

The most transcriptionally active eukaryotic plankton in the surface oceans: who is doing what, and where?

Daniel J. Richter^{1,2}, Jürgen F. H. Strassert³, Nicolas Henry², Fabien Burki³ and Colomán de Vargas²

1. Institut de Biologia Evolutiva (CSIC-UPF), Barcelona, Spain

2. Sorbonne Universités, UPMC Univ Paris 06, CNRS UMR 7144, Adaptation et Diversité en Milieu Marin, Équipe EPEP, Station Biologique de Roscoff, 29680 Roscoff, France

3. Science for Life Laboratory, Program in Systematic Biology, Uppsala University, Uppsala, Sweden

The *Tara* Oceans expedition circumnavigated the globe over a three-year period, collecting samples in the sunlit ocean from over 150 individual stations. For eukaryotic plankton ranging from 0.8 μm -2 mm, we apply phylogenetic methods to *Tara* metatranscriptomic sequence data in order to produce a summary of microbial eukaryote transcriptional activity in the world's surface oceans. First, we map *Tara* Oceans metatranscriptomic sequencing reads against a reference database of phylogenetic trees for a set of 250 conserved, constitutively expressed eukaryotic genes. Next, we combine mapping data from individual gene trees onto a single taxonomically-labeled species tree, with each branch in the eukaryotic tree of life weighted by its transcriptional activity. We find that, globally, metatranscriptomes show a similar representation of major eukaryotic lineages to estimates derived from *Tara* metabarcodes (based on the V9 hypervariable region of the 18S ribosomal locus), indicating that both data sources likely reflect the biomass of active cells in the surface ocean. Within the smaller size fractions (pico- and nanoplankton, 0.8-5 and 5-20 μm), we confirm the unexpectedly high relative abundance of dinoflagellates, and we find a substantial representation of transcriptionally active metazoans. To demonstrate the application of our phylogenetic placement method to functional genes, we study the biogeographic expression patterns of the silicon transporter (SIT), which transports environmental Si across the cell membrane to be used in the production of biosilicified structures (such as diatom frustules). We find that diatoms dominate SIT expression in the Antarctic, where Si concentrations are highest, whereas choanoflagellates and copepods perform the bulk of SIT expression in the northern Indian Ocean and the Mediterranean Sea, where Si concentrations are much lower. We propose that our catalog of eukaryotic gene expression in the surface oceans can be used to understand the ecology and biology of microbial eukaryotes in global pelagic ecosystems.