

PhytoREF: a reference database of the plastidial 16S rRNA gene of photosynthetic eukaryotes with curated taxonomy

JOHAN DECELLE,*+ SARAH ROMAC,*+ ROWENA F. STERN,‡ EL MAHDI BENDIF,§ ADRIANA ZINGONE,¶ STÉPHANE AUDIC,*+ MICHAEL D. GUIRY,** LAURE GUILLOU,*+ DÉSIRÉ TESSIER,+++ FLORENCE LE GALL,*+ PRISCILLIA GOURVIL,*+ ADRIANA L. DOS SANTOS,*+ IAN PROBERT,*+ DANIEL VAULOT,*+ COLOMBAN DE VARGAS*+ and RICHARD CHRISTEN+++‡

*UMR 7144 - Sorbonne Universités, UPMC Univ Paris 06, Station Biologique de Roscoff, Roscoff 29680, France, †CNRS, UMR 7144, Station Biologique de Roscoff, Roscoff 29680, France, ‡Sir Alister Hardy Foundation for Ocean Science, The Laboratory, Citadel Hill, Plymouth PL1 2PB, UK, §Marine Biological Association, The Laboratory, Citadel Hill, Plymouth PL1 2PB, UK, ¶Stazione Zoologica Anton Dohrn, Villa Comunale, Naples 80121, Italy, **The AlgaeBase Foundation, c/o Ryan Institute, National University of Ireland, University Road, Galway Ireland, ††CNRS, UMR 7138, Systématique Adaptation Evolution, Parc Valrose, BP71, Nice F06108, France, †††Université de Nice-Sophia Antipolis, UMR 7138, Systématique Adaptation Evolution, Parc Valrose, BP71, Nice F06108, France

Abstract

Photosynthetic eukaryotes have a critical role as the main producers in most ecosystems of the biosphere. The ongoing environmental metabarcoding revolution opens the perspective for holistic ecosystems biological studies of these organisms, in particular the unicellular microalgae that often lack distinctive morphological characters and have complex life cycles. To interpret environmental sequences, metabarcoding necessarily relies on taxonomically curated databases containing reference sequences of the targeted gene (or barcode) from identified organisms. To date, no such reference framework exists for photosynthetic eukaryotes. In this study, we built the PhytoREF database that contains 6490 plastidial 16S rDNA reference sequences that originate from a large diversity of eukaryotes representing all known major photosynthetic lineages. We compiled 3333 amplicon sequences available from public databases and 879 sequences extracted from plastidial genomes, and generated 411 novel sequences from cultured marine microalgal strains belonging to different eukaryotic lineages. A total of 1867 environmental Sanger 16S rDNA sequences were also included in the database. Stringent quality filtering and a phylogeny-based taxonomic classification were applied for each 16S rDNA sequence. The database mainly focuses on marine microalgae, but sequences from land plants (representing half of the PhytoREF sequences) and freshwater taxa were also included to broaden the applicability of PhytoREF to different aquatic and terrestrial habitats. PhytoREF, accessible via a web interface (<http://phyto-ref.fr>), is a new resource in molecular ecology to foster the discovery, assessment and monitoring of the diversity of photosynthetic eukaryotes using high-throughput sequencing.

Keywords: high-throughput sequencing, metabarcoding, photosynthesis, phytoplankton, plastidial 16S rRNA gene, protists

Received 12 November 2014; revision received 20 February 2015; accepted 2 March 2015

Introduction

Eukaryotes that acquired photosynthesis through endosymbiosis with cyanobacteria or plastid-bearing eukaryotes are distributed across most eukaryotic super-groups and exhibit a bewildering morphological diversity across more than eight orders of magnitude in organism size (Archibald 2012; Not *et al.* 2012). Most photosynthetic

eukaryotes are unicellular (referred to as protists), but a few lineages, essentially macroalgae (e.g. the rhodophyte class Florideophyceae or the chlorophyte class Ulvophyceae) and the embryophyte land plants, have evolved into multicellular forms. The radiation of photosynthetic marine protists during the Neoproterozoic arguably led to a major oxidation event in the history of the Earth system (Knoll 2014). Today, eukaryotic microalgae are key players in aquatic food webs and global biogeochemical processes. In the marine ecosystem, they are the major contributors to primary production through their capacity to perform oxygenic photosynthesis (Falkowski

Correspondence: Colomban de Vargas and Richard Christen
Fax: +33 2 98 29 23 24; E-mails: vargas@sb-roscoff.fr and Richard.christen@unice.fr