Diversity and distribution of haptophytes revealed by environmental sequencing and metabarcoding – a review

Bente Edvardsen1*, Elianne Sirnæs Egge1 and Daniel Vaulot2

* Corresponding author: bente.edvardsen@ibv.uio.no
1 University of Oslo, Department of Biosciences, P.O. Box 1066 Blindern, 0316 Oslo, Norway
2 Sorbonne Universités, UPMC Univ Paris 06, CNRS, UMR 7144, Station Biologique, Place Georges Teissier, 29680 Roscoff, France

With 2 figures and 2 tables in the text and an electronic supplement

Abstract: Microalgae of the division Haptophyta are a major component of the marine nanoplanckton and present in all seas. They are important primary producers and grazers of picoplankton in the ocean, and their metabolic products can have an impact on global climate. Due to their small size, species are difficult to distinguish by microscopy, and knowledge on their diversity and distribution is incomplete. Environmental sequencing studies have revealed a high marine protist diversity. We review the current knowledge on diversity and distribution of haptophytes revealed by Sanger sequencing of clone libraries (environmental sequencing) and by high-throughput sequencing of amplicons (metabarcoding). We also discuss the methodology used. Finally, we provide a curated haptophyte reference 18S rRNA-gene database for future taxonomic assignment of environmental sequences and metabarcoding reads.

Keywords: clone libraries; diversity; distribution; ecology; environmental sequencing; Haptophyta; high-throughput sequencing; novel lineages; metabarcoding; reference database

Introduction

Haptophytes are mostly single-celled nanophytoplankton (usually 2–30 μm) and have a worldwide distribution. They are one of the major groups of primary producers in the ocean, together with cyanobacteria, diatoms, dinoflagellates and prasinophytes. They have also been found to be important grazers on picoplankton in the open ocean (e.g. Frias-Lopez et al. 2009; Unrein et al. 2013). Some haptophytes, such as members of the genera Emiliania, Gephyrocapsa, Phaeocystis, Chrysochromulina and Prymnesium, form extensive blooms that may affect the global carbon balance and possibly climate forcing, or cause fish kills with ecological and economical impact (Jordan & Chamberlain 1997; Edvardsen & Imai 2006). However, their small cell size means that they are difficult to study. Species identification often requires electron microscopy and a high degree of taxonomic expertise. This is especially true for non-calcifying haptophytes, whereas coccolithophorids have calcified scales (coccoliths), which are easier to preserve and observe under the optical and scanning electron microscopes. Therefore knowledge on haptophyte diversity, distribution and abundances at the species level remains fragmentary. The effect of environmental factors and climate change on populations of this important phytoplankton group is also to a large extent unclear. However, the use of molecular methods, such as environmental sequencing and metabarcoding has revealed large haptophyte diversity in the ocean that was previously unsuspected. By environmental sequencing we hereafter mean PCR amplification of marker genes (amplicons) in natural plankton samples followed by clone library construction and Sanger sequencing of specific clones. In the metabarcoding approach, the clone library construction is omitted and sequencing of amplicons is performed by high-throughput sequencing (HTS, also called next generation sequencing or NGS). Metabarcoding with the 454 Life Sciences technology is also called pyrosequencing. In contrast, metagenomics refers to whole-genome sequencing of environmental samples.

In genetic surveys of plankton communities, different size fractions of the community are often separated by filtration in order to avoid the dominance by bigger organisms with larger genomes and high number of gene copies per genome (Zhu et al. 2005). These studies have revealed high, unexpected haptophyte diversity within the pico- (0.2–2 μm) and nano-plankton (2–20 μm) size ranges. Most of these sequences do not have a close similarity to cultured species (Unrein et al. 2013; Egge et al. 2015a). They may represent