Complete mitochondrial genome sequence of the freshwater diatom Asterionella formosa
Adrien Villain, Mila Kojadinovic, Carine Puppo, Laura Prioretti, Pierre Hubert, Gerald Gregori, Alain Roulet, Céline Roques, Jean-Michel Claverie, Brigitte Gontero, et al.

To cite this version:
Adrien Villain, Mila Kojadinovic, Carine Puppo, Laura Prioretti, Pierre Hubert, et al.. Complete mitochondrial genome sequence of the freshwater diatom Asterionella formosa. Mitochondrial DNA Part B, Taylor Francis Online, 2017, 2 (1), pp.97 - 98. <10.1080/23802359.2017.1285210>. <hal-01760365>

HAL Id: hal-01760365
https://hal-amu.archives-ouvertes.fr/hal-01760365
Submitted on 25 Apr 2018

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L’archive ouverte pluridisciplinaire HAL, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d’enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.
Complete mitochondrial genome sequence of the freshwater diatom Asterionella formosa

Adrien Villain, Mila Kojadinovic, Carine Puppo, Laura Prioretti, Pierre Hubert, Yizhi Zhang, Gérald Grégori, Alain Roulet, Céline Roques, Jean-Michel Claverie, Brigitte Gontero & Guillaume Blanc

To cite this article: Adrien Villain, Mila Kojadinovic, Carine Puppo, Laura Prioretti, Pierre Hubert, Yizhi Zhang, Gérald Grégori, Alain Roulet, Céline Roques, Jean-Michel Claverie, Brigitte Gontero & Guillaume Blanc (2017) Complete mitochondrial genome sequence of the freshwater diatom Asterionella formosa, Mitochondrial DNA Part B, 2:1, 97-98, DOI: 10.1080/23802359.2017.1285210

To link to this article: https://doi.org/10.1080/23802359.2017.1285210
Complete mitochondrial genome sequence of the freshwater diatom *Asterionella formosa*

Adrien Villain\(^a\), Mila Kojadinovic\(^b\), Carine Puppo\(^b\), Laura Priorett\(^b\), Pierre Hubert\(^c\), Yizhi Zhang\(^b\), Gérald Grégoiri\(^d\), Alain Roule\(^e\), Céline Roques\(^e\), Jean-Michel Claverie\(^b\)^{g}\, Brigitte Gonter\(^b\)\ and Guillaume Blanc\(^a\)

\(^a\)Information Génomique & Structurale UMR 7256, Aix Marseille Univ CNRS, IMM FR 3479, Marseille, France; \(^b\)BIP UMR 7281, IMM FR 3479, Aix Marseille Univ CNRS, Marseille Cedex 20, France; \(^c\)Laboratoire d’Ingénierie des Systèmes Macromoléculaires, Aix Marseille Univ CNRS UMR 7255 (IMM FR 3479), Marseille, France; \(^d\)Mediterranean Institute of Oceanography, Aix Marseille Univ, Univ Toulon, CNRS, Marseille, France; \(^e\)GeT-PlaGe, Genotoul, INRA, Castanet-Tolosan, France; \(^f\)UAR1209, INRA, Castanet-Tolosan, France; \(^g\)Assistance Publique des Hôpitaux de Marseille (APHM), Marseille, France

ABSTRACT

We report the complete mitochondrial genome sequence of the freshwater diatom *Asterionella formosa*. The large 61.9 kb circular sequence encodes 34 proteins and 25 tRNAs that are universally conserved in other sequenced diatoms. We fully resolved a unique 24 kb region containing highly conserved repeated sequence units, possibly collocating with an origin of replication.

**ARTICLE HISTORY**
Received 23 December 2016
Accepted 18 January 2017

**KEYWORDS**
Diatoms; Asterionella; mitogenome; pacbio

Diatoms are one of the largest and ecologically most significant groups of organisms on the Earth. These unicellular stramenopile algae are broadly distributed in marine and freshwater habitats and studied for potential biotechnological applications as well. *Asterionella formosa* Hassall (Lund 1949) is a freshwater araphid pennate diatom species forming typical star-shaped colonies. A single colony was isolated from Esthwaite Water (54.4N, 2.9W) in the English Lake District in December 2014. DNA was extracted following a hexadecyltrimethylammonium bromide (CTAB)-based protocol, and sequenced using the Pacific Biosciences RSII instrument.

Genome assembly of data from 13 SMRT cells was performed using the HGAP 2.0 protocol (Chin et al. 2013) implemented in SMRT analysis (2.3.0.140936.p0.0). A 82,419-bp long contig was identified as the mitochondrial genome and manually circularized into a 61,877-bp chromosome. Protein-coding genes were predicted by retaining all open reading frames (ORFs) > 100 codons, whereas ORFs < 100 codons were only predicted as genes when exhibiting a BLASTP (Altschul 1997) match (E-value < 1E-5) in the NCBI non-redundant (Nr) database. For all validated genes, start codon predictions were further refined by comparison with homologous sequences. Transfer RNAs were predicted using tRNAscan SE (Lowe & Eddy 1997) and ribosomal RNAs were predicted by alignment with diatom reference sequences. We annotated repeated sequences by combining the results of tandem repeats finder (Benson 1999) and local BLASTN searches. Functional annotations were gathered and manually validated within Unipro UGENE (the UGENE team 2012). The genome sequence is available in DDBJ/EMBL/GenBank under the accession no. KY021079.

The gene content of the *A. formosa* mitogenome is almost identical to previously published diatoms mitochondrial genomes (Secq & Green 2011; Ravin et al. 2010). The 62 genes include small and large rRNAs subunits, 25 tRNAs, and 35 protein-coding genes encoding 16 ribosomal protein subunits (rps), 10 NADH dehydrogenase subunits (nad), 3 ATPase subunits (atp), 3 cytochrome oxidase subunits (cox), the apocytochrome B, and the Sec-independent transllocate protein TatC. A single type II intron with an intronic reverse-transfer transcriptase domain is located in the cox1 gene, while two are found in its homolog in other diatoms. The gene cluster rpl10-rps8-rpl6-rps2-rps4-atp8-rps12-rps7-rpl14-rpl5-nad1-tatC-rps11-rpl2-rps19-rps3-rpl16-apa9-nad4l-nad11 is present and seems conserved among diatoms. As in *Phaeodactylum tricornutum*, but unlike *Thalassiosira pseudonana* and *Synechococcus acus*, most protein-coding genes (32/35 = 91%) are encoded on the same strand. This strong-bias is apparently species-specific as it does not correlate with the phylogenetic relationships shown in Figure 1. Gene density is high and illustrated by overlaps between rpl2 and rps19, and rps19 and rps3. In contrast, a 24.9 kb long region devoid of predicted coding sequences and composed of successive blocks of various tandemly arrayed repeats is located between tmQ and nad11.

A single, large repeat region has already been described in other mitogenomes and has been suggested to serve as a replication origin in two cryptophytes (Hauth 2005; Kim et al. 2008). This feature is also typical of diatom mitogenomes.
Alignments were performed with MUSCLE (Edgar 2004) and the tree was constructed with PhyML (Guindon et al. 2009) using the CpREV model, Bootstrap values of 100 permutations were inserted. Please confirm the statement is accurate.

Figure 1. Visualization with TreeGraph (Stöver & Müller 2010) of a phylogenetic tree of selected diatoms and stramenopile mitogenomes based on the concatenation of the protein sequences of the following genes: atp6, atp8, cox1, cox2, cox3, nad1, nad2, nad4, nad4l, nad5, nad6, nad7, rpl2, rps12, rps19, rps16, rpl22. Alignments were performed with MUSCLE (Edgar 2004) and the tree was constructed with PhyML (Guindon et al. 2009) 20120412 using the CpREV model, selected by ProtTest 3 (Darriba et al. 2011). Bootstrap values of 100 permutations are indicated at the nodes.

(Ravin et al. 2010; Secq & Green 2011), however, neither the sequences nor the general organizations of these repeats are conserved. A few ion torrent-sequenced diatom mitogenomes are reportedly lacking such a repeat region (An et al. 2016a, 2016b) but this absence may be due to the incapacity of short reads to resolve complex repeats, contrary to Sanger or Pacbio sequencing.

Disclosure statement

A declaration of interest statement reporting no conflict has been inserted. Please confirm the statement is accurate.

Funding

This work was supported by the A*MIDEX project (ANR-11-IDEEX-0001-02) funded by the ‘Investissements d’avenir’ French Government program, managed by the French National Research Agency (ANR), the Centre national de la recherche scientifique interdisciplinary PEPS project Exomod, the Aix-Marseille University, and la Région PACA. We acknowledge the use of the PACA-Bioinfo Platform, supported by IBISA and France-Genomique (ANR-10-INBS-0009). This work was performed in collaboration with the GeT core facility, Toulouse, France (http://get.genotoul.fr), and was supported by France Genomique National infrastructure, funded as part of ‘Investissement d’avenir’ program managed by Agence Nationale pour la Recherche (contract ANR-10-INBS-09).

References