Customizing sequence databases for system level analysis of gene expression in marine microbial plankton

Gutierrez Jayson and Hablützel Pascal

Flanders Marine Institute (VLIZ), Wandelaarkaai 7, 8400 Oostende, Belgium
E-mail: jayson.gutierrez@vliz.be

High-throughput sequencing of environmental samples is significantly improving our understanding of the molecular activities of complex microbial communities in their natural environment. For instance, by enabling taxonomic profiling and genome-wide differential gene expression analysis, meta-omics analyses have shed new light on the relationship between community structure and ecosystem functions. However, the ability to accurately mapping the activity of physiological pathways in natural microbial ecosystems from high-throughput sequencing data is constrained by the quality and completeness of reference sequence databases.

Here, we applied well-established bioinformatics tools to leverage publicly available genome sequences for creating a custom planktonic database suitable for a systems biology-oriented analysis of environmental samples using the popular HUMAnN workflow, which has proven effective at reverse-engineering molecular phenotypes putatively involved in ecosystem functions. To test the effectiveness of database customization, we reanalyzed previously published metratranscriptomics datasets and demonstrate that significant improvements can be gained in terms of mapping dominant metabolic activities that may be key drivers of community behavior.

By further expanding on the taxonomic and functional complexity of our draft database with newly released high-quality genome assemblies for marine microbes, the hope is to increasingly improve our ability to map the molecular traits that drive changes in marine microbial community composition and function through space and time.

Keywords: Systems biology; Meta-omics analysis; Metabolic pathway; Marine plankton; Molecular database