FONDAZIONE EDMUND MACH

Interreg Alpine Space Priority 3: Liveable Alpine Space. SO3.2: Enhance the protection, the conservation and the ecological connectivity of Alpine Space

Interreg Alpine Space Eco-AlpsWater European Regional Development Fund

EUROPEAN UNION

Seminario di presentazione del progetto Eco-AlpsWater
Sede ISPRA, Sala del Consiglio Federale, Roma, 16 ottobre 2019

Eco-AlpsWater

Innovative Ecological Assessment and Water Management Strategy for the Protection of Ecosystem Services in Alpine Lakes and Rivers

Formalizzazione di protocolli per l'analisi del DNA ambientale in laghi e fiumi

Nico Salmaso

Coordinatore del progetto Eco-AlpsWater

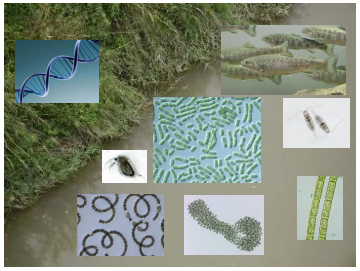
IASMA Research and Innovation Centre
Fondazione Mach-Istituto Agrario di S. Michele all'Adige
Unità di ricerca Idrobiologia
nico.salmaso@fmach.it – 0461 615323

Where were we?

- Formalization of protocols for the analysis of environmental DNA in lakes and rivers
- Steps to be considered in the development of protocols for the analysis of eDNA
- Examples of applications of the metagenomic approach to complement algal biodiversity assessment in the large lakes south of the Alps

Interreg Alpine Space Eco-AlpsWater
European Regional Development Fund

Output from Next Generation Sequencing e.g. Illumina MiSqeq sequencing



DNA extraction

PCR

Sequencing

Bioinformatic pipelines
FASTQ → FASTA

400-430 bp

Reference reads data

```
>DENOVO1 (Tychonema)
TGGGGAATTTTCCGCAATGGGCGAAAGCCTGACGGAGCAAGACCGCGTGGGGGAAGAAGGCTCTTGGGTTGTAAACTCCTT
TTCTCTGGGAAGAACAAAATGACGGTACCAGAGGAATCAGCATCGGCTAACTCCGTGCCAGCAGCCGCGGTAAGACGGAGG
ATGCAAGCGTTATCCGGAATGATTGGGCGTAAAGCGTCCGCAGGTGGCAGTTCAAGTCTGCTGTCAAAGACCGGGGCTCAA
CCTCGGAAAGGCAGTGGAAACTGAACAGCTAGAGTATGGTAGGGGCAAAGGGAATTCCTGGTGTAGCGGTGAAATGCGTAG
AGATCAGGAAGAACATCGGTGGCGAAGGCGCTTTGCTGGACCATAACTGACACT AGGGACGAAAGCTAGGGGAGCGAATG
>DENOVO2 (Family Anaerolineaceae)
TAGGGAATATTGGTCAATGGGCGAAAGCCTGAACCAGCAACGCCGCG AGGCCTTCGGGTCGTAAAGCGCTT
TTGGGAGGGATGAAATTGACAGTACCTCCCGAATAAGGATCGGC GCAGCCGCGGTAAGACGTAGGATC
CAAGCGTTATCCGGAATTACTGGGCGTAAAGGGCGTGTAC TCGGCCATGAAAGCTCCCGGCTCAACTG
GGAGAGGCTGGTCGATACTGCCTGGCTAGAGGGCAA ATTCCCGGTGTAGTGGTGAAATGCGTAGATA
TCGGGAGGAACACCAGTGGCGAAGGCGGCCTCC ACTCTGAAACGCGAAAGCATGGGGAGCAAACA
>DENOVO3 (Synechococcus)
TGGGGAATTTTCCGCAATGGGCGCAA CGCCGCGTGAGGGATGAAGGCCTCTGGGCTGTAAACCTCTT
TTATCAAGGAAGAAGATCTGAC AAGCCACGGCTAATTCCGTGCCAGCAGCCGCGGTAATACGGGAGT
GGCAAGCGTTATCCGGAATTA GCGTCCGCAGGCGGTTTTACAAGTCTGTCGTTAAAACGTGGAGCTCAAC
TCCATTTCGGCGATGGAAACTG AGAGTGTGGTAGGGGCAGAGGGAATTCCCGGTGTAGCGGTGAAATGCGTAGA
TATCGGGAAGAACACCAGTGGCG GCGCTCTGCTGGGCCATAACTGACGCTCATGGACGAAAGCCAGGGGAGCGAAAG
>DENOVO4 (Cyanobacteria;Chloroplast)
TAGGGAATTTTCCGCAATGGGCGAAAGCCTGACGGAGCAATACCGCGTGAGGGATGACGGCCTGTGGGTTGTAAACCTCTT
TTCTCAAGGAAGAAGTTCTGACGGTACTTGAGGAATAAGCATCGGCTAACTCTGTGCCAGCAGCCGCGGTAATACAGAGGA
TGCAAGCGTTATCCGGAATCACTGGGCATAAAGCGTCTGTAGGTGGTTTGGTAAGTCTGCTGTTAAAGACTGGGGCTCAAC
CCCAGAAAAGCAGTGGAAACTGCCAGACTTGAGTGTGGTAGAGGTAGAGGGAATTCCTAGTGTAGCGGTGAAATGCGTAGA
TATTAGGAAGAACACCAATGGCGAAGGCACTCTACTGGACCATAACTGACACTGAGAGACGACAGCTAGGGGAGCAAATG
(….)
(….)
```

After the application of the bioinformatic pipeline to FASTQ files, besides the table with the reads, many other information are provided and/or integrated in one unique information system

Interreg Alpine Space Eco-AlpsWater
European Regional Development Fund

Output after the application of a bioinformatic pipeline: up to 5 interlinked tables

Reference reads data

>DENOVO1 (Tychonema)
TGGGGAATTTTCCGCAATGGGCGAAAGCCTGACGGAGCAAGACCGCGTGGGGGAAGAAGGCTCTTGGGTTGTAAACTCCTTTTCTCTGGGAAGAACAAAATGACGGTAC
CAGAGGAATCAGCATCGGCTAACTCCGTGCCAGCAGCCGCGGTAAGACGGAGGATGCAAGCGTTATCCGGAATGATTGGGCGTAAAGCGTCCGCAGGTGGCAGTTCAAG
TCTGCTGTCAAAGACCGGGGCTCAACCTCGGAAAGGCAGTGGAAACTGAACAGCTAGAGTATGGTAGGGGCAAAGGGAATTCCTGGTGTAGCGGTGAAATGCGTAGAG
ATCAGGAAGAACATCGGTGGCGAAGGCGCTTTGCTGGACCATAACTGACACTCAGGGACGAAAGCTAGGGGAGCGAATG
>DENOVO2 (Family Anaerolineaceae)
TAGGGAATATTGGTCAATGGGCGAAAGCCTGAACCAGCAACGCCGCGTGCGCGATGAAGGCCTTCGGGTCGTAAAGCGCTTTTGGGAGGGATGAAATTGACAGTACCTC
CCGAATAAGGATCGGCTAACTACGTGCCAGCAGCCGCGGTAAGACGTAGGATCCAAGCGTTATCCGGAATTACTGGGCGTAAAGGGCGTGTAGGAGGTTGGGCAAGTCG
GCCATGAAAGCTCCCGGCTCAACTGGGAGAGGCTGGTCGATACTGCCTGGCTAGAGGGCAAGAGAGGGAGGTGGAATTCCCGGTGTAGTGGTGAAATGCGTAGATATCG
GGAGGAACACCAGTGGCGAAGGCGGCCTCCTGGCTTGTACCTGACTCTGAAACGCGAAAGCATGGGGAGCAAACA

Taxonomy table

	"Kingdom"	"Phylum"	"Class"	"Order"	"Family"	"Genus"	"Species"	"OTUID"
"Seq1"	"Bacteria"	"Chloroflexi"	"Anaerolineae"	"Anaerolineales"	"Anaerolineaceae"	NA	NA	NA
"Seq3"	"Bacteria"	"Proteobacteria"	"Alphaproteobacteria"	"SAR11_clade"	"Clade_III" NA	NA	NA	
"Seq4"	"Bacteria"	"Proteobacteria"	"Gammaproteobacteria"	"Betaproteobacteriales"	"Hydrogenophilaceae"	NA	NA	NA
"Seq5"	"Bacteria"	"Actinobacteria"	"Actinobacteria"	"Frankiales"	"Sporichthyaceae"	NA	NA	NA
"Seq7"	"Bacteria"	"Proteobacteria"	"Gammaproteobacteria"	"Betaproteobacteriales"	"Burkholderiaceae"	NA	NA	NA
"Seq8"	"Bacteria"	"Actinobacteria"	"Acidimicrobiia"	"Microtrichales"	"Ilumatobacteraceae"	NA	NA	NA

OTU/ASVs abundance tables (n. reads)

	Seq1	Seq2	Seq3	Seq4	Seq5	Seq6	Seq7	Seq9	Seq10	Seq11	Seq12
Bar0114D1	35756	2141	1220	2738	0	0	191	46	282	0	254
Bar0114D2	36593	2373	2191	2011	0	0	304	132	685	0	239
Bar0114D3	38022	2406	1646	3630	0	0	190	107	299	0	225
B0214D1	32275	753	1518	1783	0	0	52	161	1206	0	257
B0214D2	27096	699	1285	2448	0	0	98	0	817	0	283
B0214D3	37702	1219	1833	2786	0	0	230	154	881	0	291
B0314D1	70	5345	7346	7386	0	0	667	166	303	0	620

Sample variables and environmental data

ID	sampleID	date	station	depth	month1	year	season	temp	O2	O2sat
113-GARDA-18S-Bar0114D1	Bar0114D1	14/01/2014	bardolino	1	1	2014	winter	9.961411672	9.97	89
114-GARDA-18S-Bar0114D2	Bar0114D2	14/01/2014	bardolino	10	1	2014	winter	9.967345283	10.70	95
115-GARDA-18S-Bar0114D3	Bar0114D3	14/01/2014	bardolino	20	1	2014	winter	9.96626	10.55	94
116-GARDA-18S-B0214D1	B0214D1	11/02/2014	brenzone	1	2	2014	winter	9.125112308	10.87	95
117-GARDA-18S-B0214D2	B0214D2	11/02/2014	brenzone	10	2	2014	winter	9.1181	10.87	95
118-GARDA-18S-B0214D3	B0214D3	11/02/2014	brenzone	20	2	2014	winter	9.117487097	11.08	97
119-GARDA-18S-B0314D1	B0314D1	11/03/2014	brenzone	1	3	2014	winter	9.930163158	11.80	105

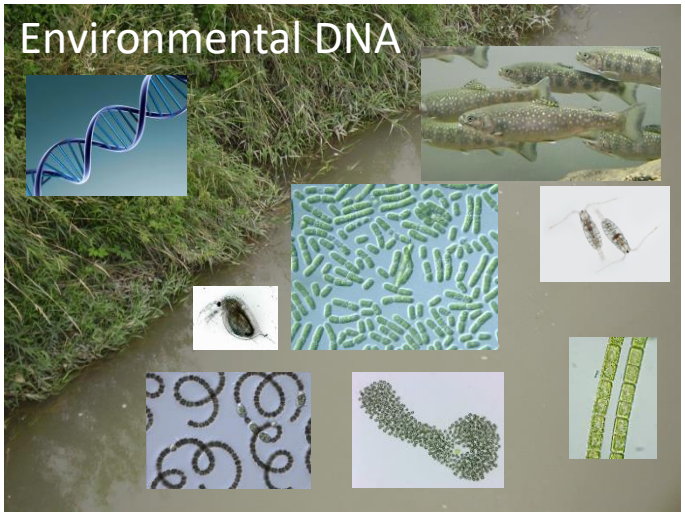
Phylogenetic tree

Formalization of protocols for the analysis of environmental DNA in lakes and rivers

A few considerations based on the pros and cons of new approaches (**focused on marker gene amplification metagenomics**)

- One universal protocol for the assessment of biodiversity of eukaryotes, eubacteria and archeobacteria does not exist
- Different kingdoms or even different groups require specific protocols (e.g. bacteria, protists, diatoms, macroinvertebrates, fishes)
- Different biological elements require specific protocols taking into account all the different steps of the analysis, from sampling (characteristic of target biota) to sequencing (e.g. primers) and bioinformatic analyses
- At present, these protocols are still under active development, even within the Eco-AlpsWater Consortium
- A few protocols are however emerging as a family of «standard», especially in microbial ecology (e.g. diatoms, using *rbcL* genes)

Marker gene amplification metagenomics (metagenetics or targeted metagenomics), is a high-throughput sequencing (HTS) application focusing on a nucleotide target (e.g. 16S rRNA genes) to describe the taxonomic content of a sample

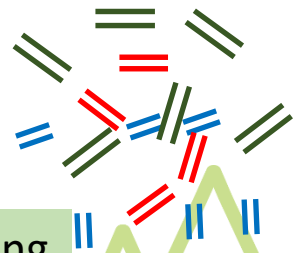


1) Collect water sample

2) Extract environmental DNA

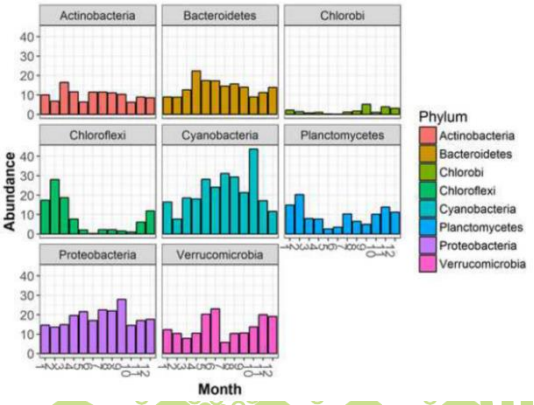
3.1) Amplify target genes with PCR (e.g. 16S, 18S *rbcL*, ITS)

3.2) Sequencing



```
FASTQ
@M00380:72:000000000-AMNE0:1:1101:16943
GGGGGGGGGGGGGGGGGGGGFCFEGGGGGGGGGGGGFGG
TCGCTCCCCTAGCTTTCGTCCATGAGCGTCAGTTATGGC
###9:@D,##:4BDFFGFGG#9CDFGF,C#39FFFGF+C
TCTACGCATTTCACCGCTCCACCTGAAGTTCCACATGCC
GGGGGGGFDGFC?F9CFDCFGFGGFB#3@@DGF#6#36<
```

4) Taxonomic inventories and Community analysis



Interreg Alpine Space Eco-AlpsWater
European Regional Development Fund

Steps to be considered in the development of protocols for the analysis of eDNA

1.1. Sampling strategy

1.2. Filtering of water/collection of biofilms

1.3. Conservation of samples/biological material

→ 1) FIELD ACTIVITY (Previous presentation)

2.1. DNA extraction and conservation

→ 2) WET LAB ACTIVITY

3.1. Selection of primers for amplification

3.2. Sequencing

→ 3) SEQUENCING FACILITY

4.1. Selection bioinformatic approaches (OTUs - ASVs)

4.2. Taxonomic classification

4.3. Downstream statistical analysis

→ 4) BIOINFORMATIC ANALYSES AND DOWNSTREAM STATISTICAL ANALYSES

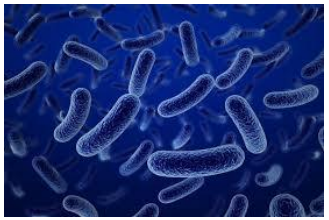
1) FIELD ACTIVITY

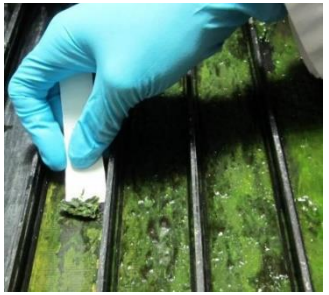
1.1) Sampling strategy

- The dimensions, life-forms, target eDNA material of biological elements have to be taken into account when developing and implementing protocols.

- eDNA material can be represented by single cells/units (microbes, protists, zooplankton), cellular materials/residues (macroinverterbates, fish), or biofilms.

- Samples can be represented by low volumes (0.2-2 L) for bacteria and protists, or higher volumes, up to several liters for fish.





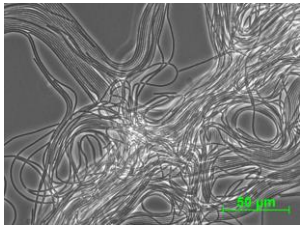
Interreg Alpine Space Eco-AlpsWater
European Regional Development Fund

1) FIELD ACTIVITY

1.2. Filtering of water/collection of biofilms

For microbes, different strategies can be adopted:

- Prefiltering (e.g. eliminate fraction > 10 µm), and then filter at 0.22 µm
- Differential filtration
- Filtration and analysis of specific dimensions
- Filtration of whole water samples (clogging!)
- Vertical filtration or Sterivex filters



Filtration of water for other biological elements

- Based on same pore size filters used for microbial communities, or even larger pore size filters (e.g. for fish even 0.45-0.8 µm filters are used)

Collection of fish samples and collection of biofilms (diatoms) are the object of new protocols prepared by the EAW Consortium. Instead, standard approaches for bacteria / cyanobacteria and protists (published or in preparation) will be used.

Lake plankton sample collection from the field for downstream molecular analysis

Interreg Alpine Space - Eco-AlpsWater project – WP1

Isabelle Domaizon[1], Rainer Kurmayer[2], Camilla Capelli[3], Cécile Chardon[1], Peter Hufnagl[4], Marine Vautier[1], Nico Salmaso[5].

[1] INRA, CARRTEL, Thonon les bains, France
[2] University of Innsbruck, Research Institute of Limnology, Innsbruck, Austria
[3] University of Applied Sciences Southern Switzerland, SUPSI, Cannobio, Switzerland
[4] Austrian Agency for Health and Food Safety, Vienna, Austria
[5] Edmund Mach Foundation, San Michele all'Adige, Trento, Italy

Interreg Alpine Space Eco-AlpsWater
European Regional Development Fund

1) FIELD ACTIVITY

1.3. Conservation of samples/biological material

- Refrigeration of water samples at 4°C for short time storage
- Freezing at -20 °C or -80°C of filters
- Use of chemical preservatives

2) WET LAB ACTIVITY

2.1. DNA extraction and conservation

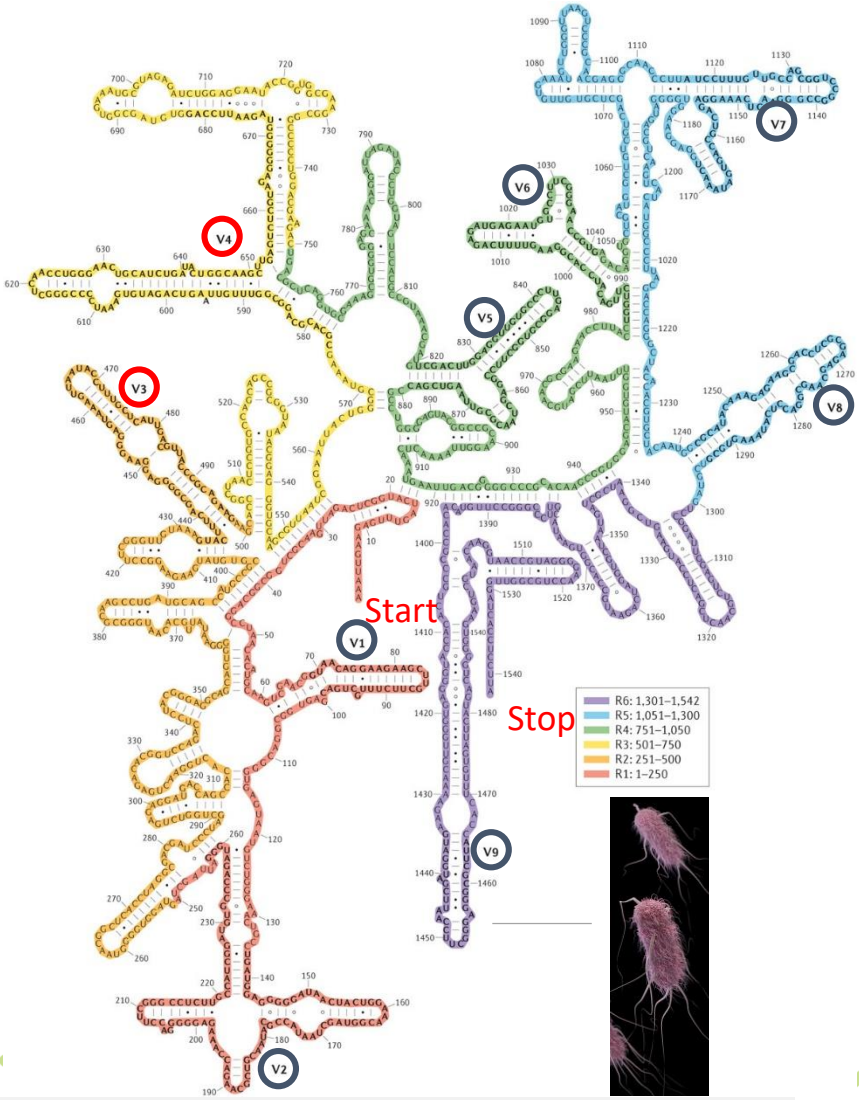
- Classical laboratory methods (cell lysis, deproteinization with phenol-chloroform, and precipitation with alcohol, e.g. isopropanol, ethanol)
- Use of commercial kits, e.g. QIAGEN DNeasy PowerWater
- Storage of eDNA, at -20°C or -80°C

Protocols with detailed instructions for sending biological materials and eDNA samples to partners responsible of DNA estraction and/or sequecing have been prepared by the EAW Concortium

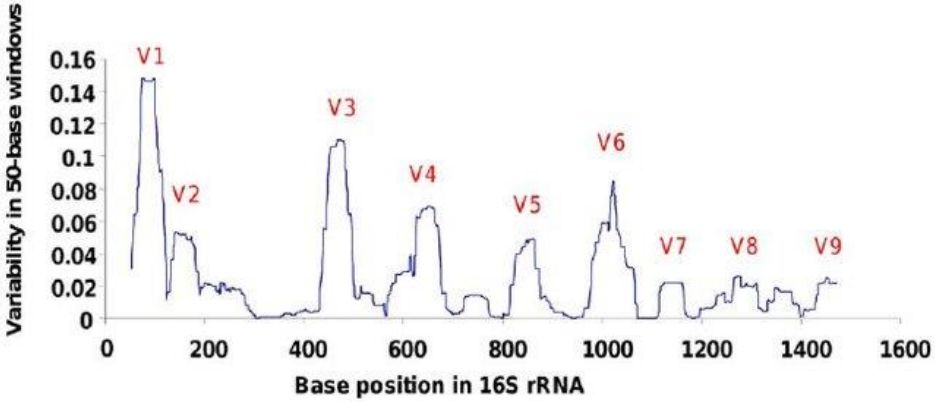
Interreg Alpine Space Eco-AlpsWater
European Regional Development Fund

3) SEQUENCING FACILITY

3.1. Selection of primers for amplification



- Selection has to be done for the different target genes of interest
- For example, for **16S rDNA**, selection is done within 1-2 hypervariable regions



- Present technologies generally allow sequencing of reads up to 300F + 300R reads

Rate of 16S rRNA divergence of 1–2% per 50 million years

Interreg Alpine Space Eco-AlpsWater
European Regional Development Fund

3) SEQUENCING FACILITY

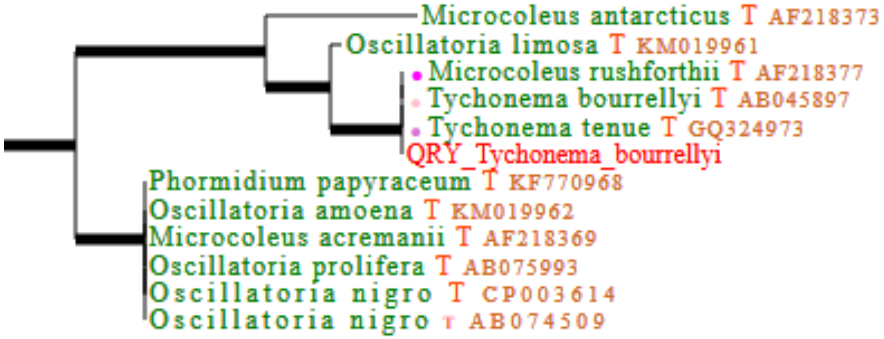
3.1. Selection of primers for amplification

V3-V4 region

>ENA LM651410 Tychonema bourrellyi NIVA-CYA 33/1 partial 16S rRNA gene

...AGTGGCGGACGGGTGAGTAACGCGTGAGAATCTACCTTCAGGACGG
AGACAACAGTTGGAAACGACTGCTAACCCCCGATGTACCGARAGGGAA
AATATTTATAGCCTGAAGAAGAGCTCGCGTCCGATTAGCTAGTTGGAGA
GGTAAAAGCTCACCAAGGCGACGATCGGTAGCTGGTCTGAGAGGACG
ATCAGCCACACTGGGACTGAGACACGGCCCAGACTCCTACGGGAGGC
AGCAGTGGGGAATTTTCCGCAATGGGCGAAAGCCTGACGGAGCAAGA
CCGCGTGGGGGAAGAAGGCTCTTGGGTTGTAAACTCCTTTTCTCTGGG
AAGAACAAAATGACGGTACCAGAGGAATCAGCATCGGCTAACTCCGTG
CCAGCAGCCGCGGTAAGA**405 nt**CAAGCGTTATCCGGAATGATT
GGGCGTAAAGCGTCCGCA TCAAGTCTGCTGTCAAAGACC
GGGGCTCAACCTCGGAAAGGCAGTGGAAACTGAACAGCTAGAGTATG
GTAGGGGCAAAGGGAATTCCTGGTGTAGCGGTGAAATGCGTAGAGATC
AGGAAGAACATCGGTGGCGAAGGCGCTTTGCTGGACCATAACTGACAC
TCAGGGACGAAAGCTAGGGGAGCGAATGGGATTAGATACCCCAGTAGT
CCTAGCCGTAAACGATGGATACTAGGTGTTGTCTGTATCGACCCGGAC
AGTGCCGTAGCTAACGCGTTAAGTATCCCGCCTGGGGAGTACGCACGC
AAGTGTGAAACTCAAAGGAATTGACGGGGGCCCGCACAAGCGGTGGA
GTATGTGGTTTAATTCGATGCAACGCGAAGAACCTTACCAGGACTTGAC
ATGTCGCGAATCYTTTTGAAAGAGARGAGTGCCTTAGGGAGCGCGAAC
ACAGGTGGTGCATGGCTGTCGTCAGCTCGTGTCGTGAGATGTTGGGTT
AAGTCCCGCAACGAGCGCAACCCTCGTGTTTAGTTGCCATCATTAAGTT
GGGAACTCTAAACAGACTGCCGGTGACAAACCGGAGGAAGGTGGGGA
TGACGTCAAGTCAGCATGCCCCTTACGTCCTGGGCTACACACGTACTA
CAATGGTAGGGACAGAGGGCAGCCAACTCGCGAGAGAGAGCTAATCC
CGTAAACCCTGCCTCAGTTCAGATTGCAGGCTGCAACTCGCCTGCATG
AAGGCGGAATCGCTAGTAATCGCAGGTCAGCATACTGCGGTGAATCCG
TTCCCGGGCCTTGTACACACCGCCCGTCACACCATGGAAGTTGGCCAC
GCCCGAAGTCATTACTCTAACCCTTCGGGGAGGAGGATGCCGAAGGCA
GGGCTGATGACTGGGGTGAAGTCGTAACAAGGTAGCCGTACCGGAAG
GTGTGGCTGGATCACCTCCTTTTAGGGAGACCATCTGAC...

Using the 16S rRNA genes, the transition to new HTS technologies provided more quantitative information at the expense of taxonomic resolution (SHORT READS....)



3) SEQUENCING FACILITY

Amplification by Illumina MiSeq

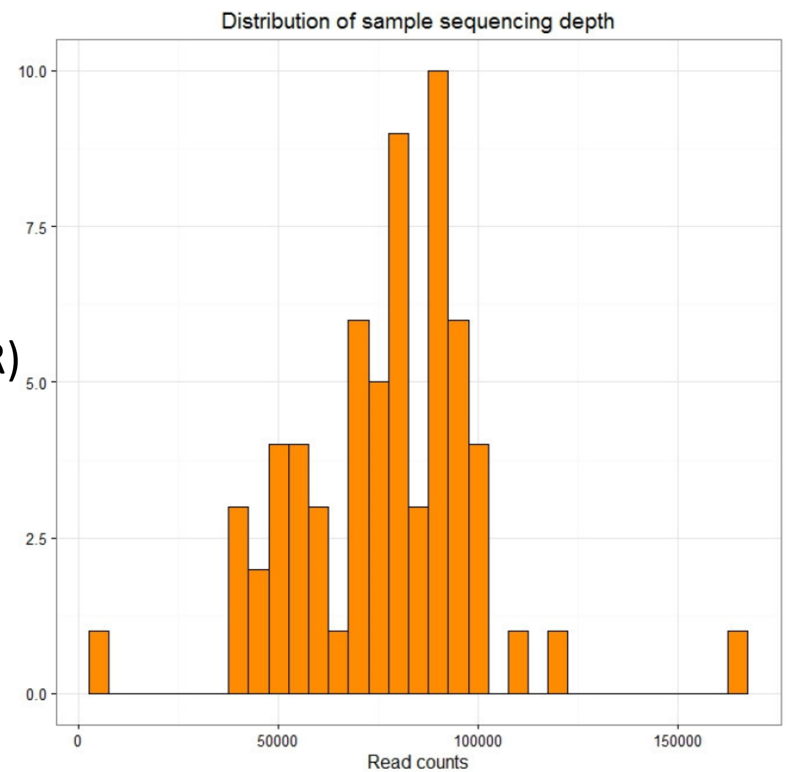
3.2. Sequencing

- In EAW, barcoded library is sequenced on Illumina® MiSeq platforms
- Selection of sequencing depths (costs!)

Example: Massive amplification of 16S rRNA from the samples collected in 2014 and 2015 (64 samples) in Lake Garda produced over 7.000.000 reads

FASTQ files (F&R)

```
1 @M00380:72:000000000-AMNE0:1:1101:12817:1299 1:N:0:35
CCTACGGGAGGCAGCAGTGGGGAATCTTGCGCAATGGGCGAAAGCCTGACGCAGCGACGCCGCGTGGGGGAAGAAGGCCTTAGGGTTGTAA
+
88BCCGGFDBA85=====CFFFGGGGGFGGGGGGGGGGGGGGGDGGGGG8F@EFGGGGGGGGGGGGEFGGGGGGGGDGGGGGGGGGGGGG
2 @M00380:72:000000000-AMNE0:1:1101:14526:1314 1:N:0:35
CCTACGGGCGGCTGCAGTAACGAATCTTCCGCAATGCGCGAAAGCGTGACGGGGTAATGCCGCGTGCAGGATGAAGCCCTTCGGGGTGTAA
+
8ACCCG6*88=5886;;C;ECC@CFGFGEEFDGGECFEGCFGGCEFEG7FGGGGFEFFECFEFGGGGDFGGGGGGFAEGGFGG@FEF@FFF
3 @M00380:72:000000000-AMNE0:1:1101:15372:1352 1:N:0:35
CCTACGGGTGGCAGCAGTAGGGAATATTGGTCAATGGGCGAAAGCCTGAACCAGCAACGCCGCGTGCGCGATGAAGGCCTTCGGGTCGTAA
+
68ACCGGFB@)8:8====CFFGGGGGGGGGGGGEFGGGGDGGGDGGGGGGGGGGGDGGGGGGGGGDGGGGGGGGGGGGGGGFGGGGGGGGG
4 @M00380:72:000000000-AMNE0:1:1101:19512:1372 1:N:0:35
CCTACGGGTGGCTGCAGTGGGGAATATTGGGCAATGGAGGAAACTCTGAACCAGCGACGCCGCGTGCGGGACGAAGGCCTTCGGGTTGGAA
+
8BCCCGGFFFF5====CCF,CFGGGGGGGG,FFGGGGGFGEGGFGGGGG<FGGGG7F+FGGGGGGGGCGGG+CF7FGG8,BB<C7?F++,C
5 @M00380:72:000000000-AMNE0:1:1101:21601:1401 1:N:0:35
CCTACGGGAGGCAGCAGTGGGGAATTTTCCGCAATGGGCGCAAGCCTGACGGAGCAACGCCGCGTGAGGGATGAAGGCCTCTGGGCTGTAA
+
8ACCCGGEFFFA======CFGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGDFFGGDFGGGGGGGGCFGGGGGGGGG
6 @M00380:72:000000000-AMNE0:1:1101:20747:1412 1:N:0:35
CCTACGGGGGGCAGCAGTGGGGAATATTGGACAATGGGCGCAAGCCTGATCCAGCCATGCCGCGTGAGTGATGAAGGCCTTAGGGTTGTAAAGCTCTTTTGTCCGGGACGATAATGACGGTACCGGAAGAATAAGCCCCGGCTAACTTCGTGCCAGCAGCCGCGG
+
ACCCCGGEAF=8:88::<CFFGGGGEGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGG>FGGFAFGGGFGGGGGGGGGGGGGGGGGGGGGGGGGGGG
7 @M00380:72:000000000-AMNE0:1:1101:21591:1419 1:N:0:35
CCTACGGGAGGCAGCAGTGGGGAATTTTCCGCAATGGGCGCAAGCCTGACGGAGCAACGCCGCGTGAGGGATGAAGGCCTCTGGGCTGTAAACCTCTTTTATCAAGGAAGAAGATCTGACGGTACCTGATGAATAAGCCACGGCTAATTCCGTGCGAGCAGCCGC
+
CCCCCGEB9@BA:=====;CEFFGG<FGCC@FGGGGGGGGDGGGGGGGGF@@CCFGGGGGGGEGGFGG8CFGGGGGGGGGGEGGG,4BFGCE,?4E?EFGE,5C<FGGFGGGDFG9?BF<?78B+,<,,,8<FF,A?CFCFB++@=,7@,:@,+3+8C>F<8DF@
```



4) BIOINFORMATIC ANALYSES AND DOWNSTREAM STATISTICAL ANALYSES

4.1. Selection bioinformatic approaches (OTUs - ASVs)

2 main available strategies

OTU (operational taxonomic units)

Sequences are clustered according to their similarity, and OTUs are identified based on an arbitrary similarity threshold (usually 97% similarity). Owing to errors in DNA sequencing, the number of OTUs may be inflated.
Tools: QIIME1, USEARCH, MOTHUR, MICCA and others…

ASVs (Amplicon Sequence Variants)

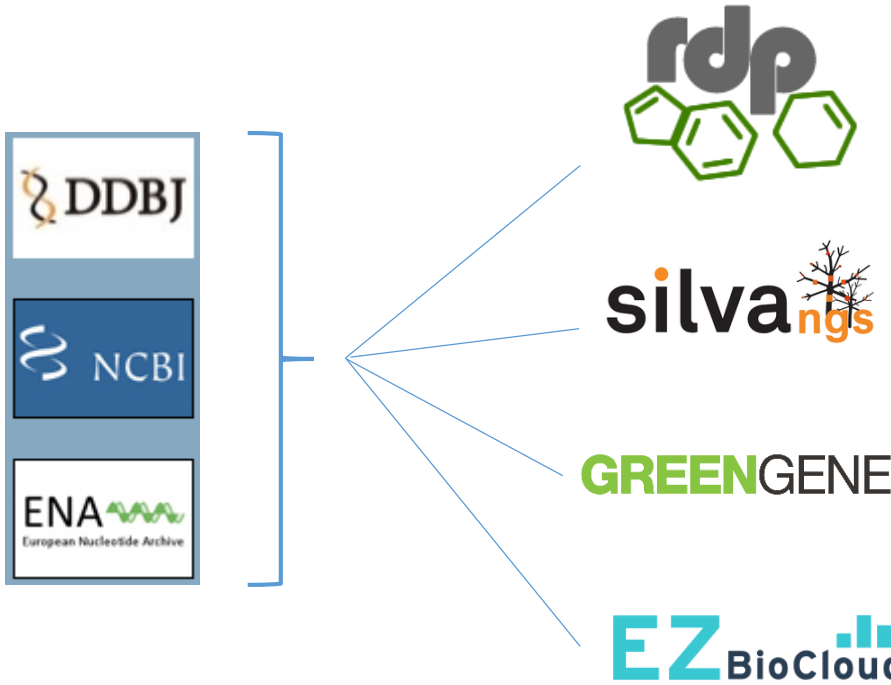
Exact amplicon sequence variants (ASVs) are ingerred from amplicon data, resolving biological differences of even 1 or 2 nucleotides. The ASVs output can be directly compared between studies, without the need to reprocess the pooled data.
Tools: DADA2, DEBLUR, UNOISE2, QIIME2 (uses either DADA2 or DEBLUR)

De novo OTUs cannot be compared across samples. Viceversa, exact sequences ASVs are comparable across samples.

4) BIOINFORMATIC ANALYSES AND DOWNSTREAM STATISTICAL ANALYSES

4.2. Taxonomic classification

- Based on comprehensive curated databases, which includes information on the main target molecular markers.



The Ribosomal Database Project (RDP) Bacteria and Fungi

16S/18S SSU and 23S/28S LSU ribosomal RNA (rRNA) sequences for all three domains of life (Bacteria, Archaea and Eukarya).

16S rRNA gene database (2013 - no more updated)

16S rRNA gene database and genomes

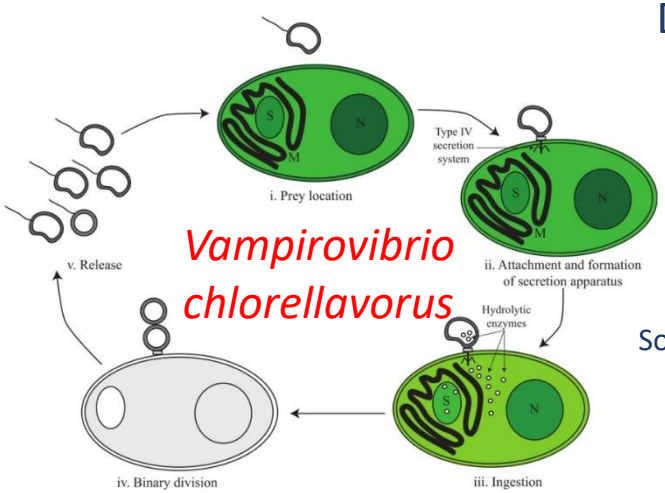
- Other more specialized databases can be used (e.g. for protists, PR2, or diatoms, rbcL…)

Interreg Alpine Space Eco-AlpsWater
European Regional Development Fund

4) BIOINFORMATIC ANALYSES AND DOWNSTREAM STATISTICAL ANALYSES

4.2. Taxonomic classification

- The taxonomic databases used in environmental metagenomics include also information on species not isolated and not described (metagenome-assembled genomes - MAGs)



Vampirovibrio chlorellavorus

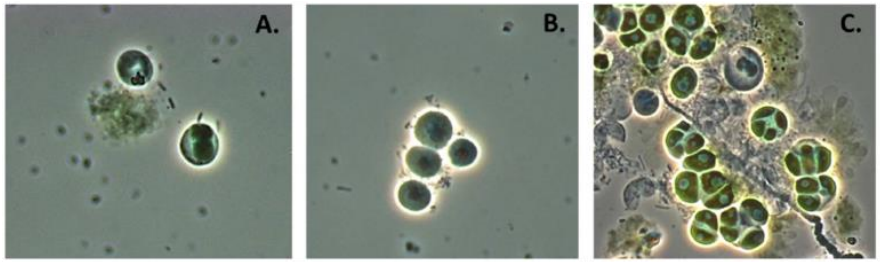
Soo et al., 2015

Discovery of new non-photosynthetic cyanobacterial groups
Up to now…only one culturable species

- Predator of *Chlorella vulgaris*
- Non-photosynthetic flagellate
- Previously classified to Proteobacteria

silva high quality ribosomal RNA databases

Genome	Order	Class
		Oxyphotobacteria
Vampirovibrio chlorellavorus	*Gastranaerophilales*	*Melainabacteria*
	Vampirovibrionales	
	Obscuribacterales	
	Caenarcaniphilales	
LMEP_6097	*SHAS531*	
SSGW_16	*V201-46*	
CBMW_12	*S15B-MN24*	*Sericytochromatia (ML635J-21)*
RAAC_196		
LSPB_72	*GL2-53*	



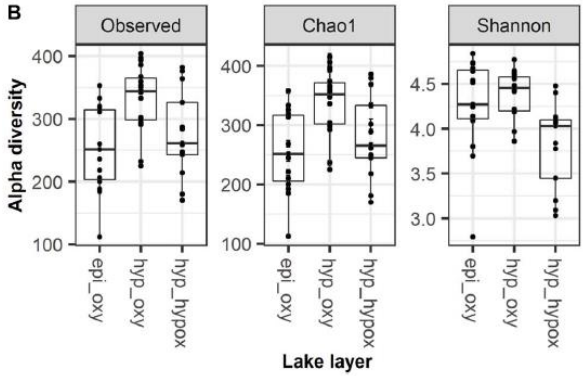
Bagwell et al., 2016

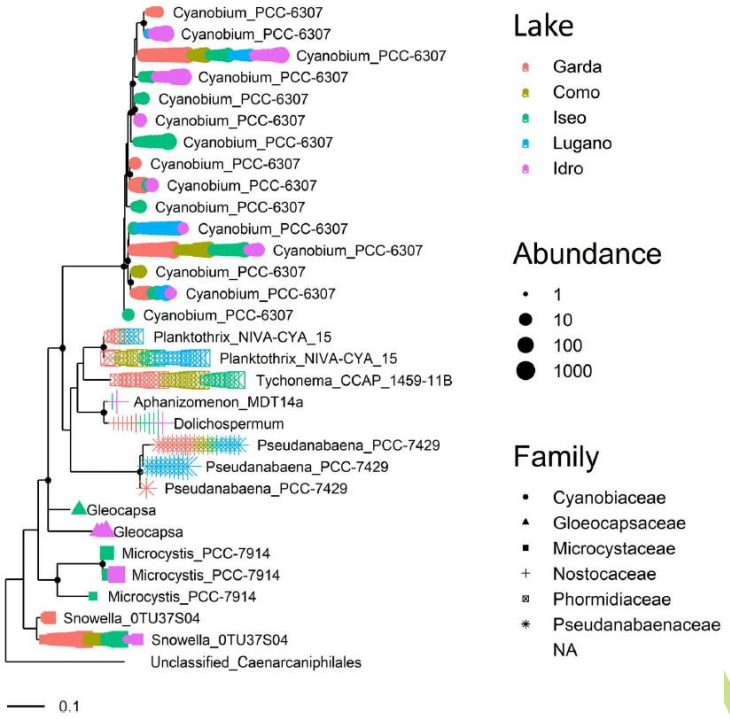
Interreg Alpine Space Eco-AlpsWater
European Regional Development Fund

4) BIOINFORMATIC ANALYSES AND DOWNSTREAM STATISTICAL ANALYSES

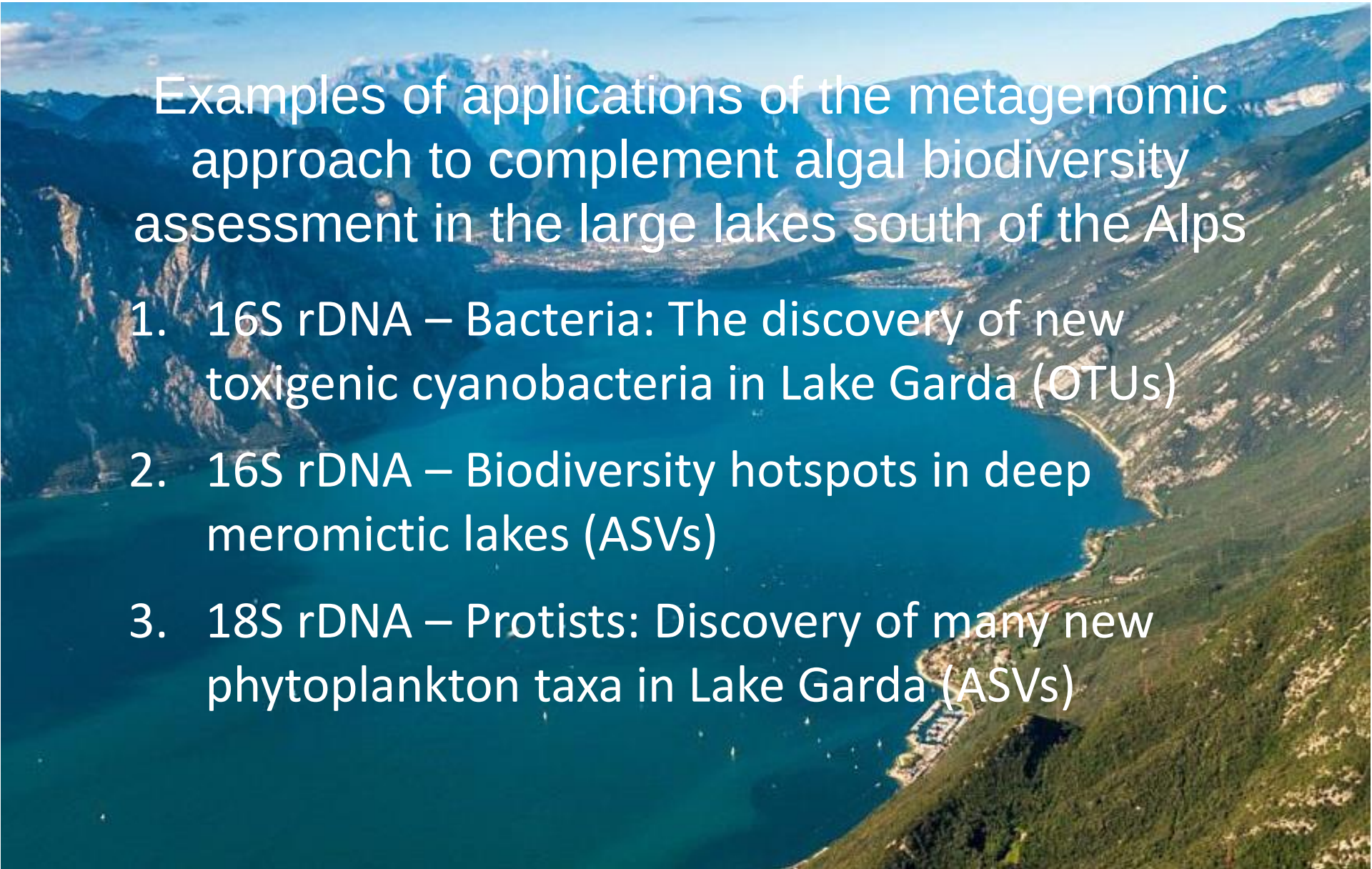
4.3. Downstream statistical analysis

- The time required for the application of standard bioinformatic pipelines (i.e. those that do not require development and adaptation to peculiare datasets of reads) is much less compared to the time required for downstream statistical analyses.
- Datasets can be very large and complex, and opened to a very wide variety of analyses
  - ✓ Biodiversity
  - ✓ Identification of nuisance/toxigenic/pathogenic species
  - ✓ Community ecology
  - ✓ Phylogenetic analyses
  - ✓ …





Interreg Alpine Space
Eco-AlpsWater
European Regional Development Fund
EUROPEAN UNION



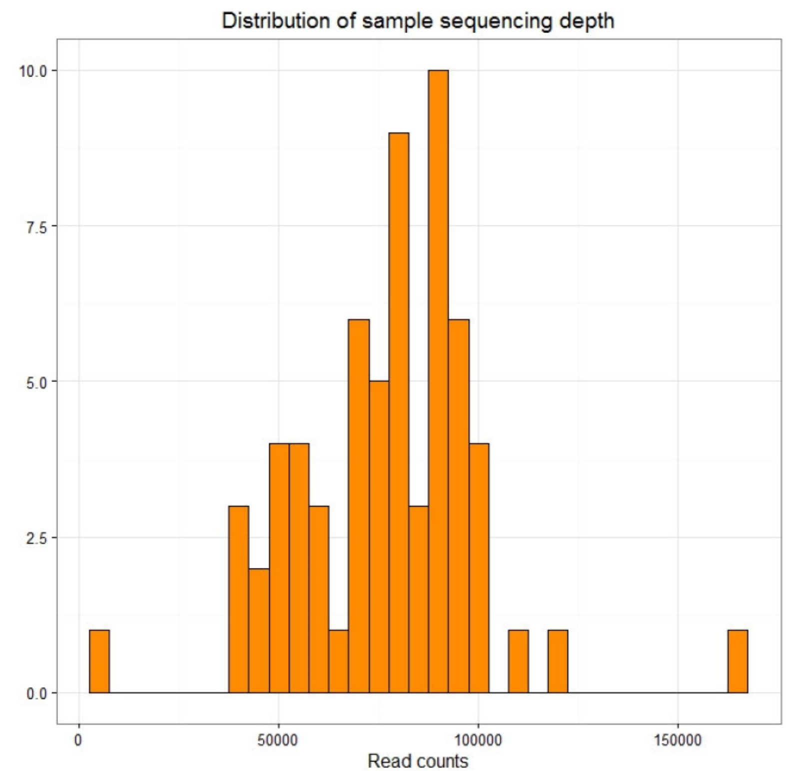
Examples of applications of the metagenomic approach to complement algal biodiversity assessment in the large lakes south of the Alps

1. 16S rDNA – Bacteria: The discovery of new toxigenic cyanobacteria in Lake Garda (OTUs)
2. 16S rDNA – Biodiversity hotspots in deep meromictic lakes (ASVs)
3. 18S rDNA – Protists: Discovery of many new phytoplankton taxa in Lake Garda (ASVs)

1) 16S rDNA – Bacteria: The discovery of new toxigenic cyanobacteria in Lake Garda

Barcoded library was sequenced on an Illumina® MiSeq (PE300) platform

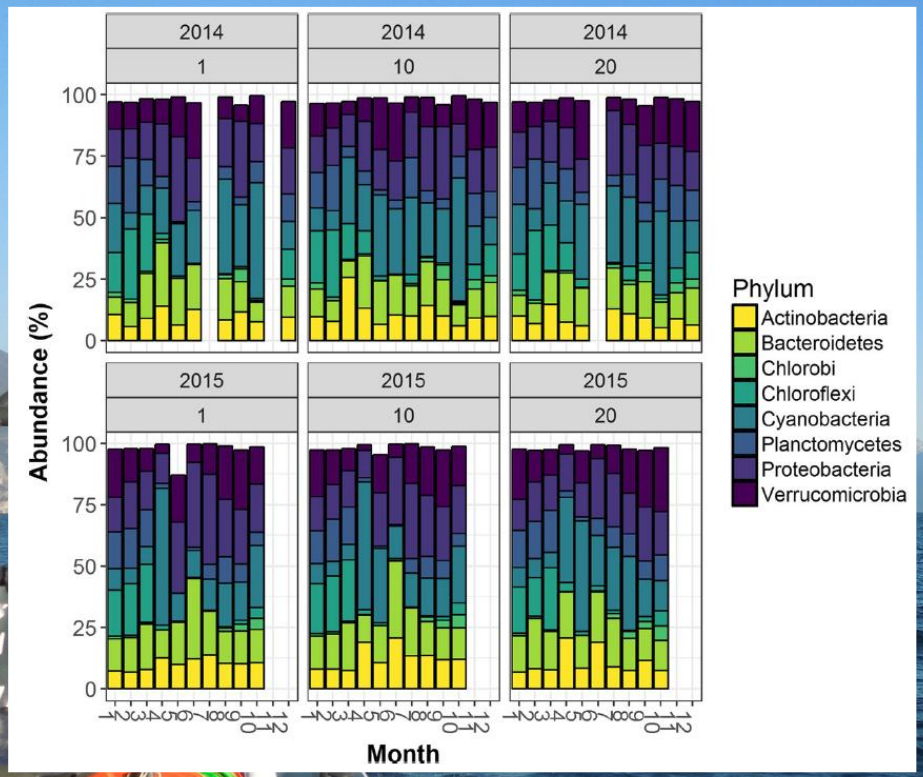
- 64 samples collected in 2014 and 2015
- Massive amplification of 16S rRNA from the samples produced over 7.000.000 reads

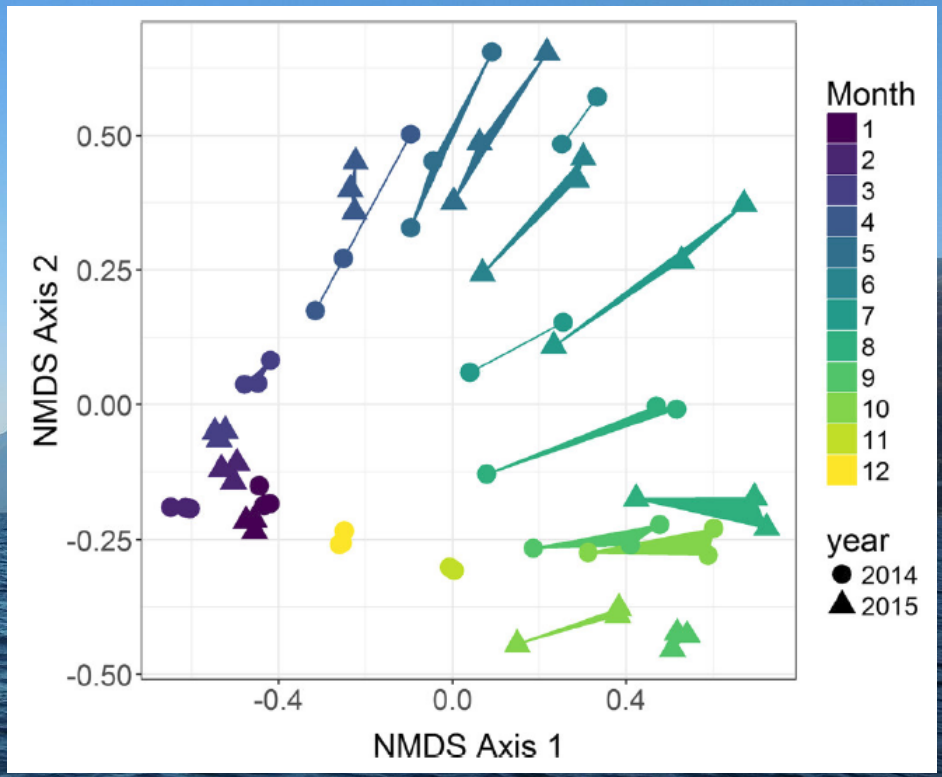



```
1 @M00380:72:000000000-AMNE0:1:1101:12817:1299 1:N:0:35
  CCTACGGGAGGCAGCAGTGGGGAATCTTGCGCAATGGGCGAAAGCCTGACGCAGCGACGCCGCGTGGGGGAAGAAGGCCTTAGGGTTGTAAACC
  +
  88BCCGGFDBA85=====CFFFGGGGGFGGGGGGGGGGGGGGGDGGGGG8F@EFGGGGGGGGGGGGEFGGGGGGGGDGGGGGGGGGGGGGGGG
2 @M00380:72:000000000-AMNE0:1:1101:14526:1314 1:N:0:35
  CCTACGGGCGGCTGCAGTAACGAATCTTCCGCAATGCGCGAAAGCGTGACGGGGTAATGCCGCGTGCAGGATGAAGCCCTTCGGGGTGTAAACT
  +
  8ACCCG6*88=5886;;C;ECC@CFGFGEEFDGGECFEGCFGGCEFEG7FGGGGFEFFECFEFGGGGDFGGGGGGFAEGGFGG@FEF@FFFFFD
3 @M00380:72:000000000-AMNE0:1:1101:15372:1352 1:N:0:35
  CCTACGGGTGGCAGCAGTAGGGAATATTGGTCAATGGGCGAAAGCCTGAACCAGCAACGCCGCGTGCGCGATGAAGGCCTTCGGGTCGTAAAGC
  +
  68ACCGGFB@)8:8====CFFGGGGGGGGGGGGEFGGGGDGGGDGGGGGGGGGGGDGGGGGGGGGDGGGGGGGGGGGGGGGFGGGGGGGGGGGG
4 @M00380:72:000000000-AMNE0:1:1101:19512:1372 1:N:0:35
  CCTACGGGTGGCTGCAGTGGGGAATATTGGGCAATGGAGGAAACTCTGAACCAGCGACGCCGCGTGCGGGACGAAGGCCTTCGGGTTGGAAACC
  +
  8BCCCGGFFFF5====CCF,CFGGGGGGGG,FFGGGGGFGEGGFGGGGG<FGGGG7F+FGGGGGGGGCGGG+CF7FGG8,BB<C7?F++,CF8F
5 @M00380:72:000000000-AMNE0:1:1101:21601:1401 1:N:0:35
  CCTACGGGAGGCAGCAGTGGGGAATTTTCCGCAATGGGCGCAAGCCTGACGGAGCAACGCCGCGTGAGGGATGAAGGCCTCTGGGCTGTAAACC
  +
  8ACCCGGEFFFA======CFGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGDFFGGDFGGGGGGGGCFGGGGGGGGGGG
6 @M00380:72:000000000-AMNE0:1:1101:20747:1412 1:N:0:35
  CCTACGGGGGGCAGCAGTGGGGAATATTGGACAATGGGCGCAAGCCTGATCCAGCCATGCCGCGTGAGTGATGAAGGCCTTAGGGTTGTAAAGC
  +
  ACCCCGGEAF=8:88::<CFFGGGGEGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGG>FGGFAFGGGFGGGGGGGGGGGGGGGGGGGGGGGGGGG
7 @M00380:72:000000000-AMNE0:1:1101:21591:1419 1:N:0:35
  CCTACGGGAGGCAGCAGTGGGGAATTTTCCGCAATGGGCGCAAGCCTGACGGAGCAACGCCGCGTGAGGGATGAAGGCCTCTGGGCTGTAAACCTCTTTTATCAAGGAAGAAGATCTGACGGTACCTGATGAATAAGCCACGGCTAATTCCGTGCGAGCAGCCGC
  +
  CCCCCGEB9@BA:=====;CEFFGG<FGCC@FGGGGGGGGDGGGGGGGGF@@CCFGGGGGGGEGGFGG8CFGGGGGGGGGGEGGG,4BFGCE,?4E?EFGE,5C<FGGFGGGDFG9?BF<?78B+,<,,,8<FF,A?CFCFB++@=,7@,:@,+3+8C>F<8DF@
```

Interreg Alpine Space Eco-AlpsWater
European Regional Development Fund

1) 16S rDNA – Bacteria: The discovery of new toxigenic cyanobacteria in Lake Garda





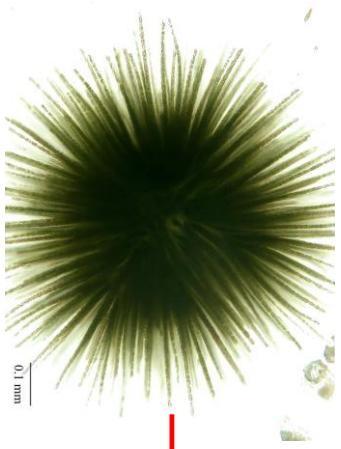
Salmaso et al., 2018, Microbial Ecology

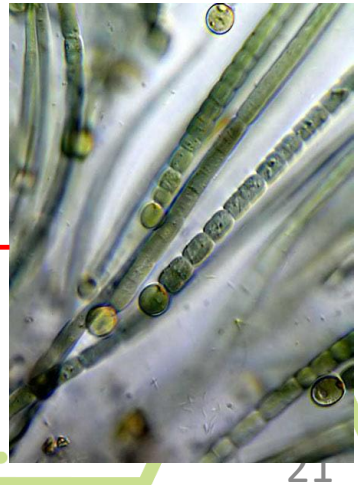
Interreg Alpine Space Eco-AlpsWater
European Regional Development Fund

1) 16S rDNA – Bacteria: The discovery of new toxigenic cyanobacteria in Lake Garda

Cyanobacterial taxa identified with traditional approaches and HTS

Taxon	Detected by microscopy-m (and phylogenetic-p)	Detected by HTS and downstream analysis (>99%)	Toxins
Tychonema bourrellyi	X mp	X	Anatoxins
Dolichospermum lemmermannii	X mp	X	-
Planktothrix rubescens	X mp	X	Microcystins
Limnothrix redekei	X m	X	Saxitoxins
Snowella sp.	X m	X	?
Microcystis aeruginosa	X m	X(?) 5 candidates	Microcystins
Limnococcus limneticus (*Chroococcus*)	(X) m	X	?
Aphanothece sp.	X m	X	?
Synechococcus rubescens		X	Microcystins
Radiocystis sp.		X	Microcystins
Chroococcus minutus		X	?
Gloeotrichia echinulata		X	?
Nostoc calcicola		X	Microcystins





More than 30 cyanobacterial OTUs did not show any clear similarity with the curated 16S rRNA archives

2) 16S rDNA – Biodiversity hotspots in deep meromictic lakes: Lake Idro, NE-Italy

DESEQ2 test, $p < 0.05$

Epilimnion

Microcystaceae	*Microcystis*	Toxic cyanobacteria
Aphanizomenonaceae	*Aphanizomenon*	
Leptospiraceae	*Leptospira*	Contains some pathogenic species
Roseiflexaceae	*Roseiflexus*	Grows photoheterotrophically
Microbacteriaceae	«*Candidatus* Planktoluna»	Freshwater habitats
Xanthomonadaceae	*Arenimonas*	

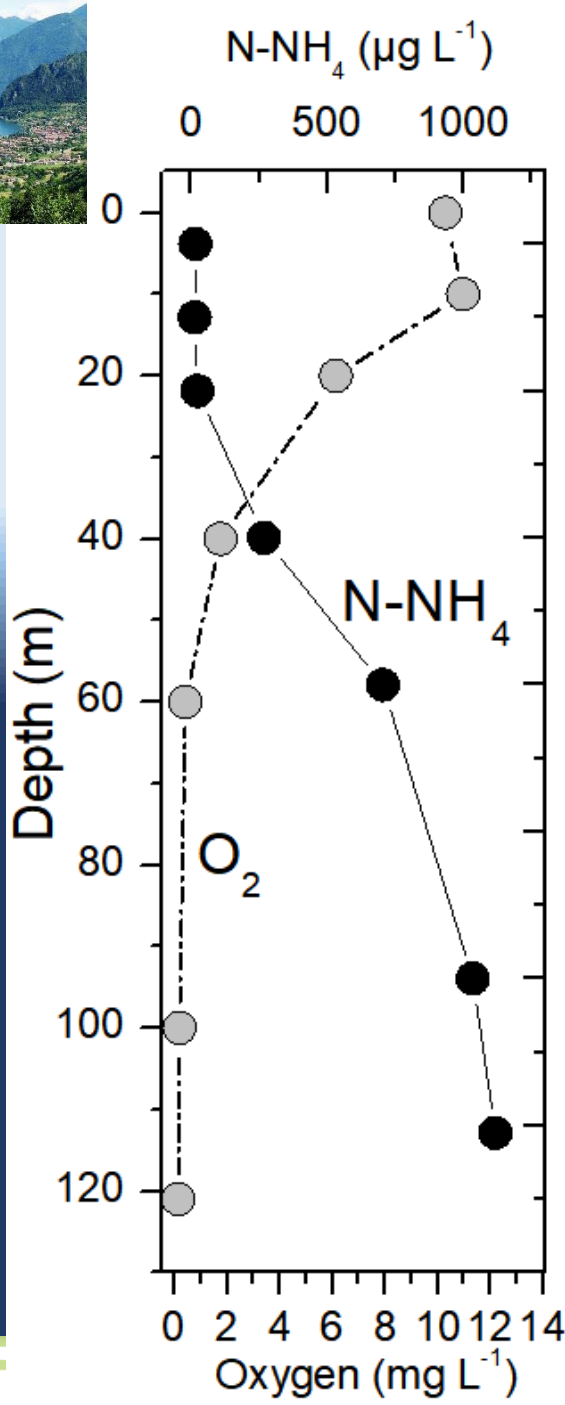
Chemolimnion

In the oxic hypolimnion, other typical families, e.g.

- methanotrophs, oxidizes methane
- Heterotrophic facultative anaerobes
- (…)

Monimolimnion

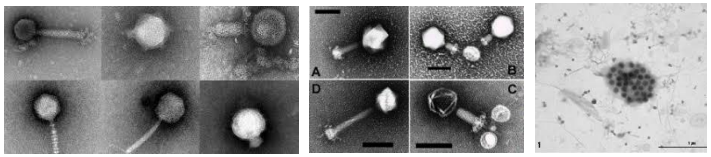
Desulfobacteraceae	*Desulfatirhabdium*	
	Desulfocapsa	Sulfate-reducing bacteria
	Desulfurivibrio	
Helicobacteraceae	*Sulfuricurvum*	Sulfide oxidation at the expense of NO2
	Sulfurimonas	
Planctomycetaceae	«*Ca*. Anammoximicrobium»	Anammox
Porphyromonadaceae	*Paludibacter*	Chemo-organotrophic
Rhodocyclaceae	*Sulfuritalea*	Denitrification (?)
Syntrophaceae	*Desulfomonile*	Sulfate-reducing bacteria
	Smithella	
	Syntrophus	Syntrophic; H2-utilizing bacteria
Xanthobacteraceae	*Pseudolabrys*	Use several organic acids



Salmaso, 2019, https://www.frontiersin.org/articles/10.3389/fmicb.2019.02257/full

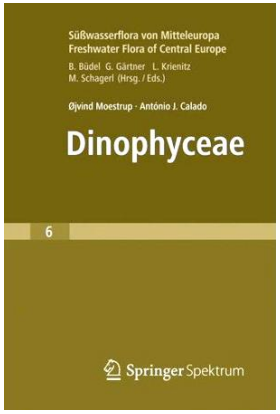
Interreg Alpine Space Eco-AlpsWater
European Regional Development Fund

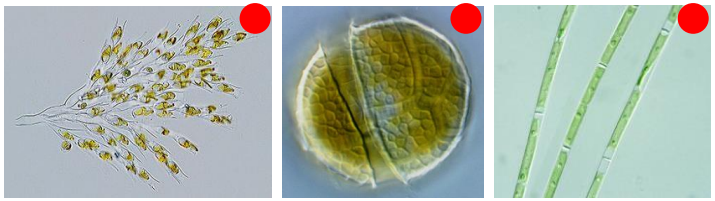
3) 18S rDNA – Protists: Discovery of many new phytoplankton taxa in Lake Garda (2014-2015)



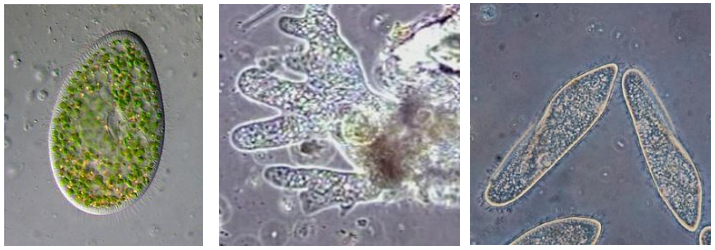
Viruses

Archaea
Bacteria & **Cyanobacteria**

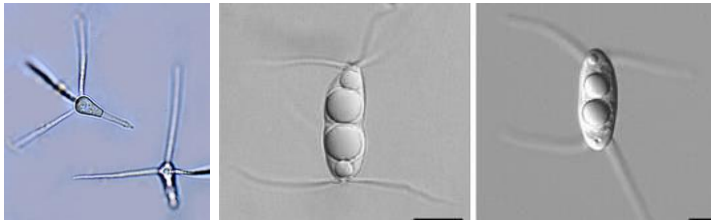




«**Phytoplankton**»



Heterotrophic protists



Fungi

Interreg Alpine Space Eco-AlpsWater European Regional Development Fund

3) 18S rDNA – Protists: Discovery of many new phytoplankton taxa in Lake Garda (2014-2015)

After classification, rarefaction and inclusion of ASVs in functional groups:

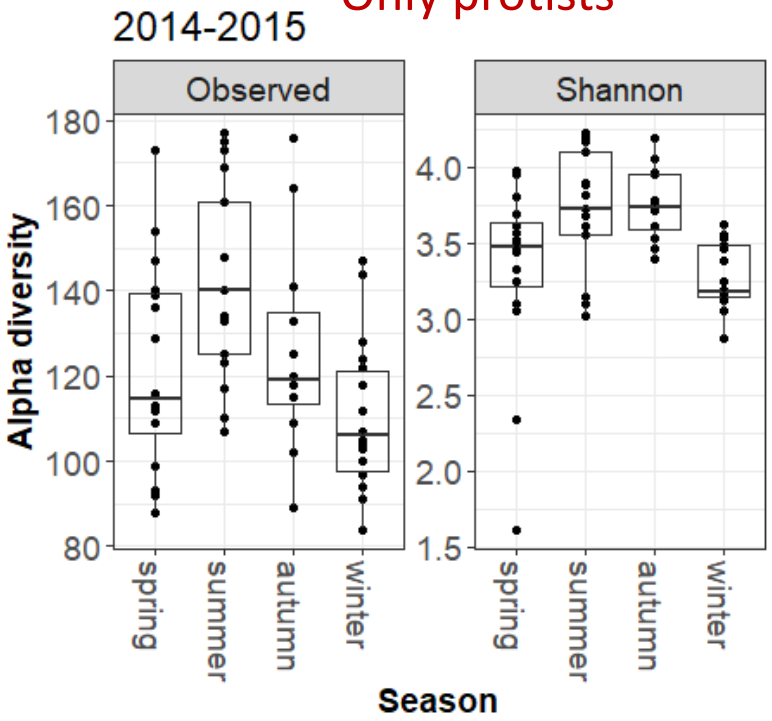
ca. 500 different heterotrophic protists ASVs

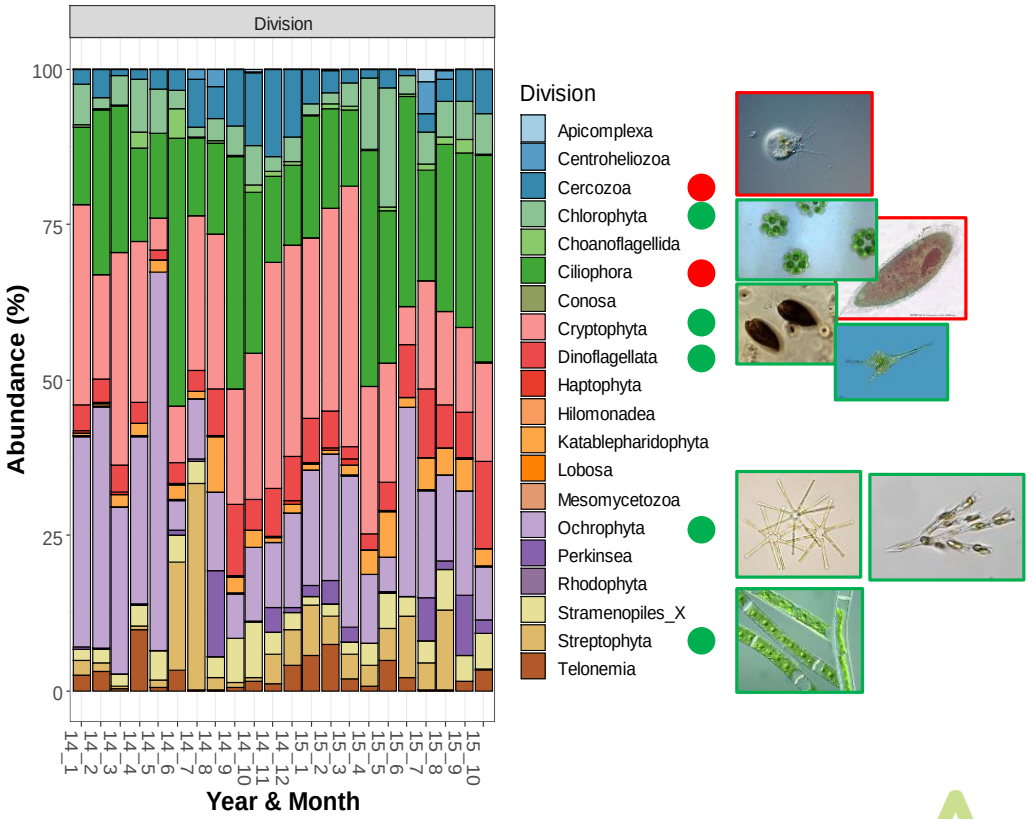
ca. 400 phytoplankton ASVs

ca. 60 zooplankton ASVs

ca. 120 fungi ASVs

Only protists





Unpublished data – In preparation

3) 18S rDNA – Protists: Discovery of many new phytoplankton taxa in Lake Garda (2014-2015)

For most phytoplankton groups, HTS allowed to:

- Confirm variation in rDNA sequences within single taxa, due to SNPs (different ASVs)
- Confirm previous microscopic based classification of genera and/or species
- Highlight the existence of species so far not described
- Suggesting existence of new taxa not identified by microscopy

PRELIMINARY RESULTS OF TWO SELECTED GROUPS

Dinoflagellates taxa

Tovellia_aveirensis
Ceratium_furcoides
Ceratium_hirundinella
Gymnodinium_sp.
Gyrodinium_helveticum
Peridinium_cinctum
Peridinium_willei
Parvodinium_inconspicuum
Scrippsiella_acuminata
Thoracosphaeraceae_sp.
Prorocentrum_sp.
Baldinia_sp.
Asulcocephalium_miricentonis

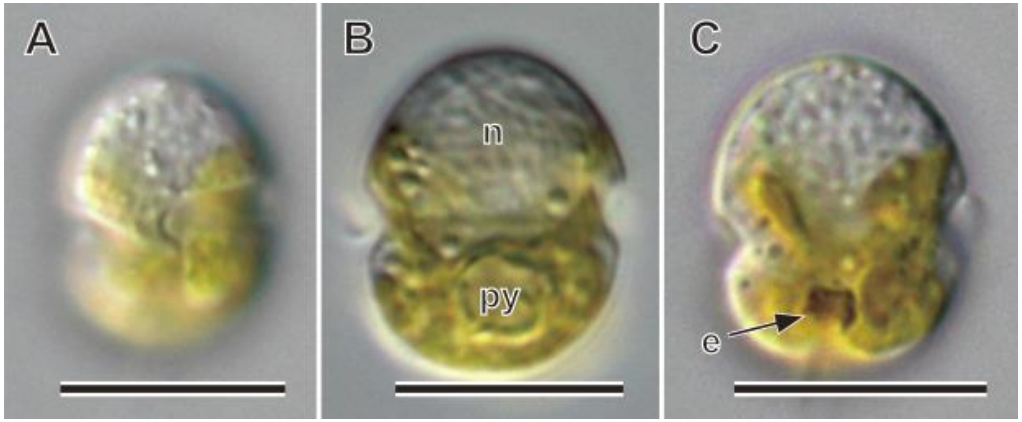
Cryptophytes taxa

Basal_Cryptophyceae-1_sp.
Cryptomonadales_sp.
Cryptomonas_curvata
Cryptomonas_tetrapyrenoidosa

Compared to previous analyses, HTS did not identify *Plagioselmis nannoplanctica*

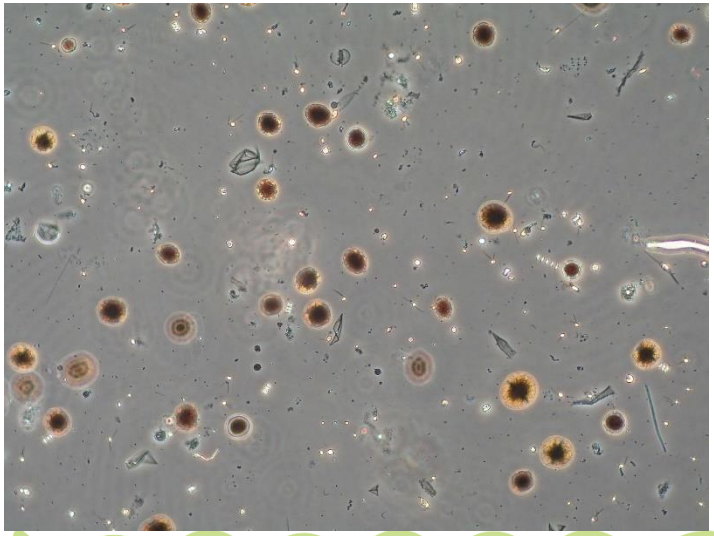
Interreg Alpine Space Eco-AlpsWater
European Regional Development Fund

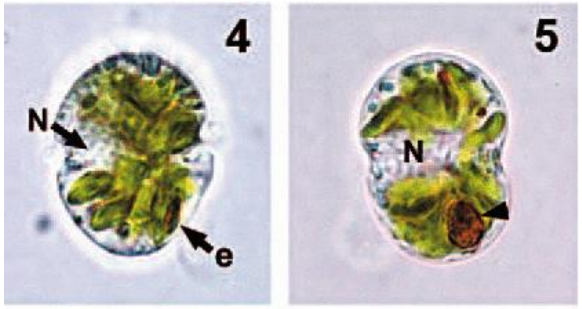
3) 18S rDNA – Protists: Discovery of many new phytoplankton taxa in Lake Garda (2014-2015)

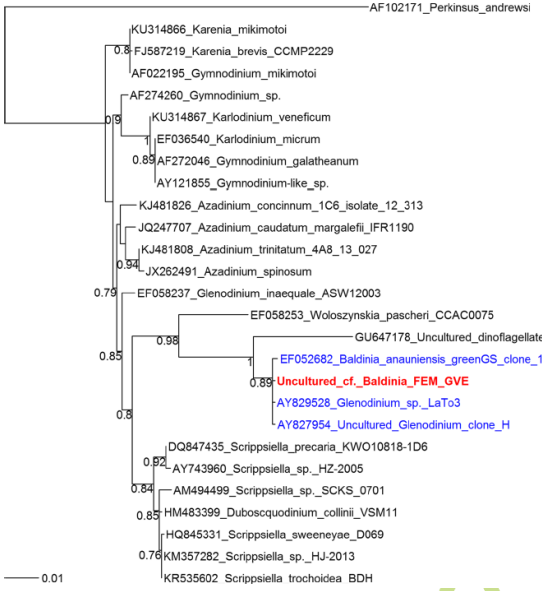


Asulcocephalium miricentonis
Described in 2015 in Japanese freshwater ponds

Takahashi et al., 2015, Protist







Baldinia anauniensis
Reported for the first time in 2015, after a huge bloom

Salmaso et al., 2018, Hydrobiologia

Interreg Alpine Space Eco-AlpsWater
European Regional Development Fund

General strategy

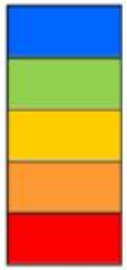
Improve the traditional monitoring approaches (WFD) by using advanced HTS techniques

- Bacteria/cyanobacteria
- Phytoplankton
- Phytobenthos (diatoms)
- Fish

Traditional approach

WPT2

Ecological status (WFD)



WPT3

Species lists

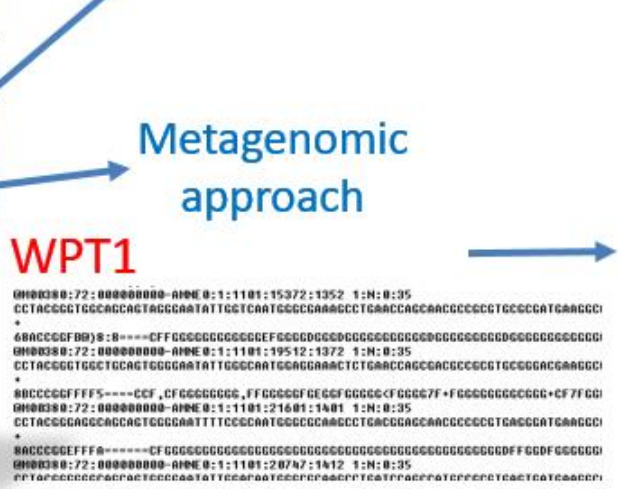
Biodiversity

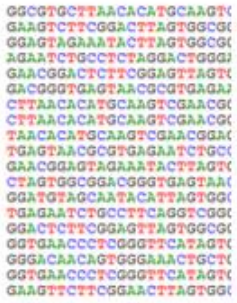
- Management plans
- Policy

WPT4

Metagenomic approach

WPT1





Discriminant features between 1st generation and next generation sequencing approaches

Process steps /Attributes	1st gen seq «genetic»	NGS «metabarcoding»
DNA extraction	Cultured isolates or strains/specimens	Environmental DNA
Target	Strain	Community (prok.-euk.)
Lenght of reads	Typically from 100 - >1500	Short reads (~400 bp)
Diversity of reads	1 type F-R (if uncontaminated)	Highly diverse (over thousands)
Taxonomic sensitivity	From Low to High	Low
Able to identify the unknown taxa	Standard application: No	Yes
Can be used for monitoring purposes	To complement	At present, No: To complement
Biological metrics available	NA	Rare examples – In progress
Quantitative	No (but alternative, qPCR)	Semiquantitative
Contribution to monitoring	Taxonomy and functional attributes of key taxa	Biodiversity, taxonomic species list check, introduced species

Considerazioni conclusive

- Le tecniche NGS/HTS stanno radicalmente cambiando il nostro modo di studiare e valutare la biodiversità degli ecosistemi acquatici.
- Vi sono tuttavia ancora molti aspetti critici da risolvere, legati anche all'attuale tecnologia utilizzata (waiting the next-next… approaches).
- Le ricadute sui sistemi di monitoraggio diventeranno fondamentali nei prossimi anni (in fase di studio e applicazione: Eco-AlpsWater, DNAqua-net). Focus su batteri/fitoplancton, macrobentos, macrofite e pesci.
- Strumenti e protocolli per l'analisi avanzata di batteri/fitoplancton e pesci sono in via di definizione nell'ambito del Consorzio EAW