



HAL
open science

Efficiency of 3 genetic markers to determine the composition of a diatom assemblage using next generation sequencing

Lenaïg Kermarrec, Frédéric Rimet, Alain Franc, Philippe Chaumeil, Jean Francois Humbert, Agnes Bouchez

► **To cite this version:**

Lenaïg Kermarrec, Frédéric Rimet, Alain Franc, Philippe Chaumeil, Jean Francois Humbert, et al.. Efficiency of 3 genetic markers to determine the composition of a diatom assemblage using next generation sequencing. 4thInternational Barcode of Life Conference, University of Adelaide. Adélaide, AUS., Nov 2011, Adélaide, Australia. 1p. hal-02805342

HAL Id: hal-02805342

<https://hal.inrae.fr/hal-02805342>

Submitted on 6 Jun 2020

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.



EFFICIENCY OF 3 GENETIC MARKERS TO DETERMINE THE COMPOSITION OF A DIATOM ASSEMBLAGE USING NEXT GENERATION SEQUENCING

KERMARREC, L. (1), Rimet, F. (1), Franc, A. (2), Chaumeil, P. (2), Humbert, J.-F. (3), BOUCHEZ, A. (1)

(1) INRA, UMR CARRETEL, Thonon-les-Bains, France, France

(2) INRA, UMR BIOGECO, Cestas, France, France

(3) INRA, UMR BIOEMCO, Paris, France, France

Thousands of diatom samples are analyzed every year to monitor water quality. Accurate and rapid species identification of this protist is challenging in the context of the European Water Framework Directive. Routine identification of diatoms based on morphological characterization of species by light microscopy requires a long training, is time consuming and can be observer dependent. DNA methods proved their ability to discriminate diatom species.

Next generation sequencing of 3 markers, SSU rRNA sequence, *rbcL* and *cox1* genes was used to establish diatom inventories with presence / absence of taxa in 3 synthetic samples. These samples were composed by strains of diatoms representing 21 species. Two were made of pooled PCR products, while the third was made directly of pooled cultures. For each sample, reads were compared for each marker to a reference database of diatom sequences constituted with sequences from our lab and GenBank. The inventories provided by the 3 markers were compared at different taxonomic levels (clades and species) for each sample.

Although the reference database of SSU rDNA was the largest, the SSU rDNA gene has a low resolution at all taxonomic level probably due to conserved regions which are common even among subdivisions. *Cox1*, where polymorphism is distributed throughout the sequenced fragment, has the higher resolution power and is suitable to distinguish diatom species. However, the inventories obtained from *cox1* sequences are the most distant from the expected results. Its reference database is smaller, due to sequencing difficulties. Inventories from *rbcL* data are closely related to expected results. As for *cox1*, polymorphism is distributed throughout the gene. The advantage of *rbcL* is to offer a high resolution power together with a large reference database. *RbcL* tends therefore to be the best gene available at present to define a barcode and to assess diatom community.

Keywords: Protists, Algae, Environmental Quality Assessment, Environmental Barcoding, Next-gen sequencing