Interreg Alpine Space

Co-funded by the European Union

DiMark

WP1 – In situ eDNA observations (D.1.2.1)

Biotic markers from plankton samples (all available eDNA molecular data) will be systematically gathered and evaluated in a manner that data from A.1.1, A.1.2 and A.1.3 will be compared in A.2.1.

Name of Author(s)	[Rainer Kurmayer, Loay Hussein]
Department/Organization/Institution	[Research Department for Limnology, University of Innsbruck]
Date of the document	[09/02/2025]
Version	[01]
Project	DiMark

Page two begins here, with a header and footer that differ from the main page's design.

Introduction

This report is about the activity 1.2 organising the collection of relevant eDNA or metabarcoding results from water bodies sampled during previous projects, respectively on algal plankton communities. Since all algae contain Chlorophyll a as the primary pigment for photosynthesis the abundance can be monitored through Chlorophyll a absorption as recorded through the satellite spectrum. This report is about 16 S and 18 S rDNA deep sequencing results representing the phytoplankton communities and informing about plankton algal composition. Since the same samples are also monitored by satellite Sentinel 3 data the aim is to provide complementary information for using the Sentinel 3 data. The available environmental metadata include additional parameters like chlorophyll a and transparency (secchi depth).

The general experience from the previous alpine space project is that overall a good qualitative relationship between HTS (16S rDNA reads) and biovolume estimates of phytoplankton (cyanobacteria) has been observed. Notably small sized picocyanobacteria are more reliably detected through sequencing than by counting using the inverted microscope. Thus using light microscopy picocyanobacteria are typically understimated while through sequencing such a bias in biovolume because of the small size can be prevented. Overall also good quantitative relationships between HTS (reads) and (LM) biovolume have been observed, i.e. (semi) quantitative conclusions seem possible.

Metabarcoding data sets

For this report most importantly the data set from the Eco Alps Water (EAW) project , which was the previous alpine space project (2018-2021) has been used (https://www.alpine-space.eu/project/eco-alpswater/). The sampling comprises 36 different lakes located in Austria, France, Germany, Italy, Slovenia and Switzerland. Out of these 36 different lakes, 16 lakes have been sampled repeatedly. The total number of samples is n=151 (Fig. 1). One can see that the pilot key lakes in the Dimark project are at least partly overlapping with the former EcoAlpswater Project (Mondsee, Lake Bled, Lake Garda) as well as some of the additional lakes (Faakersee, Starnbergersee, ...?).

As reported earlier the mixing type of the lakes sampled is monomictic (52%), dimictic (36%), meromictic (8%) and rarely polymictic (4%). During the limnological sampling period in 2019, the temperature of water at sampling campaigns was ranging from 3 to 30°C, with conductivity from 16 to 588 µS/cm. The trophic status of the lakes has been assessed by three parameters: total phosphorus, transparency and chlorophyll-a concentration and analyzed with OECD fixed boundary trophic classification system (OECD; 1982). Thus trophic status of lakes sampled in EAW was ranging from ultra-oligotrophic to eutrophic conditions (Ref, Perspectives in eDNA monitoring in alpine waters, EcoAlpsWater).

In addition the following metabarcoding data might become available in the course of this project:

1) 16S rDNA data from partner AGES on bathing water sites (6 sites, n=142). Notably this data cannot be monitored by Sentinel 3 because of the small surface areas. Thus

Sentinel 2 data with higher spatial resolution would have to be used which is currently not planned within the Dimark project.

2) 16s rDNA data from Lake Constance has been mentioned by partner UKON, which has been anticipated about one year ago, and which also could expand the Dimark data set.
3) Finally, we have 50-80 phytoplankton samples from NIB from Slovenian lakes, which would also expand the data set.

Reprocessing of 16S and 18S DNA sequences for taxonomic assignment (FEM)

The available sequences were reprocessed for taxonomic assignment with the newest reference database through partner FEM. Then the 16S rDNA data set was reduced to contain only cyanobacteria and chloroplasts (indicating eukaryotic algae). Finally, the total dataset contains 4429 amplicon sequence variants left, which are equivalent to unique genotypes. The Bayesian assignments with 95% probability was used to assign taxonomy to the lowest possible taxonomic level, i.e. species<genus<family<order), reference database SILVA 138.1. For plankton a total of 151 lake samples has been processed. Then these lake samples were sorted according to the most abundant taxa. We used the absolute read numbers for now, and calculated percentages. Since the total read numbers vary quite substantially (from 361 per sample to more than 40,000 pe sample), the rarefied data set using a standardized sequence read number should be tested as well.

Regarding plankton samples only a few families of cyanobacteria occur: Most importantly the family Cyanobiaceae representing so called picocyanobacteria, i.e. usually unicellular cells with rather small cell size occurring in high numbers (10^3 – 10^4 cells/ml) assigned to the genera Cyanobium, Synechococcus by sequencing or to Aphanocapsa or Aphanothece by the microscope. The second most abundant family comprises large filamentous cyanobacteria of the genus Planktothrix, while the other families occurred less frequently (Limnothrichaceae, Prochlorotrichaceae, Microcystaceae). Genera of Nostocaceae also occurred in plankton samples including taxa such as Dolichospermum or Aphanizomenon.

Then the taxa are ranked by read numbers (in relative terms) and compared with regard to the top no1 taxonomic assignment (Fig. 3). Notably the proportional rate of top no1 taxon is quite high as can be inferred from the range given in minimum-median-maximum %abundance. Since the median is 51% it can be concluded that for half of all samples the top no1 taxon contributes 51% or even a higher share. Thus for the half of the samples the top no1 ranked taxa were dominating. Notably for taxa ranked no2 or 3 the relative abundance dropped substantially. While for no2 taxon the median decreased to 7%, for no3 taxon it decreased further to 1%. Thus for no2 or no3 taxon half of the samples contributed 7% or 1% only.

Accordingly the majority of plankton samples (more than 50%) have picocyanobacteria (Cyanobium) as the most abundant taxon and also as the 2nd abundant taxon. Cyanobium as the 3rd abundant taxon still occurs in 1/3 of all plankton samples. Aside from Cyanobium about a quarter of plankton samples showed the filamentous cyanobacterium Planktothrix ranked no1, as well as no2 and partly as no3. Notably those samples were either dominated by Cyanobium or by Planktothrix. The cyanobacterium Snowella was ranked no 1,2,3 in a smaller number of

samples, while other taxa occurred more rare. In summary the taxonomic composition remained stable through ranking no1 or no2 or no3.

The dominance of either picocyanobacteria or Planktothrix could provide complementary information to interpret the satellite signals for Chlorophyll a. Since picocyanobacteria occur in much higher cell numbers but lower Chlorophyll content per cell size one could expect a lower but more homogeneous signal within the band width for Chlorophyll. In contrast the much larger and macroscopically visible filaments of Planktothrix might introduce more scatter within the band width used for Chlorophyll reading.

The commonly used term »algae« is defining a lifeform that thrived on earth during the last two billions of years. According to evolutionary theory more than a billion years ago an ancient cyanobacterium was ingested by a eukaryotic organism (a so called protozoan) through phagotrophy and ingested into the food vacuole. The cyanobacterium was not digested but cultivated as an intracellular symbiont which evolved to an organell during millions of years and is now called chloroplast. Probably such endosymbiosis events occurred repeatedly leading to the modern eukaryotic algae groups. Since 16S rDNA is a conserved molecule it can be used to trace this macroevolutionary processes and readily discriminate free-living cyanobacteria from chloroplasts. Within chloroplasts primary endosymbiosis can be diferentiated from different secondary endosymbiosis events and sometimes even allow for taxonomic assignment at the genus level.

Perhaps not surprisingly, a negative correlation between the top one ranked cyanobacteria taxon and the percentage of reads assigned to chloroplasts was observed. In other words according to 16S rDNA relative numbers the phytoplankton was either dominated by free-living cyanobacteria or by cyanobacteria serving as chloroplasts in eukaryotic algae (Fig. 4). Interestingly the samples showing more than 50% of the reads assigned to chloroplast did not occurr randomly but were confined to the following lakes (n = 45): Anterne > Bourget > Vogresek > Staffelsee_Sud > Staffelsee_Nord > Lugano > Starnberger_See > Bohinj > Pernica > Mondsee > Garda > Slivnica. Many of these lakes were sampled repeatedly during the year of 2019 suggesting a rather seasonally consistent dominance of chloroplasts over cyanobacteria. In contrast samples showing <10% of sequence reads assigned to chloroplasts originated from (n = 29): Pernica > Bled > Ledro > Lugano > Mondsee > Maggiore > Bourget > Bohinj > Frassino > Garda > Varese. In summary the lake samples seem to divide into cyanobacteria dominated or eukaryotic algae dominated which again could be useful information to better inpret Chlorophyll band widths in satellite signals.

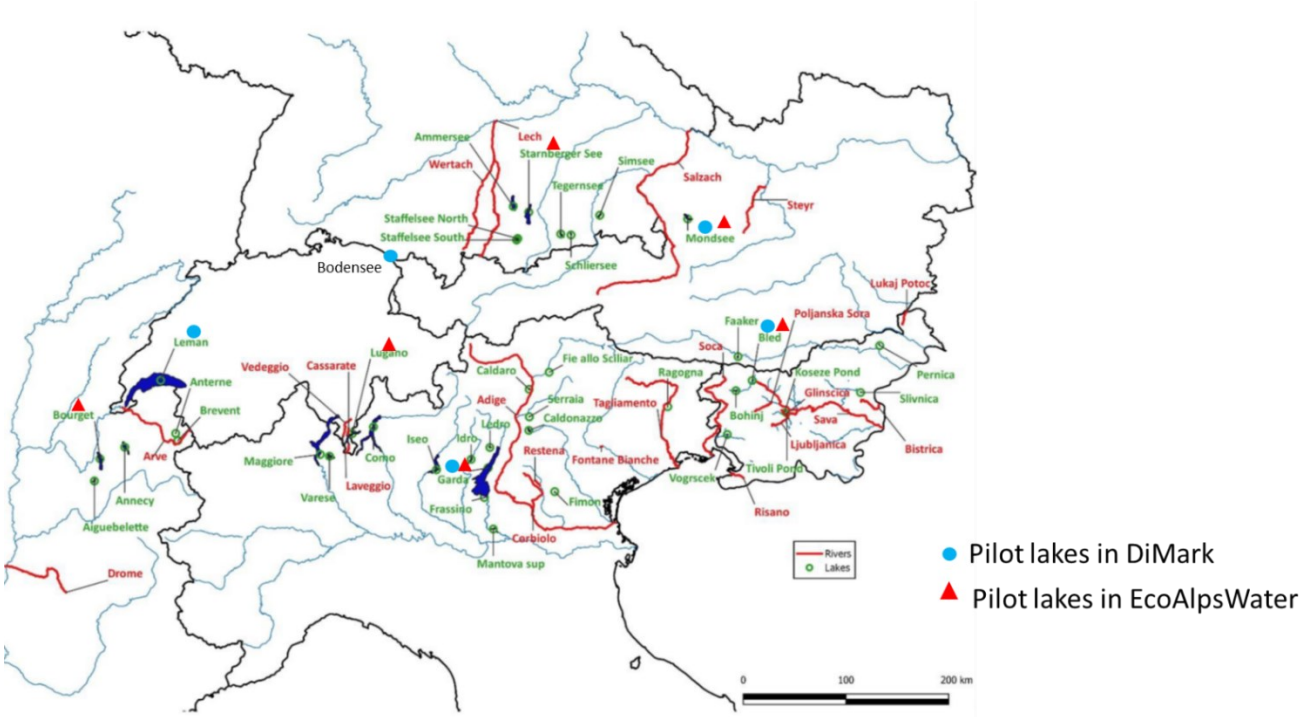

Fig. 1. Sampling map showing lakes (and rivers) sampled through the alpine space project EcoAlps-Water (2018-2021). Note that for Dimark only plankton samples from 36 lakes have been used (n=151). From Salmaso et al. 2024, Hydrobiologia

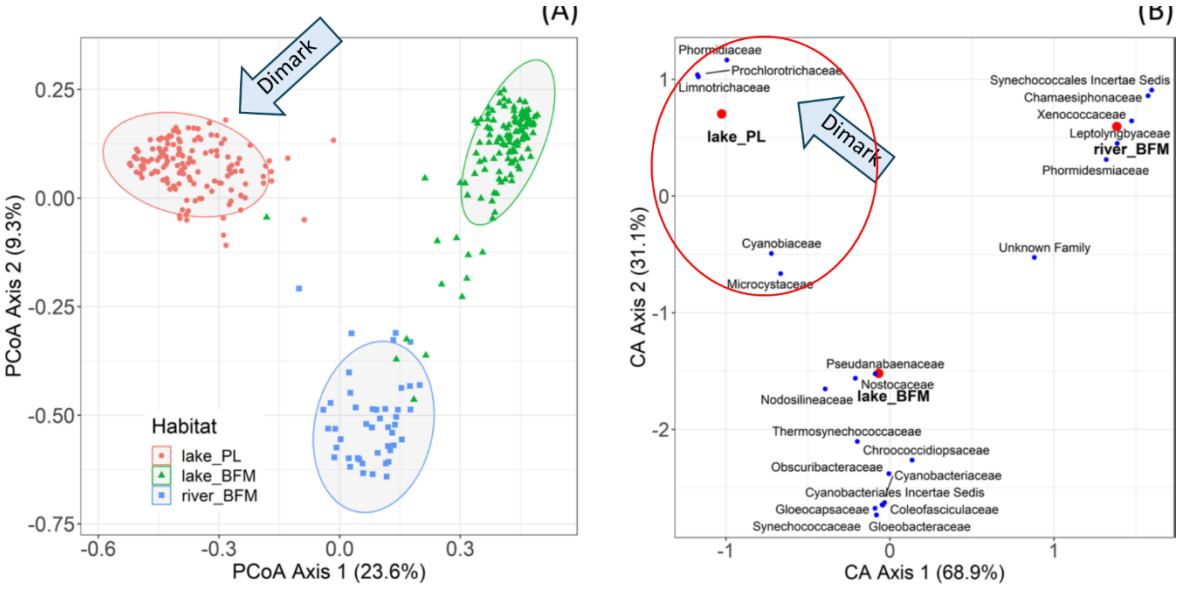

Fig. 2. Cyanobactera family composition in lakes (and rivers) sampled through the alpine space project EcoAlps-Water (2018-2021). Note that for Dimark only plankton samples (red circle) from 36 lakes have been used (n=151). From Salmaso et al. 2024, Hydrobiologia

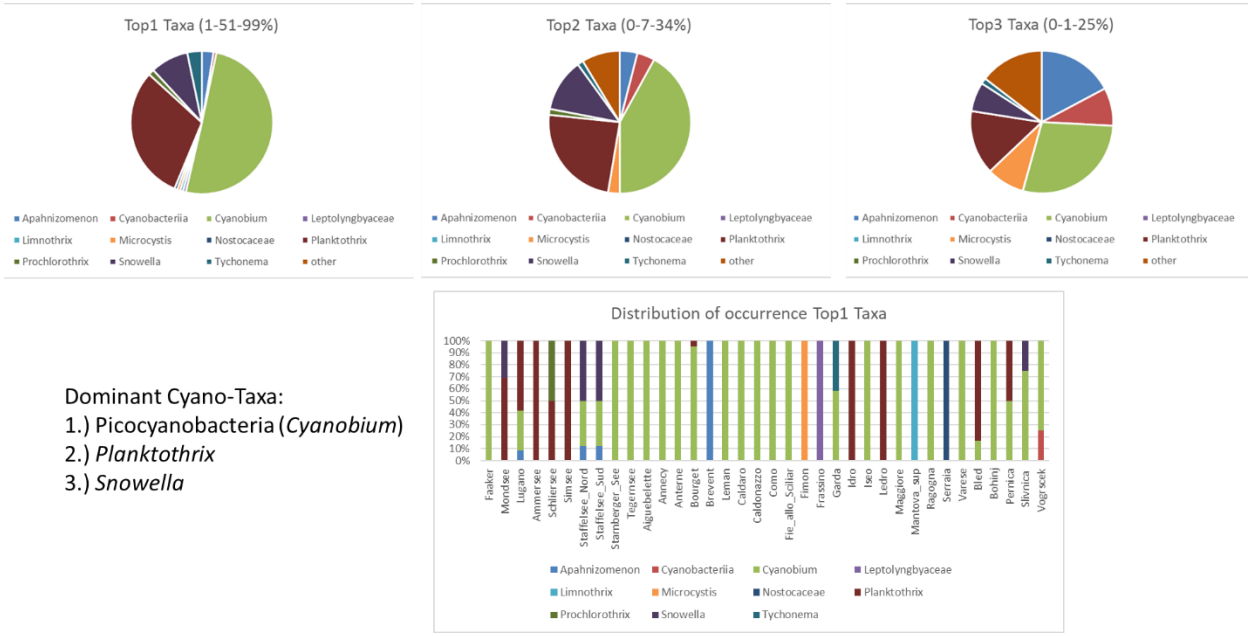

Fig. 3. Top three cyanobacteria Taxa in EAW lakes (min-med-max % abundance of sequence reads) in lakes sampled through the alpine space project EcoAlps-Water (2018-2021). Note that for Dimark only plankton samples from 36 lakes have been used (n=151). From Salmaso et al. 2024, Hydrobiologia

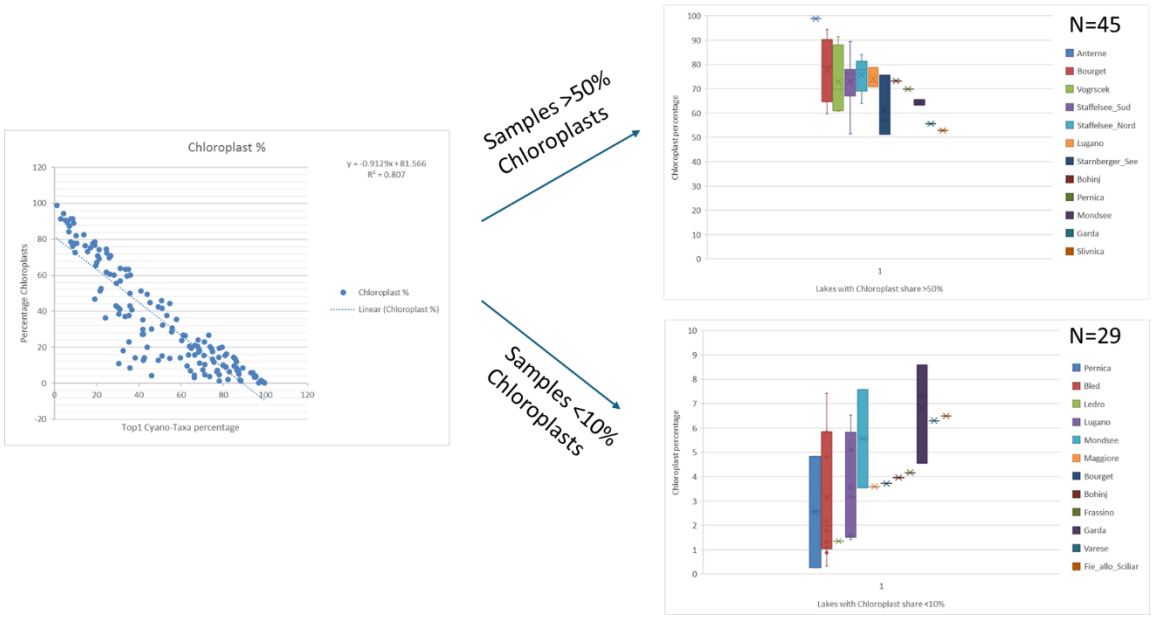

Fig. 4. Relationship between free living cyanobacteria vs eukaryotic algae (as indicated by chloroplasts). The distribution of samples sharing >50% of reads assigned to chloroplast has been indicated using bar plots for individual lakes (n = 45), upper inset. Analagously, the distribution of samples sharing < 10% of reads assigned to chloroplasts is shown for lakes (n = 29), lower inset.

Regarding the eukaryotic algae the next question would be on the composition. One option would be to use the 16S rDNA information to assign eukaryotic algal taxa. The other option is to sequence 18S rDNA from nucleus of eukaryotic algae. Analogously to 16S rDNA this 18S rDNA data set had to be reduced to contain only the phototrophic organisms containing chlorophyll. i.e. the phototrophic divisions of Alveolata, Chlorophyta, Cryptophyta (Cryptophyta:nucl), Glaucophyta, Haptophyta, Rhodophyta, Stramenopiles, and Streptophyta. Finally a reduced data set comprising 7876 ASVs or genotypes was obtained which was taxonomically assigned using the protist rreference database (PR2 version 5.0.0) and a Bayesian probability of 95%. This 18S rDNA data set has been again sorted according to the most abundant taxa using both the absolute numbers and the calculated percentages (note: we also should test the results using the rarefied data set, Nico could you please provide this as well?).

Compared with 16S rDNA the relative share of 18S sequences for taxa ranked top no1 is lower, i.e. more than half of all samples contained 24-70% of total read numbers. For taxa ranked no2 the median was 13% while for taxa ranked no3 the median was still 9%. Thus for 18S rDNA the decline in relative abundance was not as steep as for 16SrDNA. Interestingly about half of the lakes were dominated by cryptophytes composed of one taxon which is Cryptomonas. Cryptomonas occurred quite consistently including all three ranks taxa no1, no2 and no3. The second most abundant algal group comprised the chrysophytes comprising the genera Dinobryon , Epipyxis, Malllomonas, Uroglenopsis, etc. The third most abundant algal group includes dinophytes with genera Ceratium, Gymnodinium, Gyrodinium, etc. The algal group of diatoms including Taxa like Aulacoseira, Cyclotella, Fragilaria, Ulnaria and the order of Thalassiosirales is the fourth most abundant algal group. Surprrisingly green algae in general occurred rarely. In other words neither Chlorophyta sensu strictu nor Streptophyta including desmids and Zygnemales were abundant (Fig. 5).

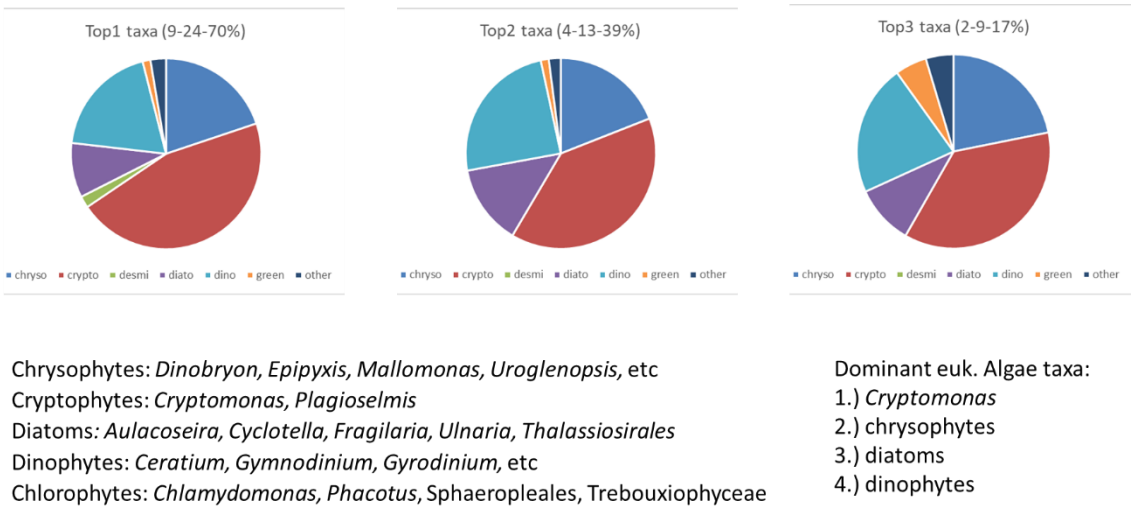

Fig. 5. Top three eukaryotic algae taxa in EAW lakes (min-med-max % abundance of sequence reads) in lakes sampled through the alpine space project EcoAlps-Water (2018-2021). Note that for Dimark only plankton samples from 36 lakes have been used (n=151). From Salmaso et al. 2022, Science of The Total Environment

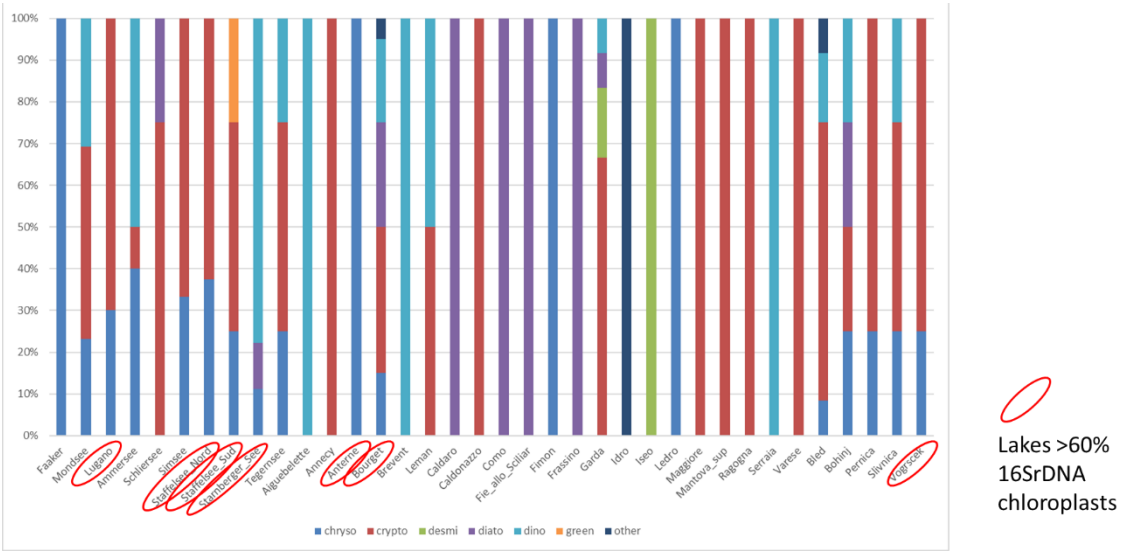

Fig. 6. Distribution of occurrence of eukaryotic algae ranked top1 (9-24-70% abundance of sequence reads)in EAW lakes in lakes sampled through the alpine space project EcoAlps-Water (2018-2021). Note that for Dimark only plankton samples from 36 lakes have been used (n=151). From Salmaso et al. 2022, Science of The Total Environment

Regarding the distribution of the eukaryotic algae groups for the ranked top one taxa, the question is whether we can see a pattern comparable to the 16S rDNA. The answer is no, basically it is either cryptophytes or chrysophytes or dinophytes which iss dominating and no relationship with samples showing >60% of chloroplast for 16S RDNA could be found (Fig.6).

In summary, the 16 S rDNA data can be used to characterize the phytoplankton composition and also help to interpret the satellite signals. For example to differentiate between the dominance of microscopic picocyanobacteria vs. macroscopic filamentous cyanobacteria and it would be interesting to see whether differences in satellite signal according to this this cyanobacteria community composition could be found. In addition it might be helpful to differentiate between lakes either dominated by cyanobacteria or by eukaryotic algae. For the latter the band widths from satellites should be broader because of co-occurring Chlorophyll a and c structural variants. Fro the 18S rDNA data we conclude a rather high proportion of cryptomonads.

In the case that phycobilins would be analysed from satellites as well it would be important to know that the phycoerythrin signal could originate from both cryptomonads or cyanobacteria. Because of the small proportion of green algae the role of chlorophyll b structural variants would be expected negligible.

The remaining steps for the EAW data set would include rarefying the samples resulting in equal sample sizes for both 16S and 18S rDNA and assigning chloroplast genotypes taxonomically using reference databases (PhytoREF database). Finally when additional data sets become available a broader pespective also including smaller more eutrophic lakes will be possible.

Acknowledgements

The project team of the EcoAlpsWater project supporting sampling and analysis: Project leader: Nico Salmaso (IT), Rainer Kurmayer (AT), Giorgio Franzini (IT), Tina Elersek (SI), Aleksandra Krivograd Klemenčič (SI), Isabelle Domaizon (FR), Michael Schubert (DE), Peter Hufnagl (AT), Claudia Greco (IT), Ute Mischke (DE), Jean Marc Baudoin (FR), Camilla Capelli (CH).

References

Kurmayer, R., T. Eleršek, I. Domaizon, C. Capelli, U. Mischke, G. Riccioni, L. Cerasino, S. Bernabei, C. Greco, C. Martone, M. Insolvibile, P. Tomassetti, F. Grassi, A. Giorgio & N. Salmaso, 2021. Perspectives on eDNA monitoring in Alpine waters. Lubljana, SI, 53. https://www.alpine-space.eu/project/eco-alpswater/

Jablonska, M., L. Cerasino, A. Boscaini, C. Capelli, C. Greco, A. Krivograd Klemenčič, U. Mischke, N. Salmaso & R. Kurmayer, 2024. Distribution of toxigenic cyanobacteria in Alpine lakes and rivers as revealed by molecular screening. Water Research 258:121783 doi:https://doi.org/10.1016/j.watres.2024.121783.

Salmaso, N., S. Bernabei, A. Boscaini, C. Capelli, L. Cerasino, I. Domaizon, T. Elersek, C. Greco, A. Krivograd Klemenčič, P. Tomassetti & R. Kurmayer, 2024. Biodiversity patterns of cyanobacterial oligotypes in lakes and rivers: results of a large-scale metabarcoding survey in the Alpine region. Hydrobiologia 851(4):1035-1062 doi:10.1007/s10750-023-05423-z.

Salmaso, N., V. Vasselon, F. Rimet, M. Vautier, T. Elersek, A. Boscaini, C. Donati, M. Moretto, M. Pindo, G. Riccioni, E. Stefani, C. Capelli, F. Lepori, R. Kurmayer, U. Mischke, A. K. Klemenčič, K. Novak, C. Greco, G. Franzini, G. Fusato, F. Giacomazzi, A. Lea, S. Menegon, C. Zampieri, A. Macor, D. Virgilio, E. Zanut, R. Zorza, F. Buzzi & I. Domaizon, 2022. DNA sequence and taxonomic gap analyses to quantify the coverage of aquatic cyanobacteria and eukaryotic microalgae in reference databases: Results of a survey in the Alpine region. Science of The Total Environment 834:155175 doi:https://doi.org/10.1016/j.scitotenv.2022.155175.