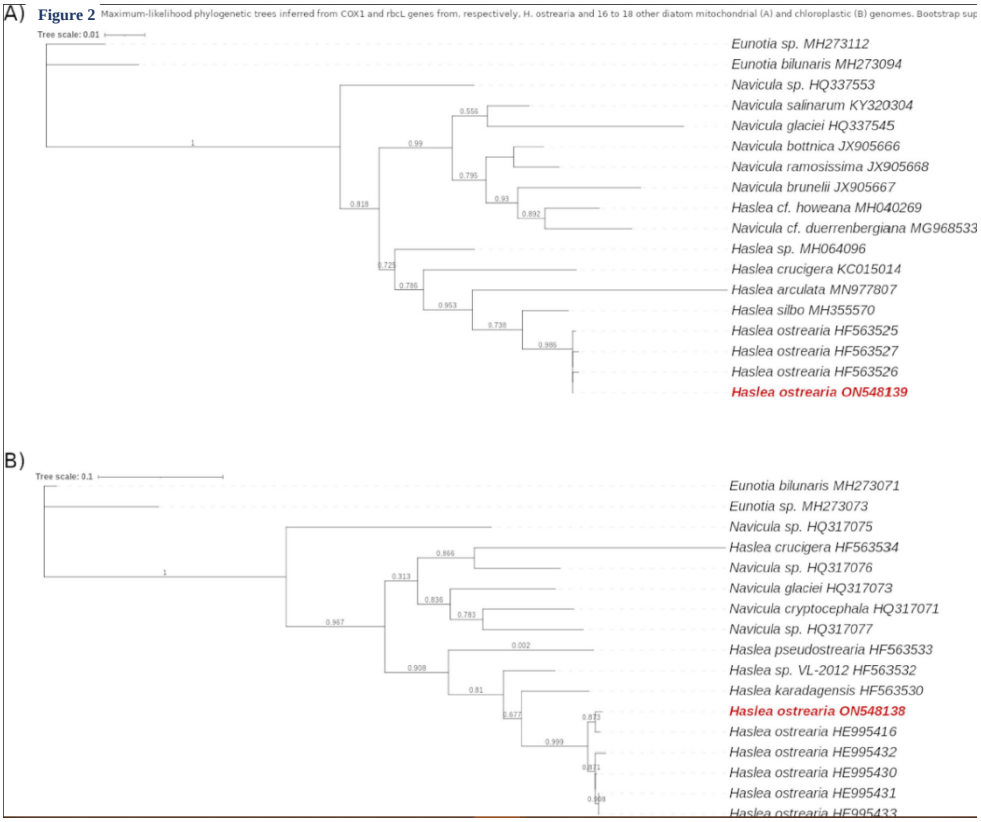


taxa: 5 *H. ostrearia* strains, 4 other *Haslea* species, 5 *Navicula* species and 2 external species from the *Eunotia* family. The chloroplast dataset also included different diatom sequences: 3 *H. ostrearia* strains, 5 other *Haslea* species, 8 *Navicula* species and 2 external species from the *Eunotia* family. All sequences were downloaded from NCBI website (https://www.ncbi.nlm.nih.gov/).

The complete circular mitochondrial genome is 38,696 bases long with a GC content (%GC) of 28.66% (36.07% A, 35.26% T, 14.76% C, 13.91% G), close to the *H. nusantara* mitochondrion(36,288 bases, 29.24% GC; Prasetiya et al. 2019). The sequence depth associated with this assembly is 1,325X. The 64 annotated genes were more numerous than *H. nusantara* ones with 3 additional genes, for a total of 39 CDS, 2 rRNA genes and 23 tRNA genes (Figure 1A). The *H. ostrearia COX1* gene, previously partially sequenced (Gastineau et al. 2013), was found complete in this study and annotated as *ctaD* by Prokka. The draft chloroplast genome was 130,200 bases long with 31.04% GC (34.19 A, 34.76% T, 15.31% C, 15.73% G), almost the same as the *H. nusantara* chloroplast (120,448 bases and 31.10%GC ; Prasetiya et al. 2019), with a sequence depth of 2,371X. One hundred and thirty one CDS, 6 rRNA genes, 31 tRNA genes and 1 tmRNA gene were identified within this genome (Figure 1B), quite close to the annotated genes of *H. nusantara* chloroplast. The known *H. ostrearia rbcL* gene, annotated as *cbbL*, and the *psbC* partial gene (Gastineau et al. 2013) were also found and completed in this study. Phylogenetic trees of both mitochondrial and chloroplastic sequences support the hypothesis that *H. ostrearia* strains are very close to each other and have a different evolutionary history from other diatoms or *Haslea* species (Figure 2). However, due to a lack of information about genes in the genus *Haslea*, only the *COX1* and the *rbcL* genes were tested.

Even if the chloroplast genome is not completed, it seems that only a few bases are missing since its size and its composition (%GC and genes) are similar to the plastid from the close species *H. nusantara*. To fill the last gaps, further analyses are needed, probably with other sequencing data type. Since all *H. ostrearia* strains for which DNA sequences are available present a close proximity to

Figure 2 Maximum-likelihood phylogenetic trees inferred from COX1 and rbcL genes from, respectively, H. ostrearia and 16 to 18 other diatom mitochondrial (A) and chloroplastic (B) genomes. Bootstrap sup



CONTACT Aurélie Peticca aurelie.peticca@univ-lemans.fr; Benoît Chénais bchenais@univ-lemans.fr

each other, the mitochondrial and chloroplast genomes may be used as a reference for the species.

Disclosure statement

The authors declare that they have no competing interests.

Funding

This research was supported by the Région Pays de la Loire (Lancom project) and Ph.D. grant for A.P. was co-funded by Région Pays de la Loire (Lancom project) and Le Mans Université. M.F. benefited from the Horizon 2020 Research and Innovation Program GHaNA (The Genus Haslea, New marine resources for blue biotechnology and Aquaculture, grant agreement No [734708/GHANA/H2020-MSCA-RISE-2016] J.-L.M.).

ORCID

Aurélie Peticca https://orcid.org/0000-0002-8972-0973

Acknowledgments

We express our sincere thanks to Mr. Christophe Klopp for helping us to reconstruct H. ostrearia genome, and to Mr Bruno Cognie for isolating the strain studied here. All sequencing analyses have been carried out by the Genotoul platform (Toulouse, France).

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. ON548138-ON548139. The associated BioProject, SRA and Bio-Sample numbers are PRJNA843895, SRR19450090/SRR19450089, and SAMN28772203 respectively.

References

Bolger, Anthony M., Marc Lohse, and Bjoern Usadel. 'Trimmomatic: A Flexible Trimmer for Illumina Sequence Data'. *Bioinformatics* 2014, 30(15), 2114–2120. https://doi.org/10.1093/bioinformatics/btu170.

Danecek, Petr, James K Bonfield, Jennifer Liddle, John Marshall, Valeriu Ohan, Martin O Pollard, Andrew Whitwham, et al. 'Twelve Years of SAMtools and BCFtools'. *GigaScience* 2021, 10(2), giab008. https://doi.org/10.1093/gigascience/giab008.

De Brouwer, J. F. C., K. Wolfstein, and L. J. Stal. 'Physical Characterization and Diel Dynamics of Different Fractions of Extracellular Polysaccharides in an Axenic Culture of a Benthic Diatom'. *European Journal of Phycology* 2002, 37(1), 37–44. https://doi.org/10.1017/S0967026201003419.

Gastineau, Romain, Vincent Leignel, Boris Jacquette, Yann Hardivillier, Angela Wulff, Pierre Gaudin, Djamel Bendahmane, Nicolaï A. Davidovich, Irena Kaczmarska, and Jean-Luc Mouget. 'Inheritance of Mitochondrial DNA in the Pennate Diatom *Haslea Ostrearia* (Naviculaceae) during Auxosporulation Suggests a Uniparental Transmission'. *Protist* 2013, 164(3), 340–351. https://doi.org/10.1016/j.protis.2013.01.001.

Gastineau, Romain, Nikolai Davidovich, Gert Hansen, Jan Rines, Angela Wulff, Irena Kaczmarska, James Ehrman, et al. 'Haslea Ostrearia-like Diatoms'. In: Sea Plants, Bourgougnon N., editor. *Advances in Botanical Research*, Academic Press, Vol. 71 t, p. 441–465, 2014a. https://doi.org/10.1016/B978-0-12-408062-1.00015-9.

Gastineau, Romain, François Turcotte, Jean-Bernard Pouvreau, Michèle Morançais, Joël Fleurence, Eko Windarto, Fiddy Prasetiya, et al. 'Marennine, Promising Blue Pigments from a Widespread Haslea Diatom Species Complex'. *Marine Drugs* 2014b, 12(6), 3161–3189. https://doi.org/10.3390/md12063161.

Harrison, Paul J., Rosemary E. Waters, and F. J. R. Taylor. 'A BROAD SPECTRUM ARTIFICIAL SEA WATER MEDIUM FOR COASTAL AND OPEN OCEAN PHYTOPLANKTON [1]'. *Journal of Phycology* 1980, 16(1), 28–35. https://doi.org/10.1111/j.0022-3646.1980.00028.x.

Kolmogorov, Mikhail, Derek M. Bickhart, Bahar Behsaz, Alexey Gurevich, Mikhail Rayko, Sung Bong Shin, Kristen Kuhn, et al. 'MetaFlye: Scalable Long-Read Metagenome Assembly Using Repeat Graphs'. *Nature Methods* 2020, 17(11), 1103–1110. https://doi.org/10.1038/s41592-020-00971-x.

Lemoine, Frédéric, Damien Correia, Vincent Lefort, Olivia Doppelt-Azeroual, Fabien Mareuil, Sarah Cohen-Boulakia, and Olivier Gascuel. 'NGPhylogeny.Fr: New Generation Phylogenetic Services for Non-Specialists'. *Nucleic Acids Research* 2019, 47(W1), W260–W265. https://doi.org/10.1093/nar/gkz303.

Prasetiya, Fiddy Semba, Romain Gastineau, Michel Poulin, Claude Lemieux, Monique Turmel, Agung Dhamar Syakti, Yann Hardivillier, et al. 'Haslea Nusantara (Bacillariophyceae), a New Blue Diatom from the Java Sea, Indonesia: Morphology, Biometry and Molecular Characterization'. *Plant Ecology and Evolution* 2019, 152(2), 188–202. https://doi.org/10.5091/plecevo.2019.1623.

Puppo, Carine, Tomas Voisin, and Brigitte Gontero. 'Genomic DNA Extraction from the Pennate Diatom Asterionella Formosa Optimised for next Generation Sequencing. V1'. Preprint, 21 September 2017. https://doi.org/10.17504/protocols.io.jytcpwn.

Seemann, T. 'Prokka: Rapid Prokaryotic Genome Annotation'. *Bioinformatics* 2014, 30(14), 2068–2069. https://doi.org/10.1093/bioinformatics/btu153.

Simonsen, R. *The Diatom Plankton of the Indian Ocean Expedition of RV 'Meteor' 1964 - 1965*. 'Meteor' Forschungsergebnisse Reihe D, Biologie, No. 19. Berlin Stuttgart: Borntraeger, 1974.

Vaser, Robert, Ivan Sović, Niranjan Nagarajan, and Mile Šikić. 'Fast and Accurate de Novo Genome Assembly from Long Uncorrected Reads'. *Genome Research* 2017, 27(5), 737–746. https://doi.org/10.1101/gr.214270.116.

Vasimuddin, Md., Sanchit Misra, Heng Li, and Srinivas Aluru. 'Efficient Architecture-Aware Acceleration of BWA-MEM for Multicore Systems'. In *2019 IEEE International Parallel and Distributed Processing Symposium (IPDPS)*, 314–24. Rio de Janeiro, Brazil: IEEE, 2019. https://doi.org/10.1109/IPDPS.2019.00041.

CONTACT Aurélie Peticca aurelie.peticca@univ-lemans.fr; Benoît Chénais bchenais@univ-lemans.fr