environmental microbiology

Environmental Microbiology (2017) **00**(00), 00–00 doi:10.1111/1462-2920.13952

Comparison of coastal phytoplankton composition estimated from the V4 and V9 regions of the 18S rRNA gene with a focus on photosynthetic groups and especially Chlorophyta

Margot Tragin,[1] Adriana Zingone[2] and Daniel Vaulot[1]*

[1]*Sorbonne Universités, UPMC Univ Paris 06, UMR 7144, CNRS, Station Biologique, Place Georges Teissier, 29680 Roscoff, France.*

[2]*Department of Integrative Marine Ecology, Stazione Zoologica Anton Dohrn, Villa Comunale, Naples, Italy.*

Summary

We compared the composition of eukaryotic communities using two genetic markers (18S rRNA V4 and V9 regions) at 27 sites sampled during Ocean Sampling Day 2014, with a focus on photosynthetic groups and, more specifically green algae (Chlorophyta). Globally, the V4 and V9 regions of the 18S rRNA gene provided similar images of alpha diversity and ecological patterns. However, V9 provided 20% more OTUs built at 97% identity than V4. 34% of the genera were found with both markers and, of the remnant, 22% were found only with V4 and 44% only with V9. For photosynthetic groups, V4 and V9 performed equally well to describe global communities at different taxonomic levels from the division to the genus and provided similar Chlorophyta distribution patterns. However, at lower taxonomic level, the V9 dataset failed for example to describe the diversity of Dolichomastigales (Chlorophyta, Mamiellophyceae) emphasizing the lack of V9 sequences for this group and the importance of the reference database for metabarcode analysis. We conclude that in order to address questions regarding specific groups (e.g., a given genus), it is necessary to choose the marker based not only on the genetic divergence within this group but also on the existence of reference sequences in databases.

Received 10 April, 2017; accepted 30 September, 2017. *For correspondence. E-mail: vaulot@sb-roscoff.fr; Tel. (+33) 2 98 29 23 23; Fax (+33) 2 98 29 23 24

Introduction

Planktonic organisms are distributed throughout all branches of the tree of life (Baldauf, 2008) but share 'universal' genes presenting certain degrees of genetic variability, which allow them to be used as barcode markers to investigate biological diversity (Chenuil and Anne, 2006). The development of high-throughput sequencing (HTS) allows the acquisition of large metabarcoding datasets (i.e., one marker gene is amplified and sequenced for all organisms), which complement the time-consuming and expertise-demanding morphological inventories to explore the diversity and distribution of protist groups in the ocean. The 18S rRNA gene is commonly used to investigate eukaryotic diversity and community structures (López-García *et al.*, 2001; Moon-van der Staay *et al.*, 2001). The complete 18S rRNA gene (around 1,700 base pairs) from environmental clone libraries can only be sequenced by the Sanger method (Sanger and Coulson, 1975) using a combination of primers. In contrast, HTS provides a very large number of reads but allows only small fragments to be sequenced (van Dijk *et al.*, 2014). Small hypervariable regions of the 18S such as V9 (around 150 bp located near the end of the 18S rRNA gene) or V4 (around 450 bp in the first half of the gene) can be targeted depending on the sequence length allowed by the sequencing technology used. Initially, the Illumina technology only allowed to sequence the V9 region because of its relatively small size (Amaral-Zettler *et al.*, 2009). In recent years longer reads became possible (up to 2 $\times$ 300 bp with current Illumina technology, van Dijk *et al.*, 2014) allowing the sequencing of the V4 region. Both the V4 and V9 regions have been used recently to describe diversity and ecological patterns of protists in several large scale studies (Massana *et al.*, 2014; de Vargas *et al.*, 2015).

The performance of the 18S RNA hypervariable regions as barcodes and the interpretation of results produced remain a matter of debate. Hu and colleagues (2015) showed that the V4 region provides an image of diversity similar to that obtained from the entire 18S rRNA gene. The choice between V4 and V9 depends on the taxonomic