

DNA sequence-based identification of cyanobacteria and microalgae in aquatic biomonitoring: strengths and weaknesses

Supporting files

This dataset has been prepared as part of the Interreg Alpine Space project Eco-AlpsWater (ASP569) - *Innovative Ecological Assessment and Water Management Strategy for the Protection of Ecosystem Services in Alpine Lakes and Rivers*, <https://www.alpine-space.eu/projects/eco-alpswater/en/home>

Individual archives include 16S rRNA (cyanobacteria) and 18S rRNA (microalgae) FASTA sequences and associated blastn results obtained from the high throughput sequencing of plankton and biofilm eDNA samples collected in 2019 in 37 lakes and 22 rivers across the Alpine region (Salmaso et al., submitted).

A) FASTA sequences

Amplicon sequence variants (ASVs) and taxonomic classifications (SILVA 138 and PR2) have been obtained from FASTQ files (Illumina MiSeq) using DADA2 bioinformatic pipelines, following the protocols described in Salmaso et al. (2021a, 2021b). The raw FASTQ files have been deposited to the European Nucleotide Archive (ENA) with study accession numbers PRJEB49047 (16S rRNA genes) and PRJEB49184 (18S rRNA genes). In DADA2, the Illumina runs were processed independently using the same function parameters, and the output tables merged later.

File name	Ribosomal marker	Region	Habitat	N. ASVs
fastacyano_eawf2_PL.fasta	16S rRNA	V3-V4	Plankton	364
fastacyano_eawf2_BFM.fasta	16S rRNA	V3-V4	Biofilm	2619
fastaprot_Microalgae_PL_eawf2.fasta	18S rRNA	V4	Plankton	2836
fastaprot_Microalgae_BFM_eawf2.fasta	18S rRNA	V4	Biofilm	3742

B) blastn analyses

blastn queries on the NCBI database performed on a subset of “nt” sequences

File name	Blastn query	Habitat
EAW_Cyano_blastn_query_genbank_Plankton.tsv	nt, txid1798711	Plankton
EAW_Cyano_blastn_query_genbank_Biofilm.tsv	nt, txid1798711	Biofilm
EAW_EukAlgae_blastn_query_genbank_Plankton.tsv	nt, 18S rRNA, SSU and “small subunit”	Plankton
EAW_EukAlgae_blastn_query_genbank_Biofilm.tsv	nt, 18S rRNA, SSU and “small subunit”	Biofilm

Legend (see also NCBI, 2020)

Column name	Description
Seqquery	ASVs number (Seq number)
Sacc	subject accession
Staxids	unique Subject Taxonomy ID(s)

Sscinames	unique Subject Scientific Name(s)
scomnames	unique Subject Common Name(s)
pident	percentage of identical matches
nident	number of identical matches
mismatch	number of mismatches
gaps	total number of gap
length	alignment length
bitscore	bit score
evaluate	expect value
ReadsBac / ReadsEuk	Reads abundance (number of sequences for each ASV)
FreqBac / FreqEuk	Frequency (number of occurrences)
Family	NCBI taxonomy
Order	NCBI taxonomy
Class	NCBI taxonomy
Phylum	NCBI taxonomy
Kingdom	NCBI taxonomy
superkingdom	NCBI taxonomy

C) References

NCBI, 2020. BLAST® Command Line Applications User Manual. National Center for Biotechnology Information (US), Bethesda (MD).

Salmaso, N., Vasselon, V., Rimet, F., Vautier, M., Elersek, T., Boscaini, A., Donati, C., Moretto, M., Pindo, M., Riccioni, G., Stefani, E., Capelli, C., Lepori, F., Kurmayer, R., Mischke, U., Krivograd Klemenčič, A., Novak, K., Greco, C., Franzini, G., Fusato, G., Giacomazzi, F., Lea, A., Menegon, S., Zampieri, C., Macorm, A., Virgilio, D., Zanut, E., Zorza, R., Buzzi, F., Domaizon, I. (submitted). DNA sequence-based identification of cyanobacteria and microalgae in aquatic biomonitoring: strengths and weaknesses.

Salmaso, N., Riccioni, G., Pindo, M., Kurmayer, R., Vasselon, V., Domaizon, I., 2021a. Metabarcoding protocol – Analysis of protists using the 18S rRNA gene and a DADA2 pipeline. Zenodo.
<https://doi.org/https://doi.org/10.5281/zenodo.5233527>

Salmaso, N., Riccioni, G., Pindo, M., Vasselon, V., Domaizon, I., Kurmayer, R., 2021b. Metabarcoding protocol – Analysis of Bacteria (including Cyanobacteria) using the 16S rRNA gene and a DADA2 pipeline. Zenodo.
<https://doi.org/https://doi.org/10.5281/zenodo.5232772>