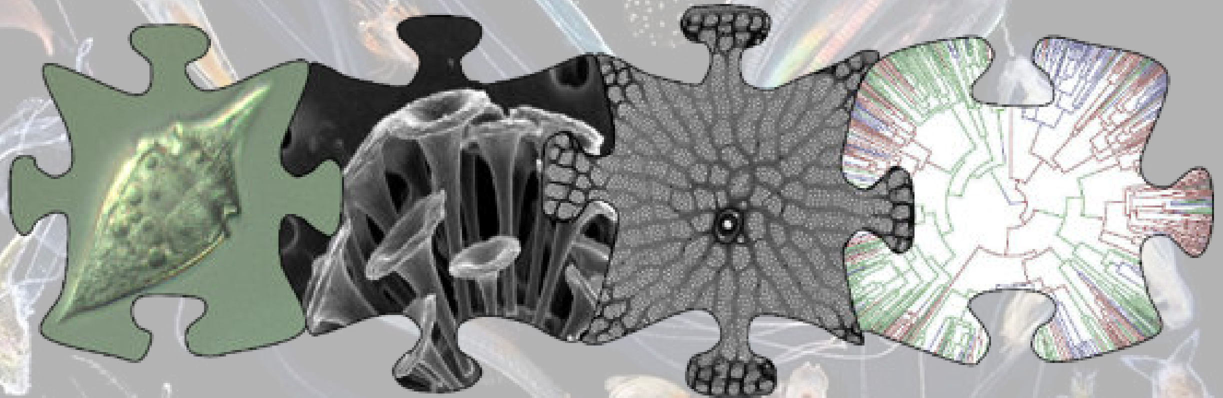


APC 13 - Naples

Metabarcoding for (Phyto)plankton

2024-10-22

Daniel Vaulot (vaulot@gmail.com)

CNRS • SORBONNE UNIVERSITÉ
Station Biologique

UiO : University of Oslo

Station Biologique de Roscoff

Station Biologique de Roscoff

- 1872 - Antoine Lacaze-Duthiers (150 years ago)
- CNRS and Sorbonne Université
- Staff: 350
- Students : 1,000-2,000 per year

Ecologie of Marine Plankton team (ECOMAP)

- Scientists: 10
- Staff: ~ 40
- Research themes:
    - Viruses
    - Bacteria
    - Cyanobacteria
    - Symbioses
    - Parasitism

Station Biologique de Roscoff

ECOLOGY OF MARINE PLANKTON - ECOMAP

PERMANENT STAFF

Christian JEANTHON
DR2 CNRS
Ecology of marine bacteria and their interactions with phytoplankton

Laure GUILLOU
DR1 CNRS
Associations within marine plankton: symbiosis and parasitism

Anne-Claire BAUDOUX
CRCN CNRS
Marine viruses: ecology, diversity and evolution

Aurélie CHAMBOUVET
CRCN CNRS
Diversity and Ecology of metazoan parasitic protists

Colomban DE VARGAS
DR1 CNRS & DU FR Tara Oceans GOSEE
Plankton system biology, biodiversity, symbiosis & evolution

Laurence GARCZAREK
DR1 CNRS
Genetic and functional diversity of marine cyanobacteria

Fabrice NOT
DR2 CNRS & DU UMR7144
Plankton symbiosis: diversity, ecology & evolution

Frédéric PARTENSKY
DR1 CNRS
Function and evolution of cyanobacterial light-harvesting complexes

Nathalie SIMON
MdC SU
Phytoplankton biodiversity and temporal dynamics

Christophe SIX
MdC SU
Adaptation of phytoplankton to temperature

Daniel VAULOT
DRCE CNRS
Emeritus
Diversity and ecology of eukaryotic picoplankton

Céline MANCEAU
AI CNRS
Aide aux projets
RH de proximité
Communication
33%ECOMAP

Dominique MARIE
IRHC CNRS
Flow cytometry
Security and safety

Charlotte BERTHELIER
IE CNRS
Bioanalyses of Marine Plankton
40%ECOMAP

OBSERVATION

Nicolas HENRY
CDI IR FR2424
Plankton eco-morpho-genetic data analysis
ECOMAP 80%
FR2424 20%

Fabienne RIGAUT-JALABERT
IR CNRS FR2424
Observation team manager
PHYTOBS / SOMLIT

Estelle BIGEARD
IE CNRS
Research on marine parasites (eukaryotes & viruses)
Lab Management (LabCollector)

Florence LE GALL
IEHC CNRS
Laboratory management and Research
Permanent Training

ROSCOFF CULTURE COLLECTION

Ian PROBERT
IR CNRS
FR2424
RCC Manager
Diversity of haptophytes

Priscillia GOURVIL
IE CNRS
FR2424
Curator of the RCC

Morgane RATIN
IR CNRS
Molecular biology and genetics of cyanobacteria
Security and safety

Sarah ROMAC
IE CNRS
Molecular biology & Lab management

Martin GACHENOT
CDD SU
FR2424
RCC
Strain maintenance & Flow Cytometry Platform

Léna GOUHIER
CDD SU
FR2424
RCC
Cryopreservation

POST-DOCS, PhD STUDENTS & TEMPORARY STAFF

Benjamin ALRIC
Post-Doc
Structure of phytoplanktonic communities & ecosystem functioning

Cédric BERNEY
Post-Doc
UniEuk project : Linking protist taxonomy and genetic databases

Gregory FARRANT
Post-doc
Phytoplankton multi-gene marker database

Charles BACHY
Post-doc
Characterization of phytoplankton by FISH

Marine LANDA
Post-doc
Interactions between diatoms and bacteria

Jade LECONTE
Post-Doc
Genomics and Metagenomics of marine picocyano-bacteria

Sebastian METZ
Post-Doc
Associations between host-parasites and the symbiont community

Marie WALDE
Post-Doc
Imaging of symbioses in marine plankton

Edwin DACHE
PhD student
Study of meiofauna communities by imaging methods
50%ECOMAP
50% Ifremer

Louison DUFOUR
PhD student
Chromatic acclimation of marine *Synechococcus* cyanobacteria

Mathilde FERRIEUX
PhD student
Role of iron in the genetic diversity of marine picocyanobacteria

Sarah GARRIC
PhD student
Adaptive physiology of cryptophyte microalgae

Iris RIZOS
PhD student
Life cycle of photosymbiotic

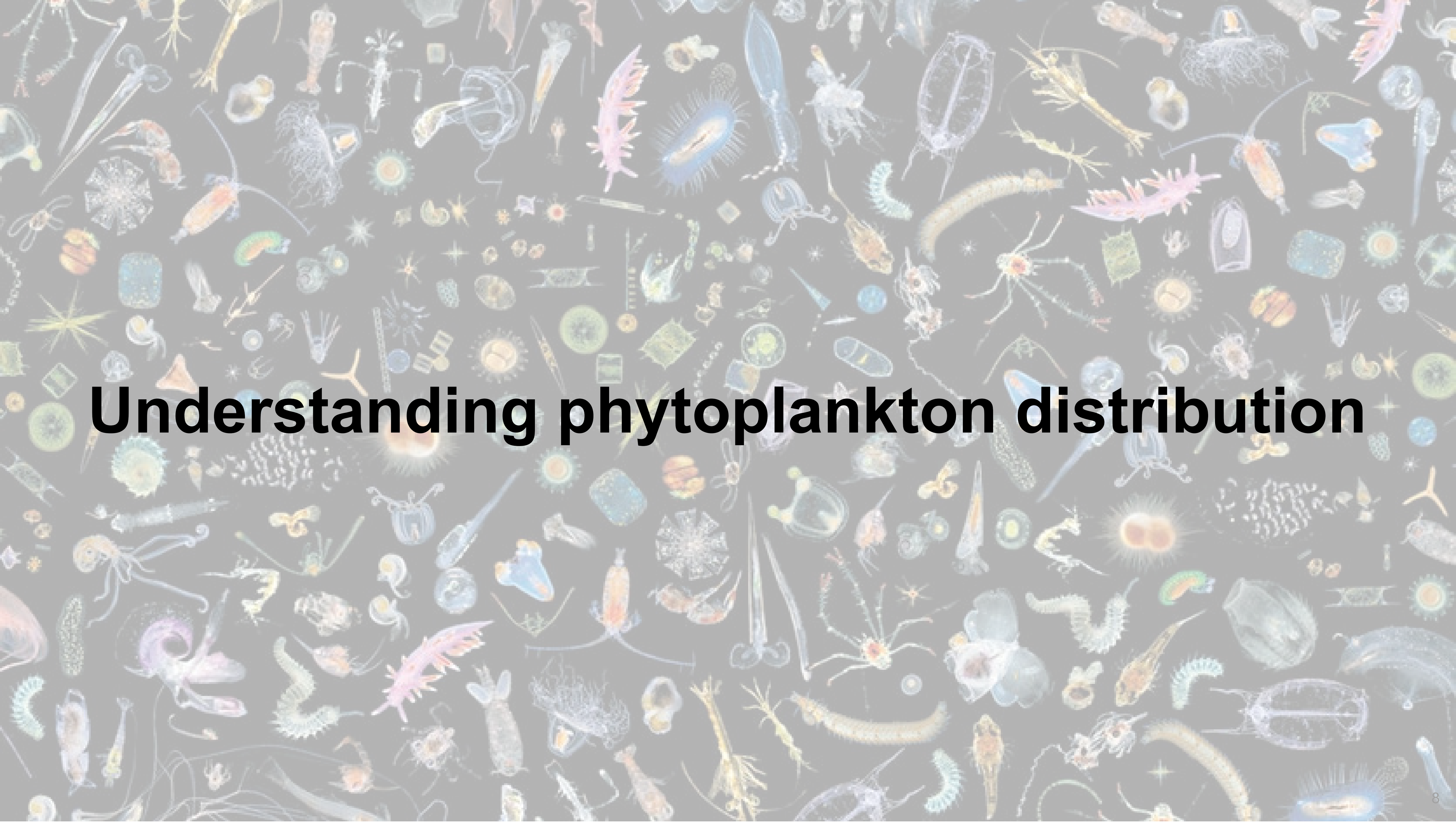
Jeremy SZYMCZAK
PhD student
Chemical interactions

Rafaele ATTIA
CDD IR
Development of microfluidic tools

Valentin FOULON
CDD IR
Imaging of meiofauna in marine sediment

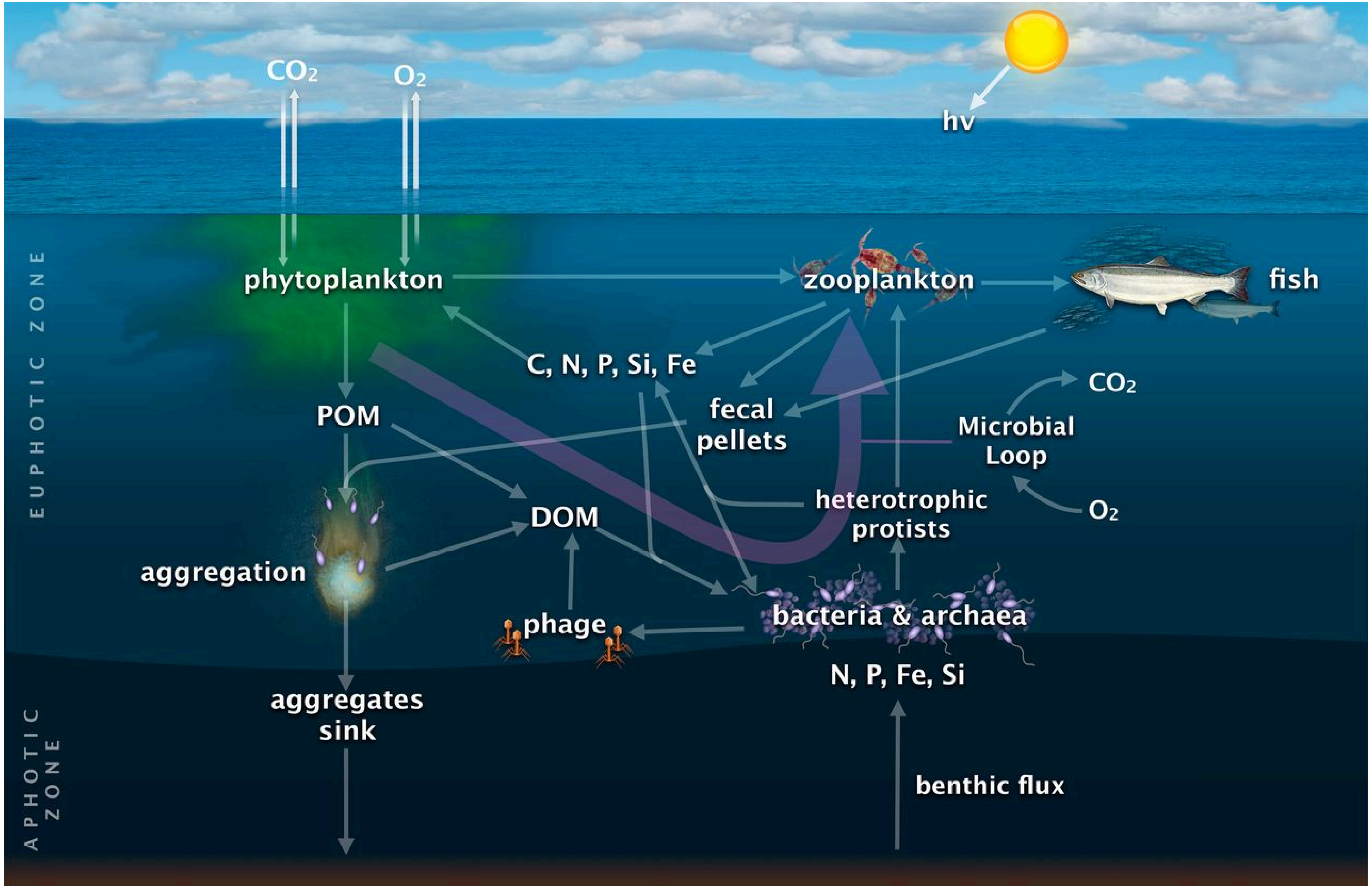
Outline

- Marine phytoplankton distribution and controls
- Metabarcoding for plankton
- Examples of application
  - Green phytoplankton
  - Biogeography of polar scpecies
- metaPR2: a database of metabarcodes

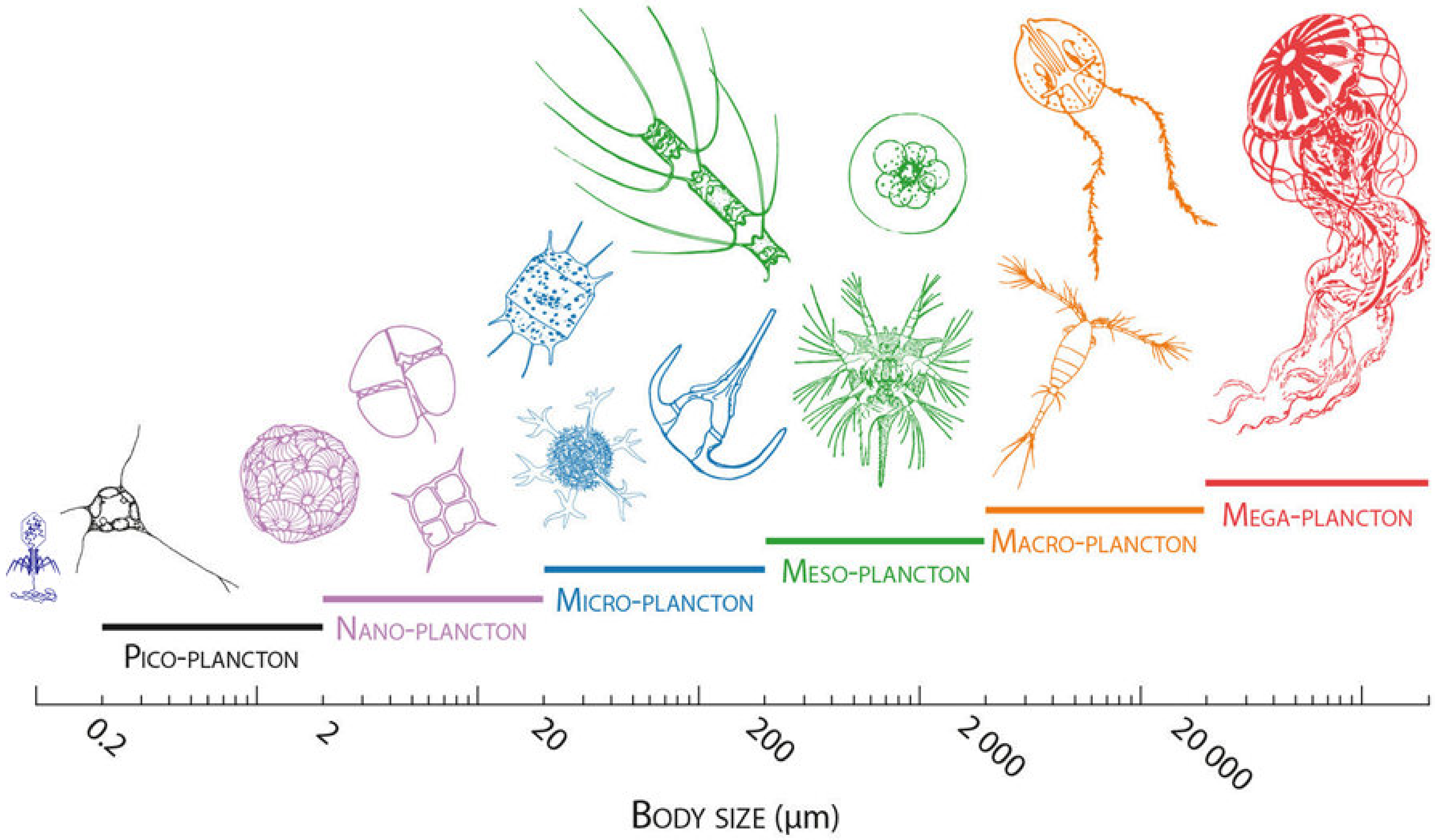


Understanding phytoplankton distribution

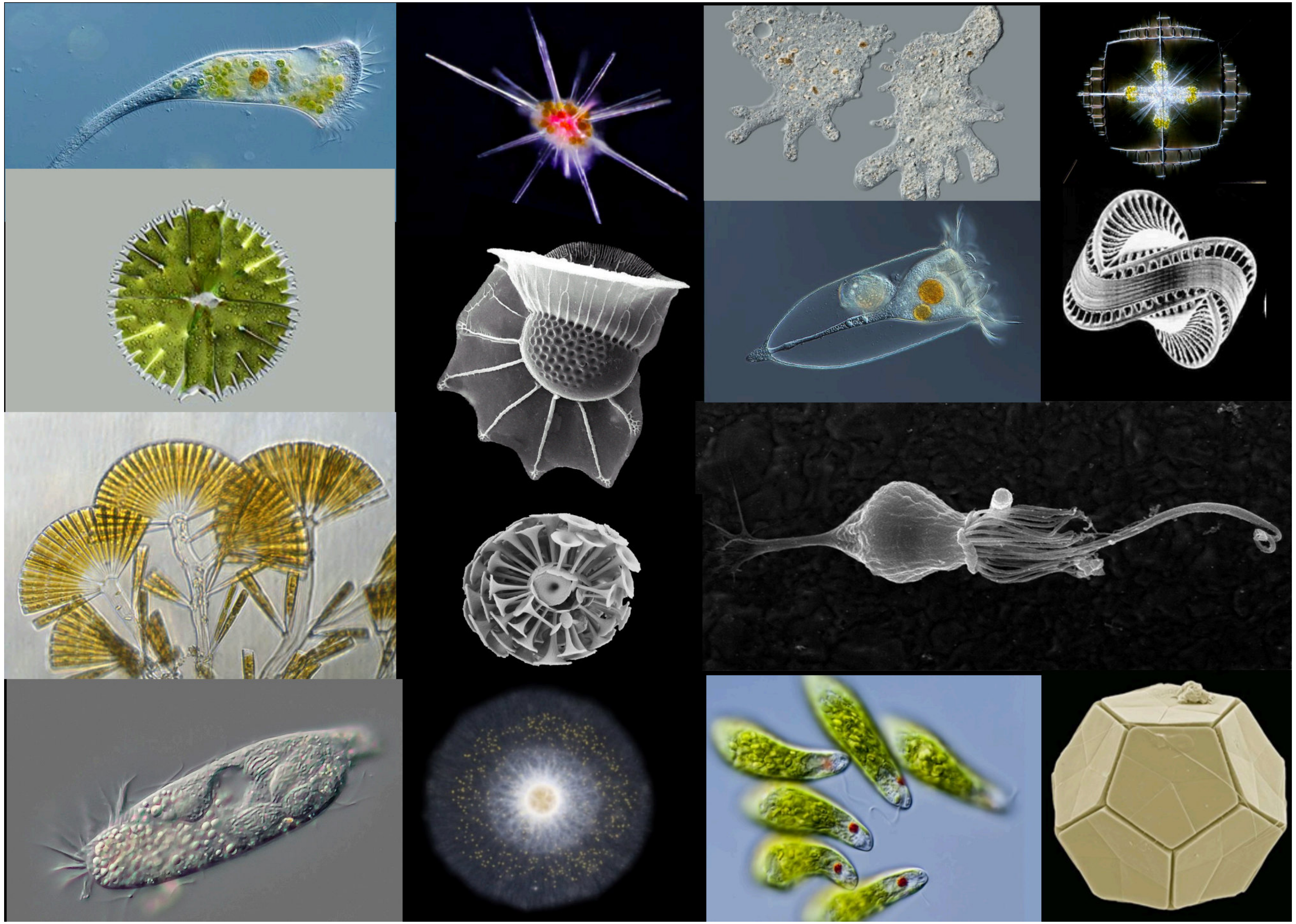
Marine food webs



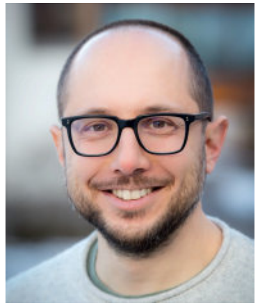
Size classes

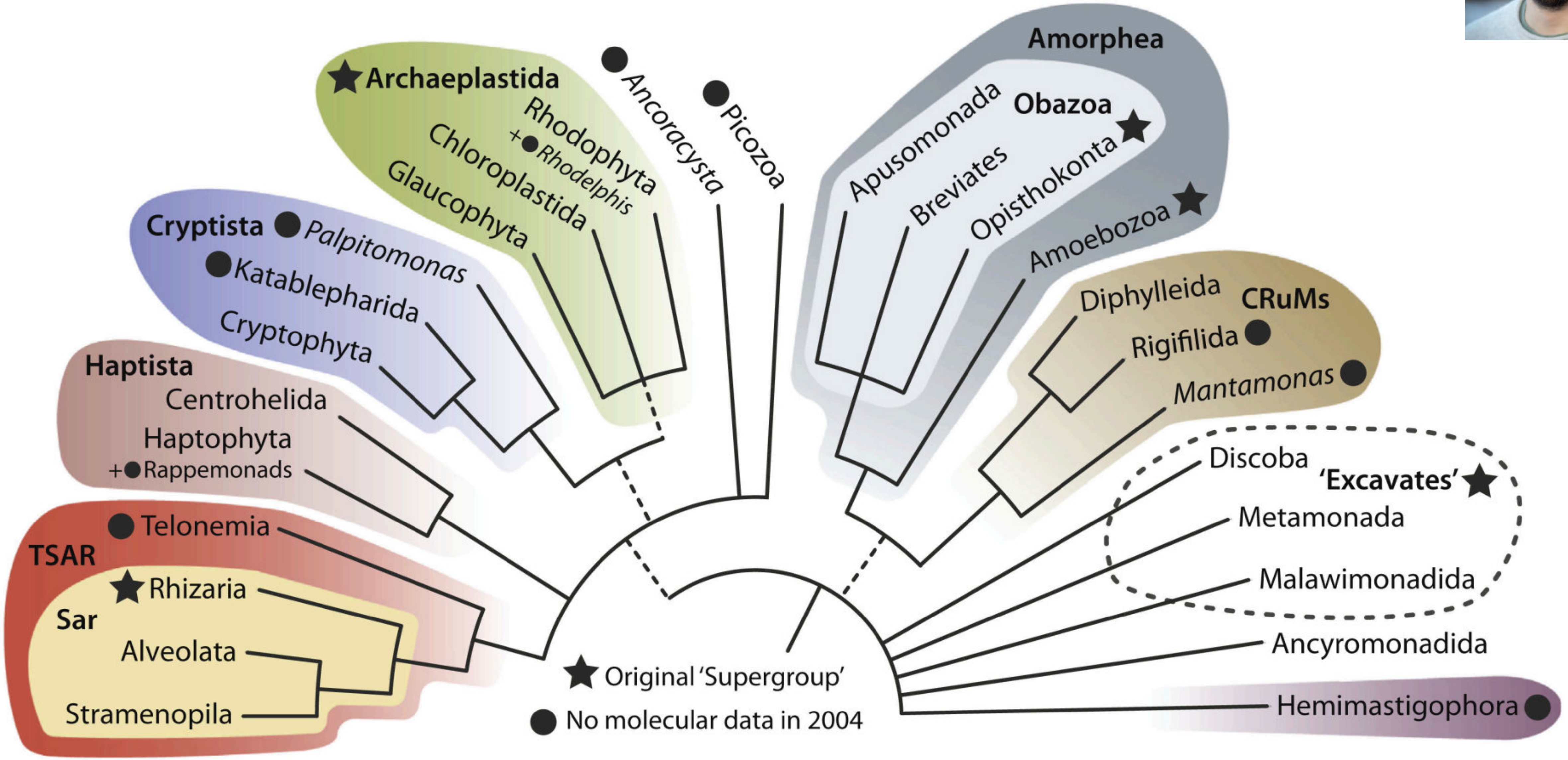


Form and function



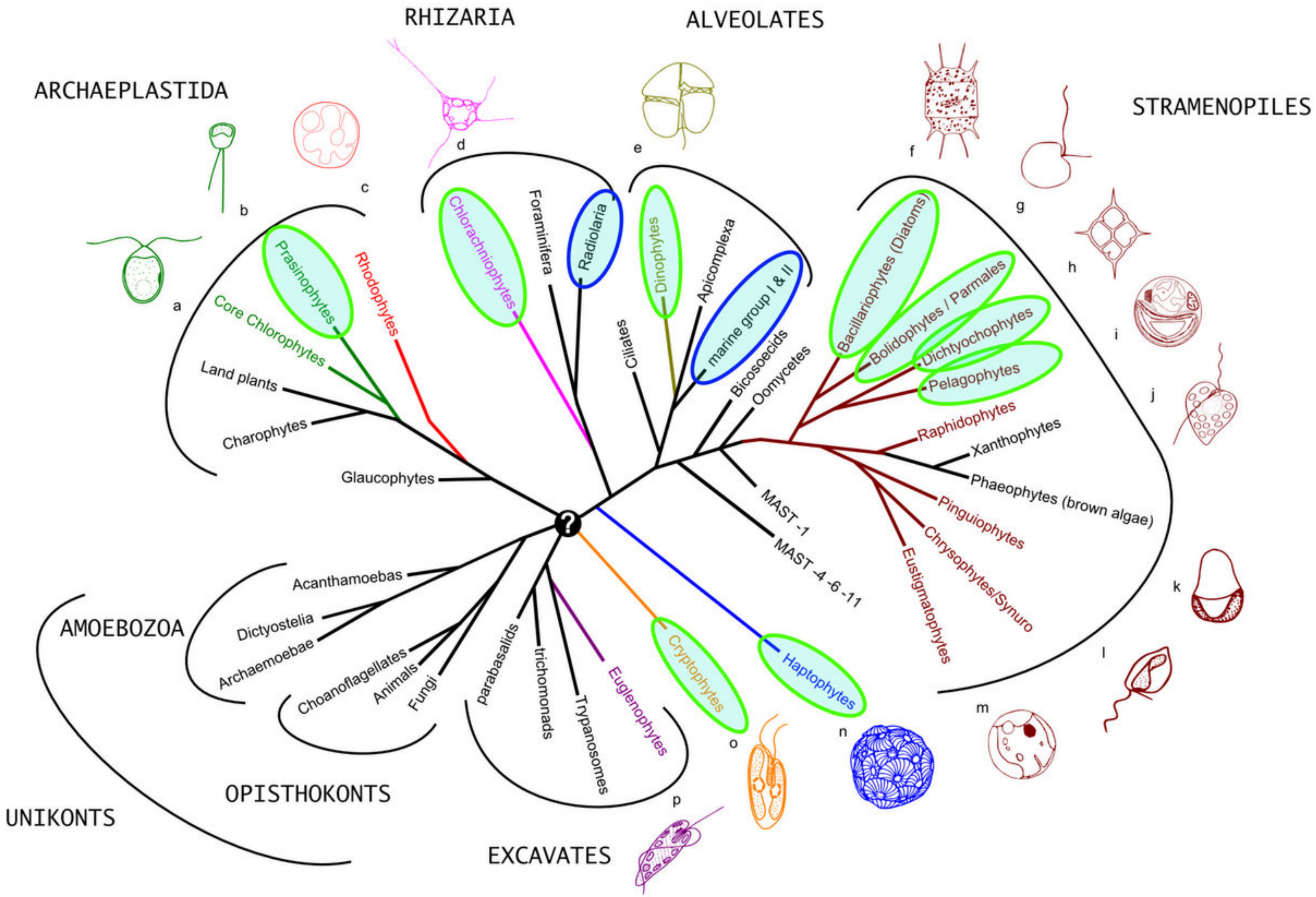
Molecular Phylogeny





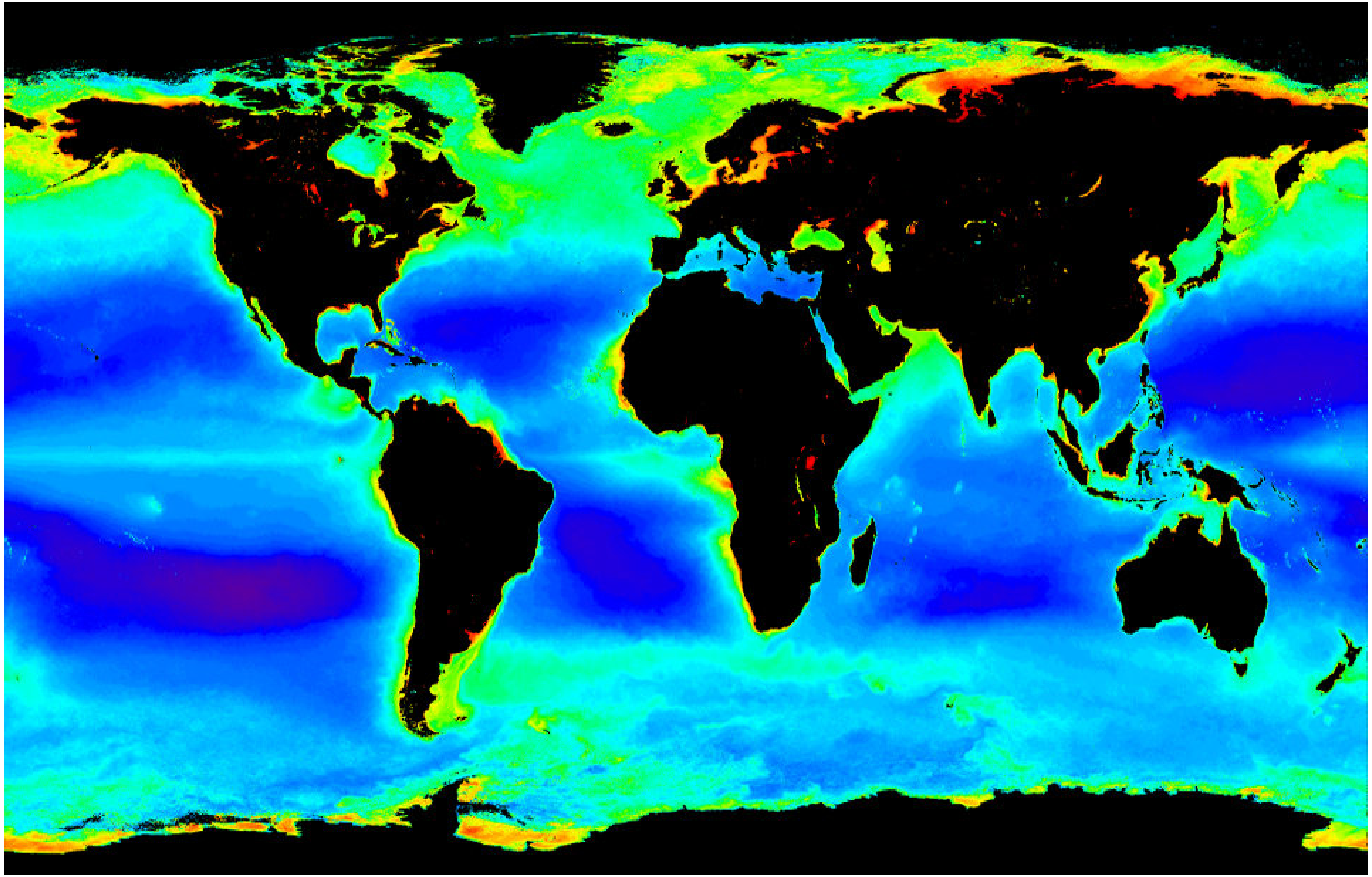
Trends in Ecology & Evolution

Phytoplankton is not monophyletic

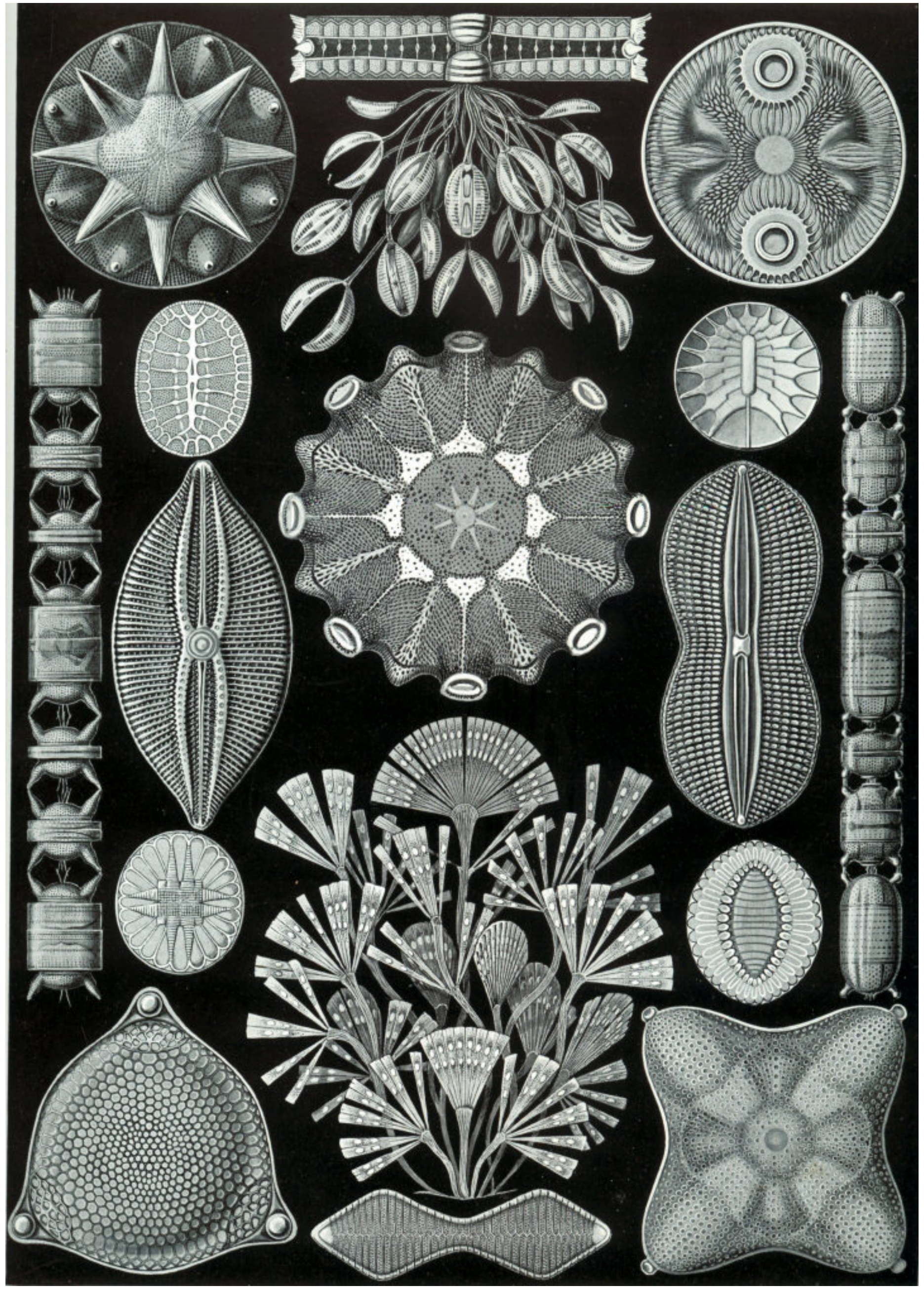


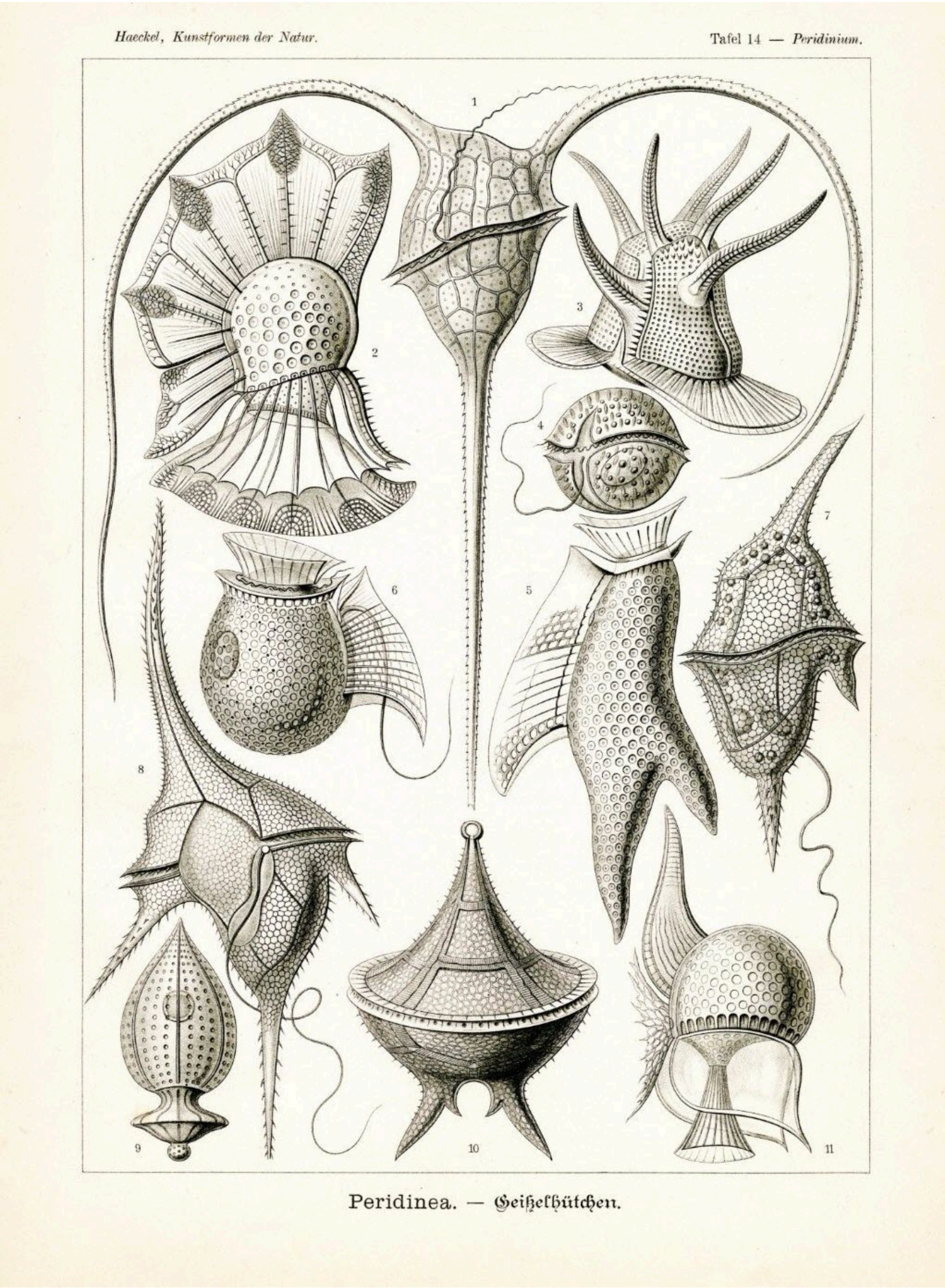
Phytoplankton rich regions

- Chl *a* estimated from satellite

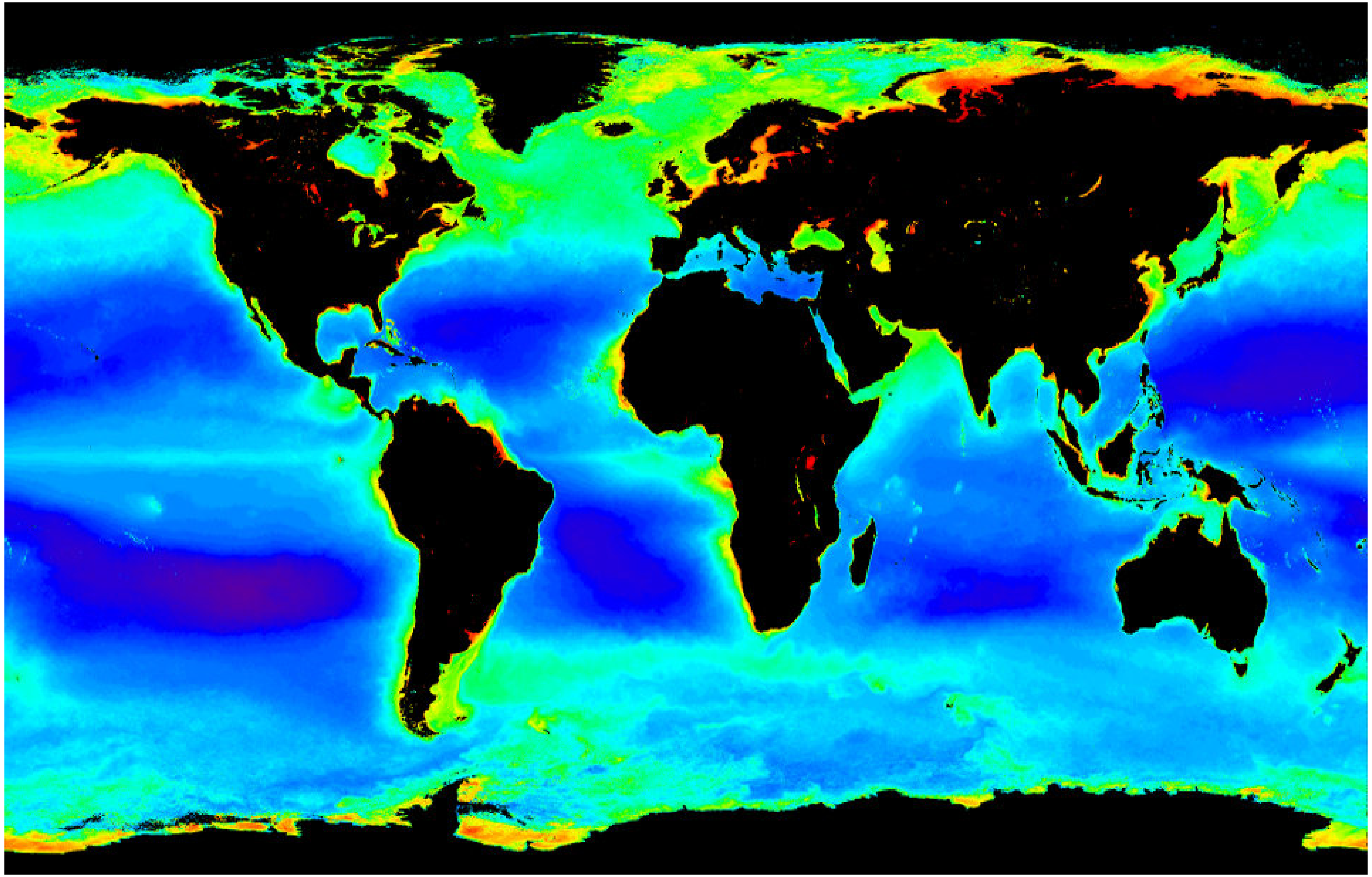


Diatoms and dinoflagellates: 20-200 µm



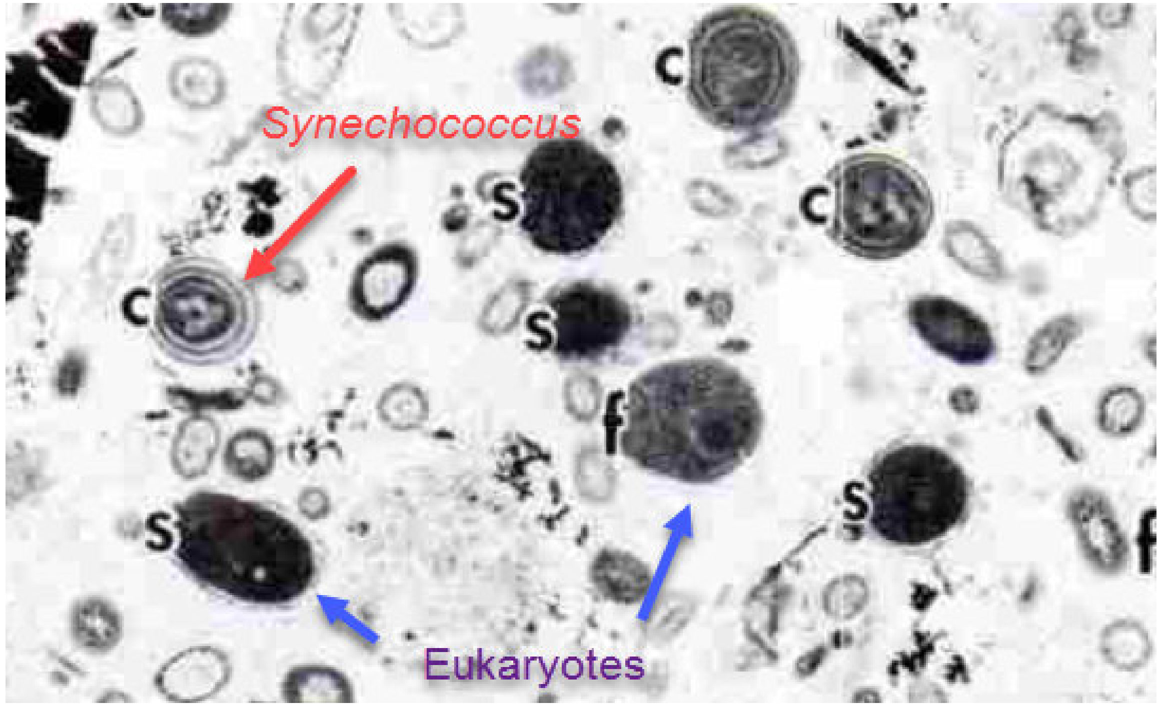


Oceanic deserts



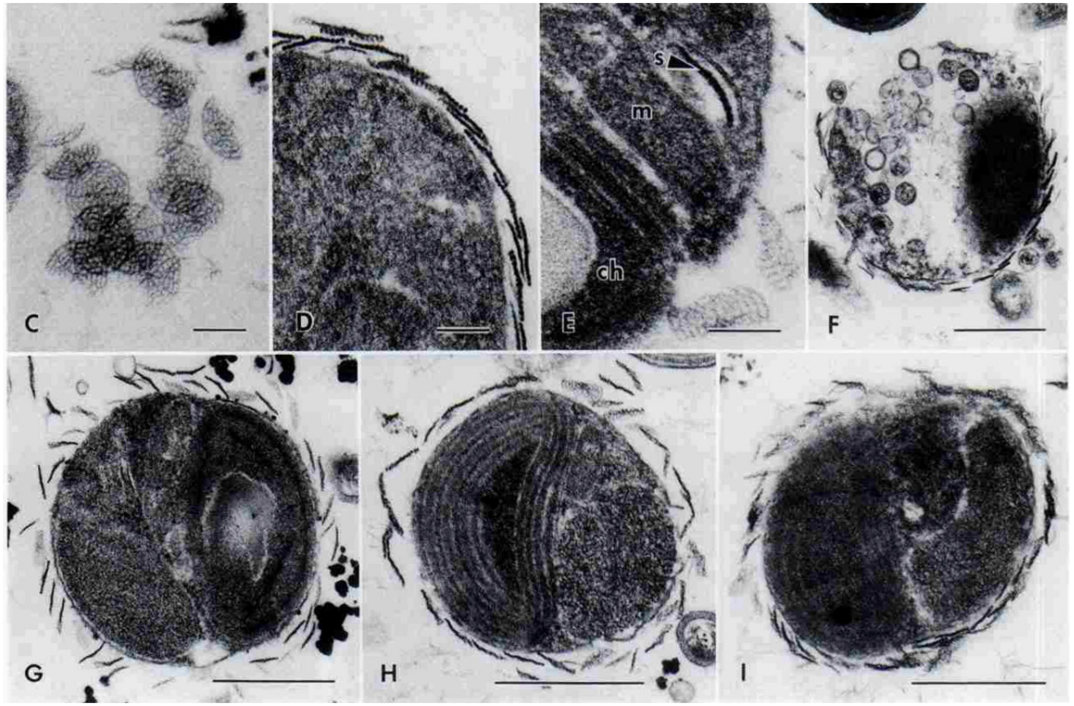
Picoplankton

1982 - John Sieburth - Electron microscopy



R.I. Alumni Bulletin
Oct 1962

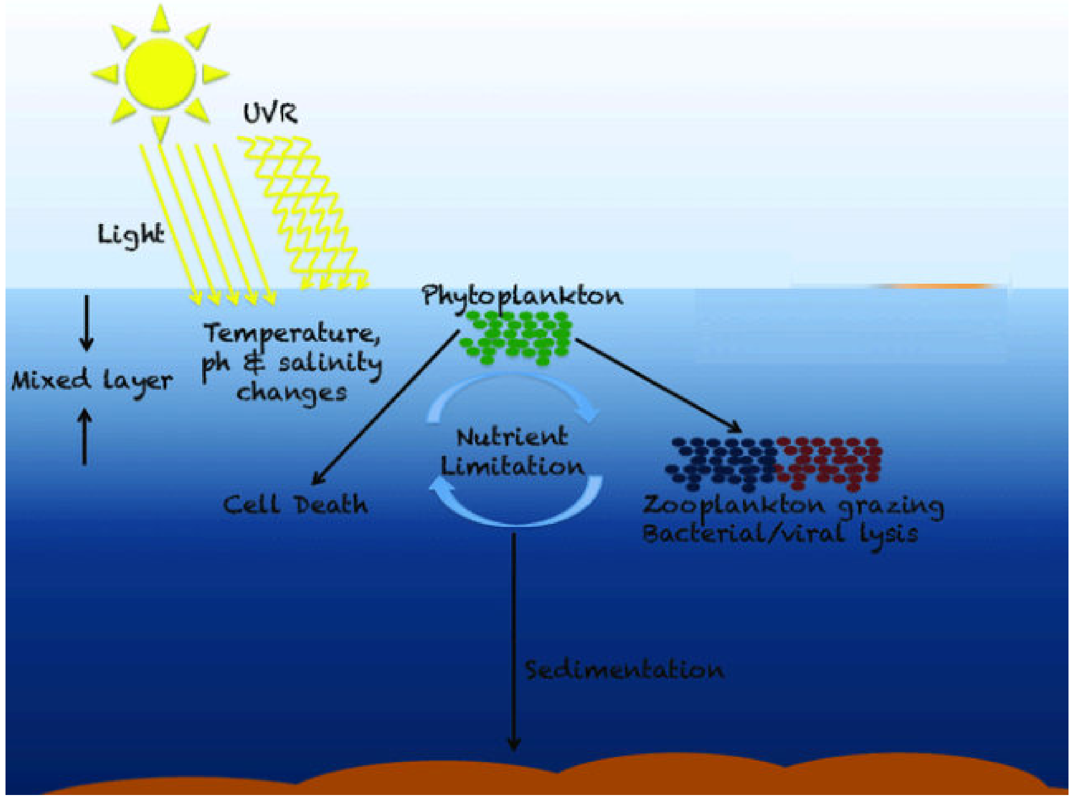
Preparing for Trident voyage are, left to right, Dr. John M. Sieburth, Stanley Chenoweth, Dr. John T. Conover, and Merrill Bracci.



What controls phytoplankton abundance ?

What factors control phytoplankton abundance and diversity ?

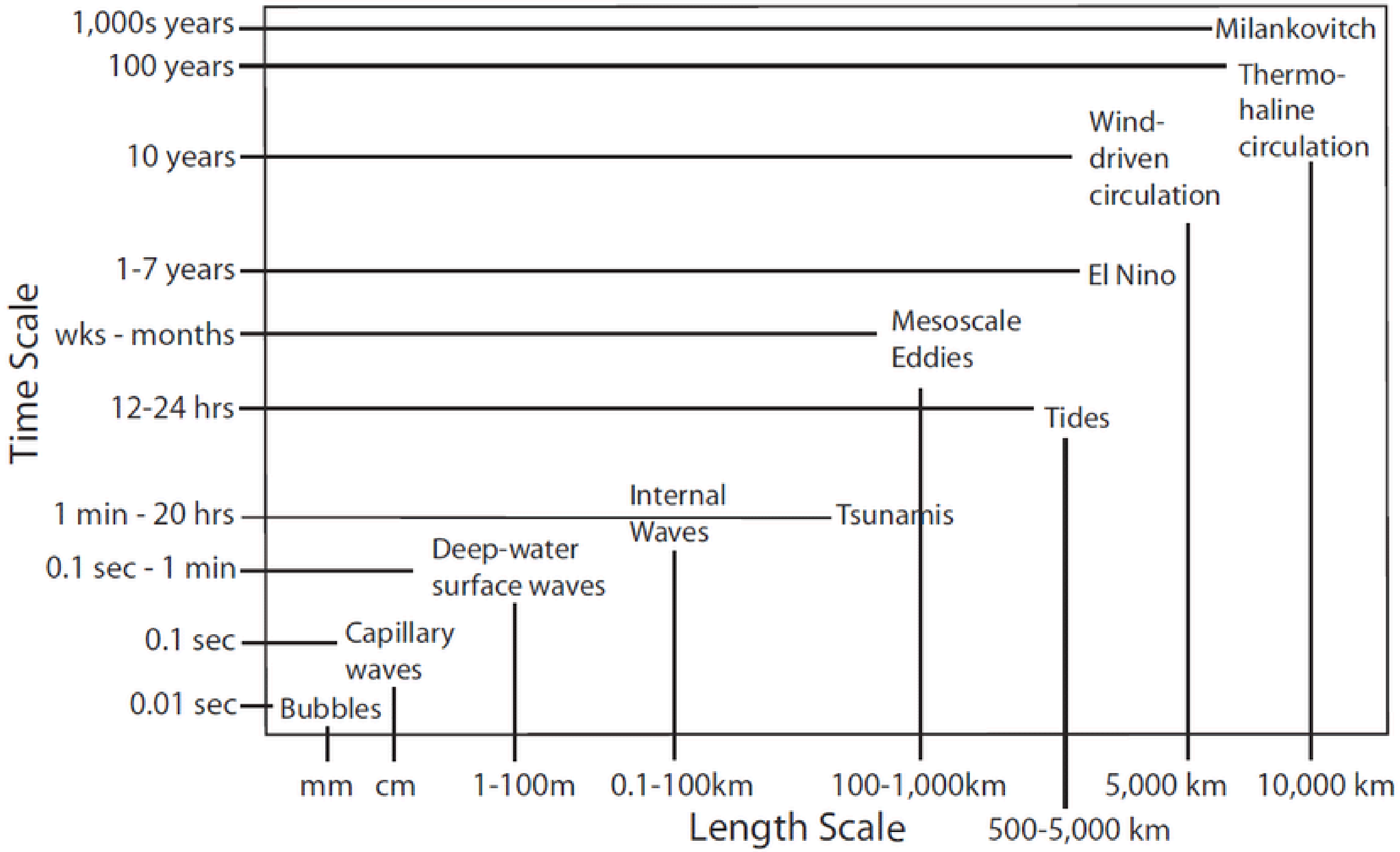
- Resources - Bottom up
  - Light
  - Nutrients (Nitrogen, Phosphorus)
  - Trace elements (Iron)
  - Temperature
  - Salinity
- Top down
  - Predation
  - Parasites (e.g. viruses)
  - Death
- Species selection
- These factors act at different scales



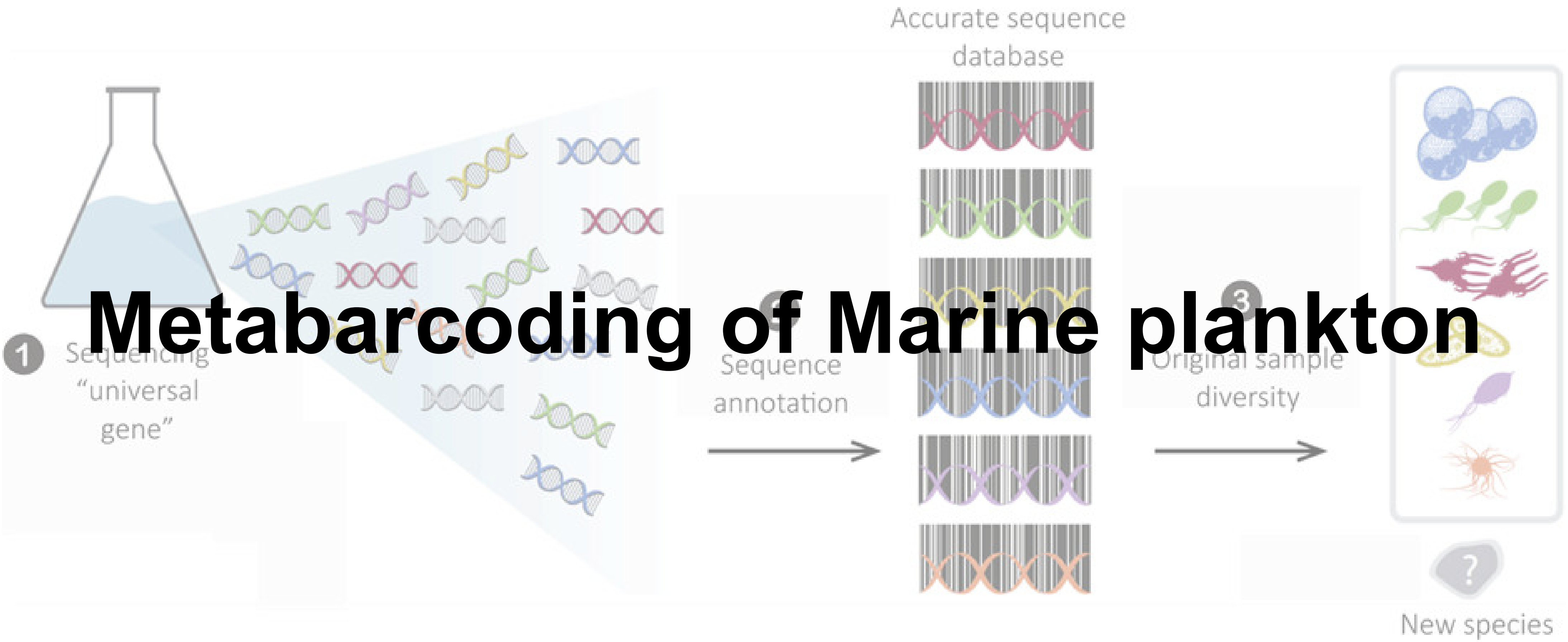
Scales - Space and Time

Can you name some scales in the ocean?

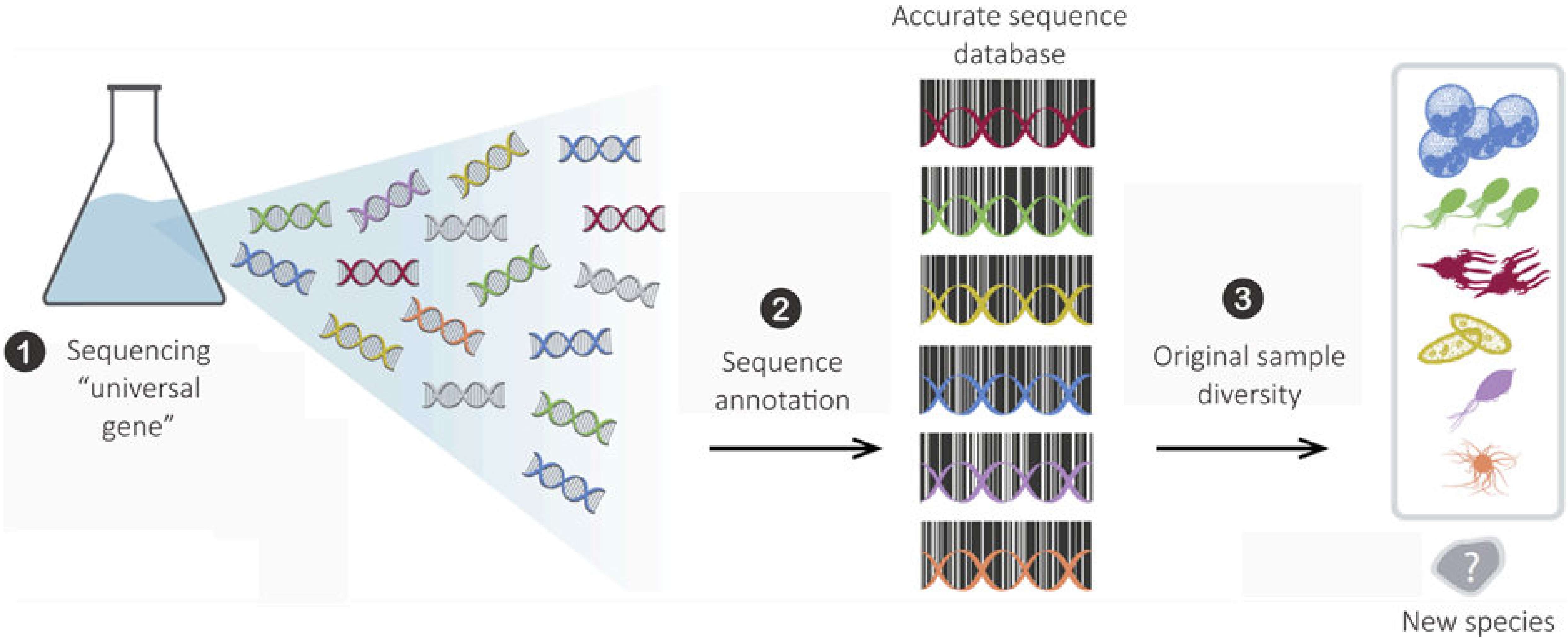
- Spatial
  - Horizontal
    - Basin
    - Region (upwelling)
    - Mesoscale
  - Vertical
    - Water column
    - Euphotic zone
- Temporal
  - Geological
  - Climate change
  - Climate oscillation
  - Eddies
  - Tides
  - Waves



Metabarcoding of Marine plankton

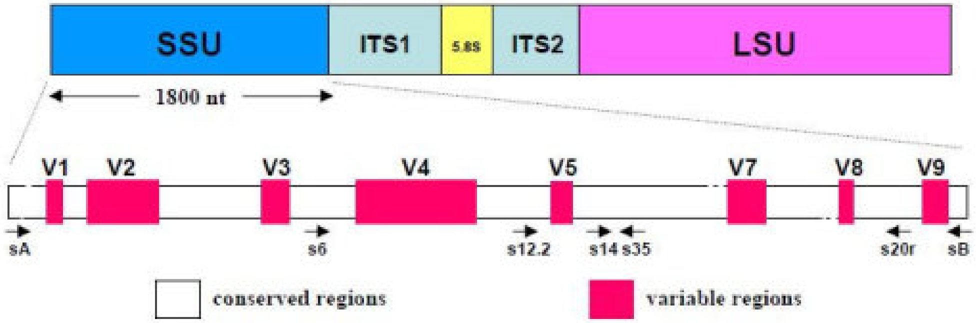


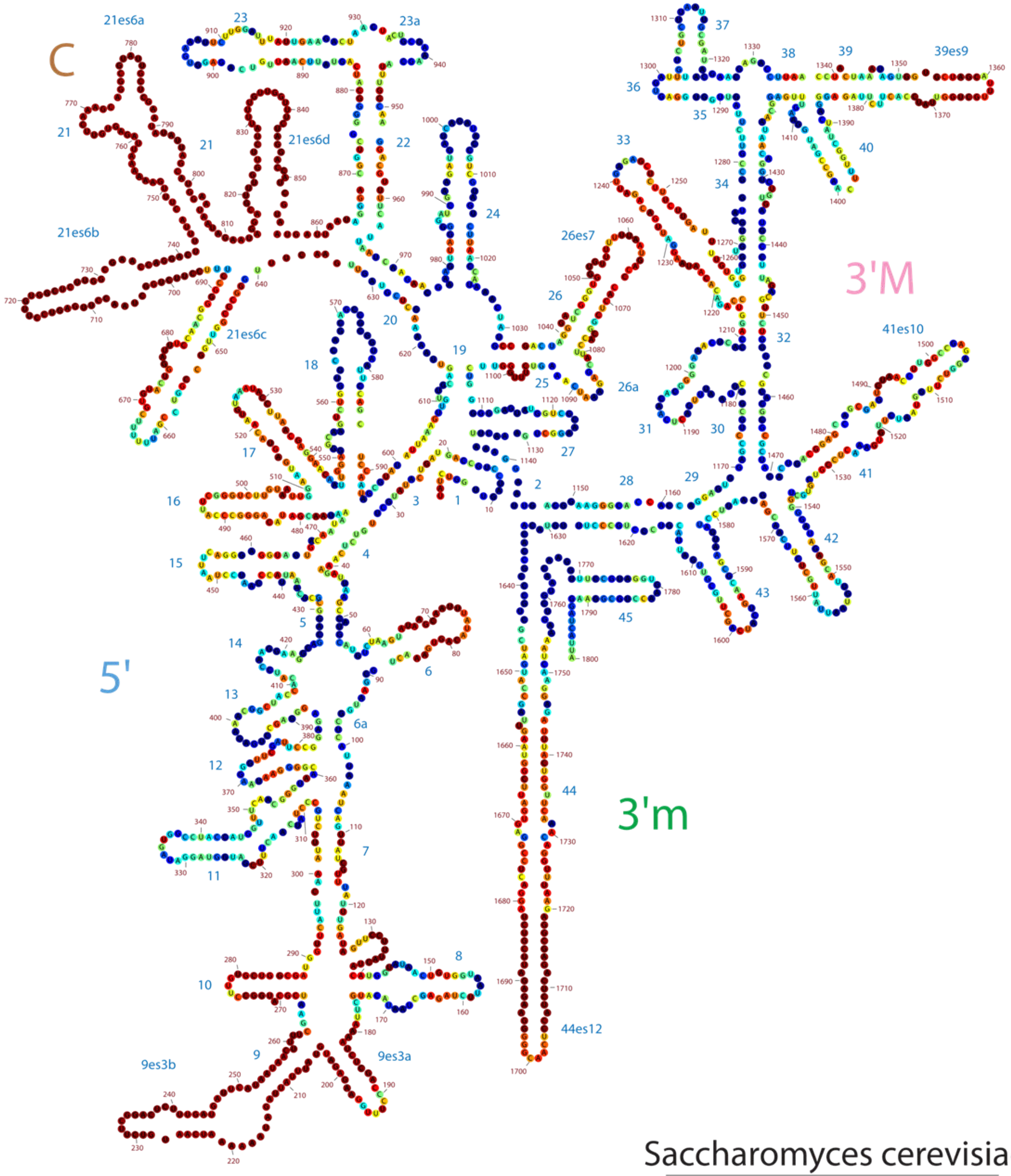
Principle


Accurate sequence database
1 Sequencing "universal gene"
2 Sequence annotation
3 Original sample diversity
?
New species

Target gene

- 18S rRNA
  - ITS (Internally Transcribed Spacer)
  - 16S plastid
  - rbcL



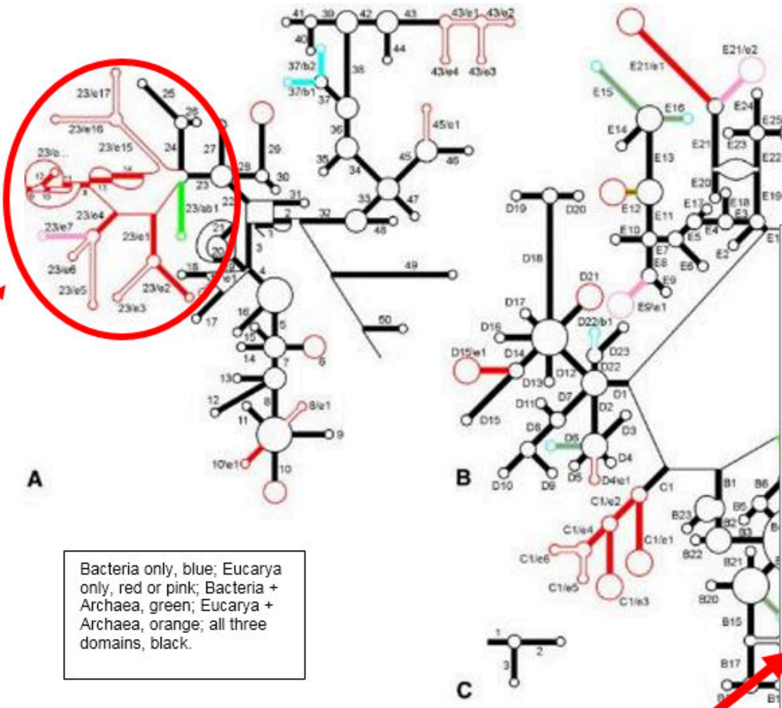


Saccharomyces cerevisiae
small subunit ribosomal RNA

Amplified region

V4

- Long region
- Presence of introns
- Resolutive
- Many reference sequences available

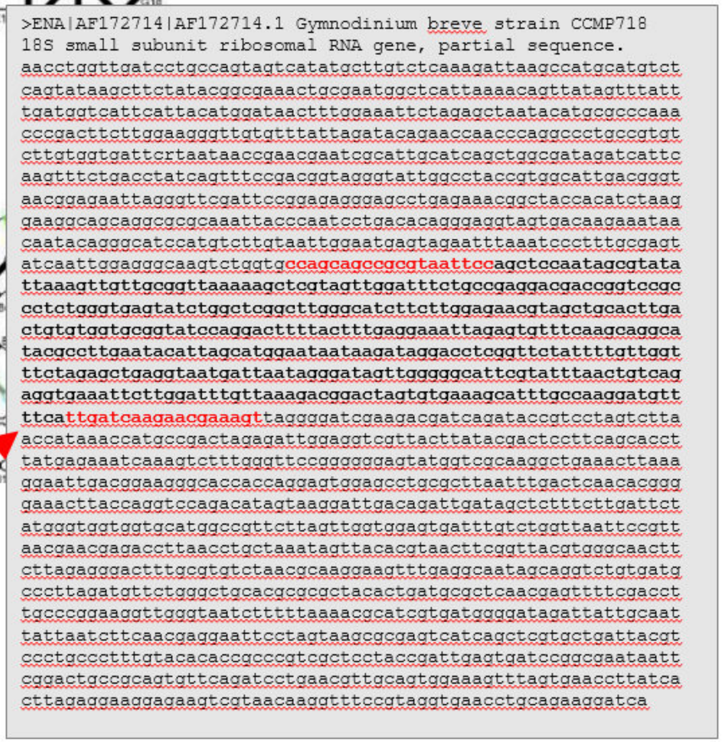


V4

Bacteria only, blue; Eucarya only, red or pink; Bacteria + Archaea, green; Eucarya + Archaea, orange; all three domains, black.

Wuyts J et al. Nucl. Acids Res. 2001;29:5017-5028

Av. size 382 nt

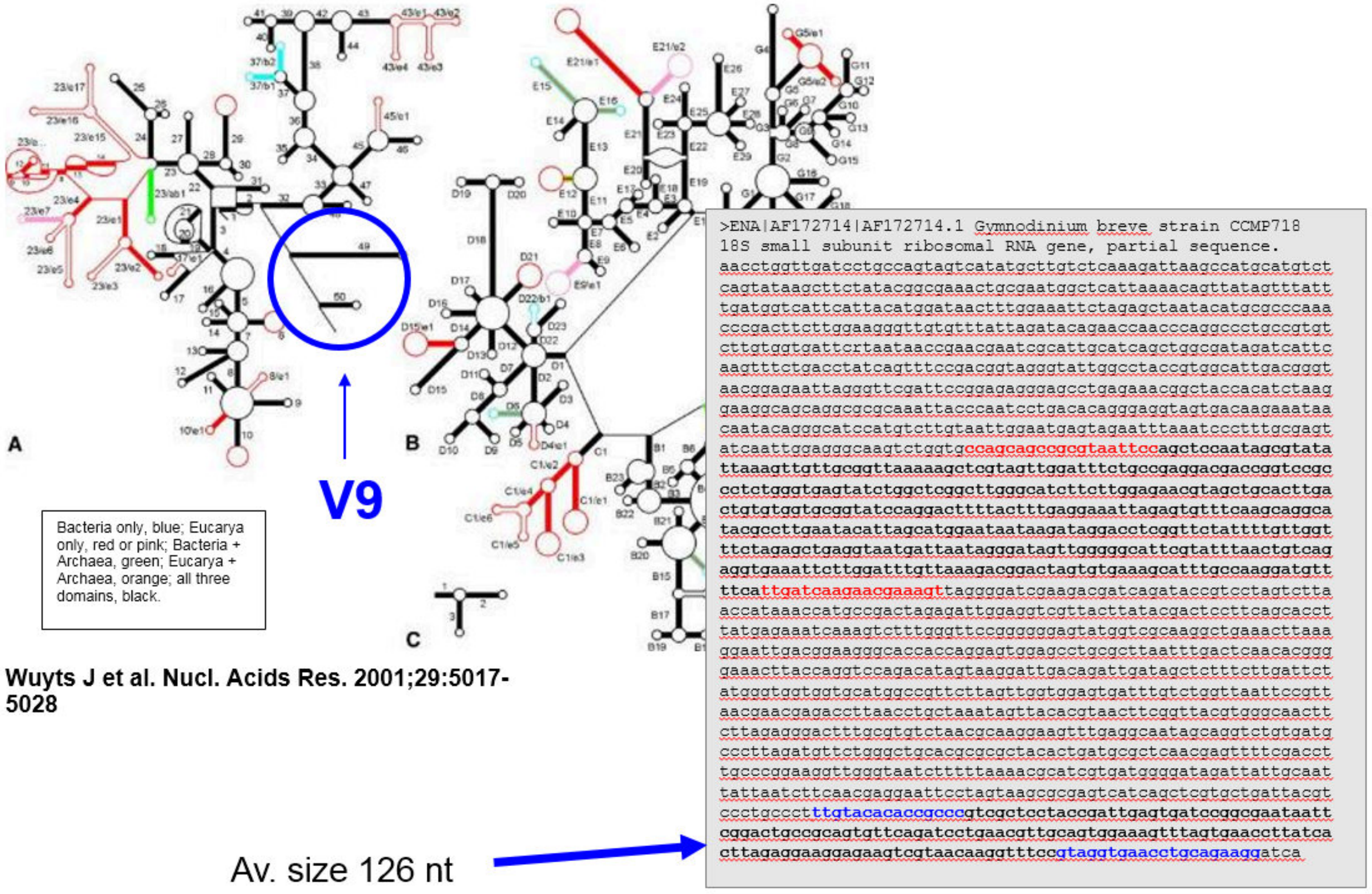


>ENA|AF172714|AF172714.1 Gymnodinium breve strain CCMP718
18S small subunit ribosomal RNA gene, partial sequence.
aacctggttgatcctgccagtagtcatatgcttgtctcaaagattaagccatgcatgtct
cagtataagcttctatacggcgaaactgcgaatggctcattaaaacagttatagtttatt
tgatggtcattcattacatggataactttggaaattctagagctaatacatgcgcccaaa
cccgacttcttggaagggttgtgtttattagatacagaaccaacccaggccctgccgtgt
cttgtggtgattcrtaataaccgaacgaatcgcattgcatcagctggcgatagatcattc
aagtttctgacctatcagtttccgacggtagggtattggcctaccgtggcattgacgggt
aacggagaattagggttcgattccggagagggagcctgagaaacggctaccacatctaag
gaaggcagcaggcgcgcaaattacccaatcctgacacagggaggtagtgacaagaaataa
caatacagggcatccatgtcttgtaattggaatgagtagaatttaaatccctttgcgagt
atcaattggagggcaagtctggtg**ccagcagccgcgtaattccagctccaatagcgtata**
ttaaagttgttgcggttaaaaagctcgtagttggatttctgccgaggacgaccggtccgc
cctctgggtgagtatctggctcggcttgggcatcttcttggagaacgtagctgcacttga
ctgtgtggtgcggtatccaggacttttactttgaggaaattagagtgtttcaagcaggca
tacgccttgaatacattagcatggaataataagataggacctcggttctattttgttggt
ttctagagctgaggtaatgattaatagggatagttgggggcattcgtatttaactgtcag
aggtgaaattcttggatttgttaaagacggactagtgtgaaagcatttgccaaggatgtt
ttcattgatcaagaacgaaagttaggggatcgaagacgatcagataccgtcctagtctta
accataaaccatgccgactagagattggaggtcgttacttatacgactccttcagcacct
tatgagaaatcaaagtctttgggttccgggggagtatggtcgcaaggctgaaacttaaa
ggaattgacggaagggcaccaccaggagtggagcctgcgcttaatttgactcaacacggg
gaaacttaccaggtccagacatagtaaggattgacagattgatagctctttcttgattct
atgggtggtggtgcatggccgttcttagttggtggagtgatttgtctggttaattccgtt
aacgaacgagaccttaacctgctaaatagttacacgtaacttcggttacgtgggcaactt
cttagagggactttgcgtgtctaacgcaaggaagtttgaggcaatagcaggtctgtgatg
cccttagatgttctgggctgcacgcgcgctacactgatgcgctcaacgagttttcgacct
tgcccggaaggttgggtaatctttttaaaacgcatcgtgatggggatagattattgcaat
tattaatcttcaacgaggaattcctagtaagcgcgagtcatcagctcgtgctgattacgt
ccctgccctttgtacacaccgcccgtcgctcctaccgattgagtgatccggcgaataatt
cggactgccgcagtgttcagatcctgaacgttgcagtggaaagtttagtgaaccttatca
cttagaggaaggagaagtcgtaacaaggtttccgtaggtgaacctgcagaaggatca

Amplified region

V9

- Short region (cheaper)
- Fewer introns
- Sometimes more resolutive than V4
- Fewer reference sequences available



Bacteria only, blue; Eucarya only, red or pink; Bacteria + Archaea, green; Eucarya + Archaea, orange; all three domains, black.

Wuyts J et al. Nucl. Acids Res. 2001;29:5017-5028

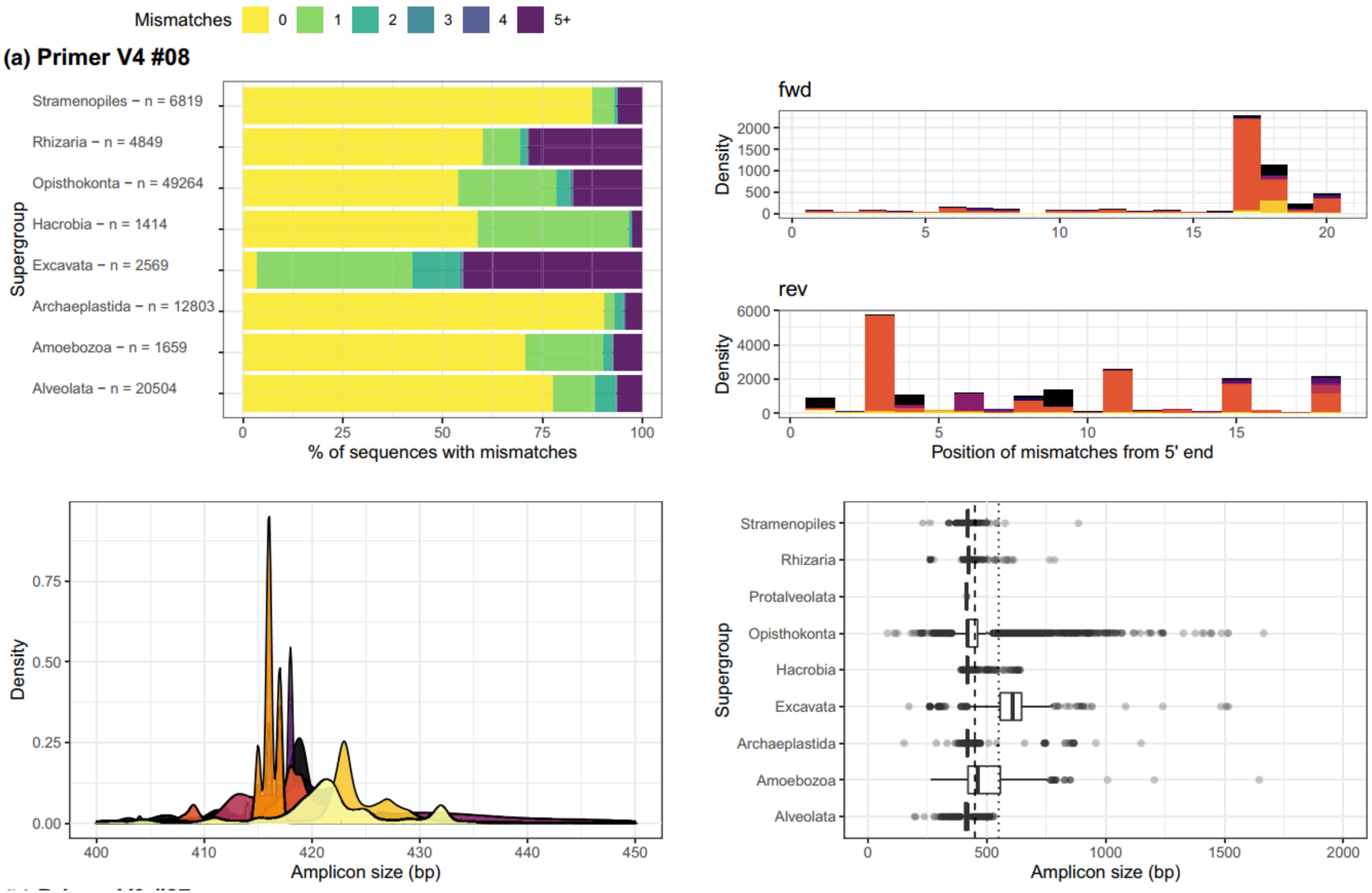
```
>ENA|AF172714|AF172714.1 Gymnodinium breve strain CCMP718
18S small subunit ribosomal RNA gene, partial sequence.
aacctggttgatcctgccagtagtcatatgcttgtctcaaagattaagccatgcatgtct
cagtataagcttctatacggcgaaactgcgaatggctcattaaaacagttatagtttatt
tgatggtcattcattacatggataactttggaaattctagagctaatacatgcgcccaaa
cccgacttcttggaagggttgtgtttattagatacagaaccaacccaggccctgccgtgt
cttgtggtgattcrtaataaccgaacgaatcgcattgcatcagctggcgatagatcattc
aagtttctgacctatcagtttccgacggtagggtattggcctaccgtggcattgacgggt
aacggagaattagggttcgattccggagagggagcctgagaaacggctaccacatctaag
gaaggcagcaggcgcgcaaattacccaatcctgacacagggaggtagtgacaagaaataa
caatacagggcatccatgtcttgtaattggaatgagtagaatttaaatccctttgcgagt
atcaattggagggcaagtctggtgccagcagccgcgtaattccagctccaatagcgtata
ttaaagttgttgcggttaaaaagctcgtagttggatttctgccgaggacgaccggtccgc
cctctgggtgagtatctggctcggcttgggcatcttcttggagaacgtagctgcacttga
ctgtgtggtgcggtatccaggacttttactttgaggaaattagagtgtttcaagcaggca
tacgccttgaatacattagcatggaataataagataggacctcggttctattttgttggt
ttctagagctgaggtaatgattaatagggatagttgggggcattcgtatttaactgtcag
aggtgaaattcttggatttgttaaagacggactagtgtgaaagcatttgccaaggatgtt
ttcattgatcaagaacgaaagttaggggatcgaagacgatcagataccgtcctagtctta
accataaaccatgccgactagagattggaggtcgttacttatacgactccttcagcacct
tatgagaaatcaaagtctttgggttccggggggagtatggtcgcaaggctgaaacttaaa
ggaattgacggaagggcaccaccaggagtggagcctgcgcttaatttgactcaacacggg
gaaacttaccaggtccagacatagtaaggattgacagattgatagctctttcttgattct
atgggtggtggtgcatggccgttcttagttggtggagtgatttgtctggttaattccgtt
aacgaacgagaccttaacctgctaaatagttacacgtaacttcggttacgtgggcaactt
cttagagggactttgcgtgtctaacgcaaggaagtttgaggcaatagcaggtctgtgatg
cccttagatgttctgggctgcacgcgcgctacactgatgcgctcaacgagttttcgacct
tgcccggaaggttgggtaatcttttttaaaacgcatcgtgatggggatagattattgcaat
tattaatcttcaacgaggaattcctagtaagcgcgagtcatcagctcgtgctgattacgt
ccctgcccttttgtacacaccgcccgtcgctcctaccgattgagtgatccggcgaataatt
cggactgccgcagtgttcagatcctgaacgttgcagtggaaagtttagtgaaccttatca
cttagaggaaggagaagtcgtaacaaggtttccgtaggtgaacctgcagaaggatca
```

Av. size 126 nt

Primers

Many choices

- V4
- V4 specific
- V9
- Universal (euk+prok)



Vaulot, D., Mahé, F., Bass, D. & Geisen, S. pr2-primer: an 18S rRNA primer database for protists. Molecular Ecology Resources 22, 168–179

Assignment

Reference database

- Genbank
  - Taxonomy very bad
- Silva
  - OK for prokaryotes
  - Eukaryotes bad

SILVA

Welcome to the SILVA rRNA database project

A comprehensive on-line resource for quality checked and aligned ribosomal RNA sequence data.

SILVA provides comprehensive, quality checked and regularly updated datasets of aligned small (16S/18S, SSU) and large subunit (23S/28S, LSU) ribosomal RNA (rRNA) sequences for all three domains of life (*Bacteria*, *Archaea* and *Eukarya*).

SILVA are the official databases of the software package ARB.

For more background information → Click here

SILVAngs

Check out our service for Next Generation Amplicon data

SILVA Alignment, Classification and Tree (ACT) Service

The SILVA ACT service combines alignment, search and classify as well as reconstruction of trees in a single web application.

SILVA ACT is available at: → www.arb-silva.de/act

SILVA Tree Viewer

The **SILVA Tree Viewer** is a web application to browse and query the SILVA guide trees.

A technical preview is available at ↗www.arb-silva.de/treeviewer. The Tree Viewer for SILVA 132 LSU can be found ↗here.

News

21.12.2019

Merry Christmans & Happy New Year 2020

The SILVA Team wishes you a Merry Christmas & Happy New Year. Many thanks for all your feedback and support to improve SILVA and SILVAngs. Looking forward to see you again in 2020.

16.12.2019

SILVA 138 SSU released

Finally it is done... SILVA 138 SSU is released. LSU will follow in Spring 2020. Please note: The Treeviewer and SILVAngs are still on SILVA 132. Updates are planned for January 2020. Many thanks for your patience, the next release will come faster (hopefully).

24.10.2019

SILVA SSU Release 138 in November

We know you all are looking forward to the release of SILVA SSU 138. We are currently in the last steps of creating the release and plan to release SILVA SSU 138 in the week of November 11. We are sorry for the additional delay and the inconveniences it may cause.

03.09.2019

Update on SILVA Release 138

We are sorry to inform you that the SILVA Release 138 is delayed further and we have decided to release the SSU and LSU datasets separately to compensate for the delay. We estimate the SSU datasets to be available in October and the LSU datasets by the end of the year. This news article will give you some background information on the release.

go to Archive ->

User satisfaction survey

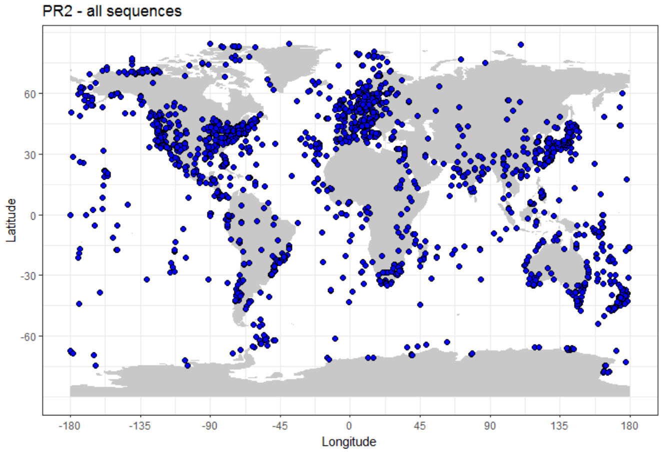
SILVA is now part of the German Network for Bioinformatics Infrastructure **de.NBI**.

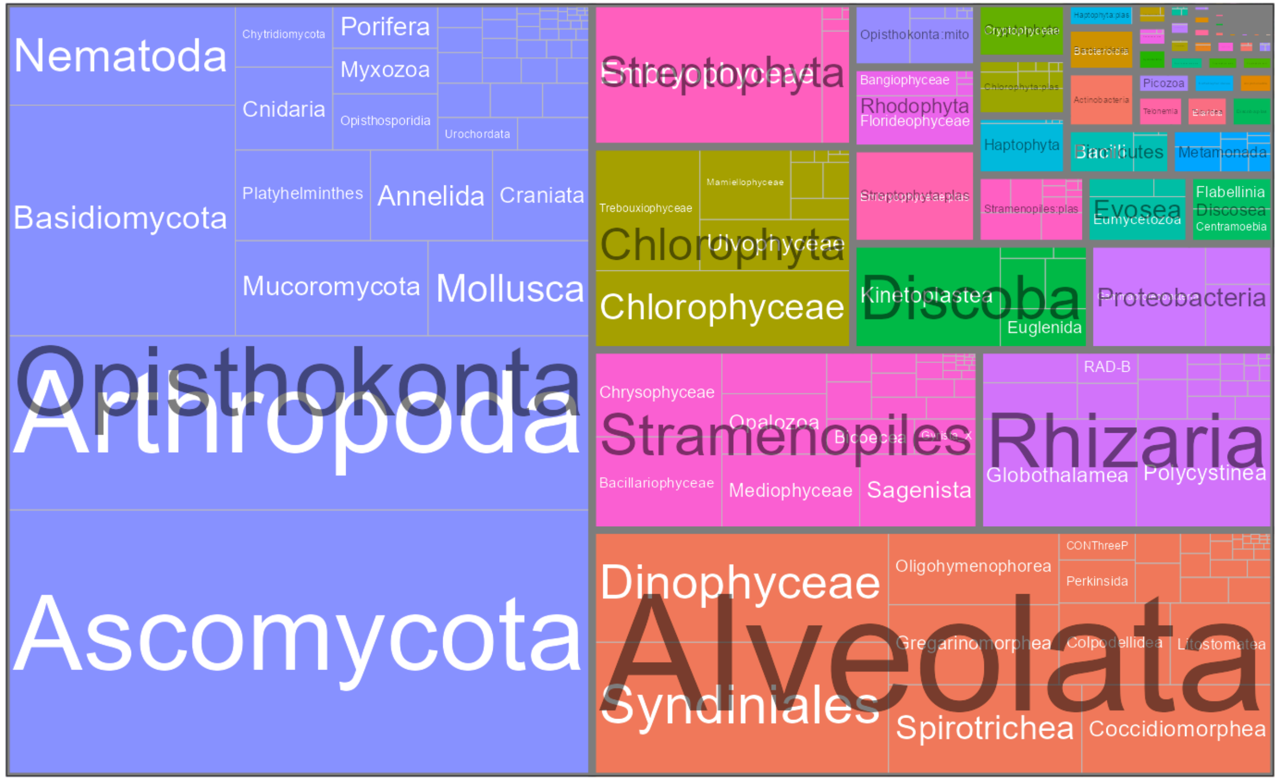
To evaluate and improve our quality of service we need your feedback. **Please help us** by participating in this short ↗survey.

Assignment

The PR2 database

- 221 091 sequences
    - nuclear 18S rRNA
    - plastid 16S rRNA (PhytoRef)
    - bacteria and archaea 16S to check contamination
- Unified taxonomy (9 ranks from kingdom to species)
- Metadata (e.g. coordinates, environment)
- Available as web interface, flat files or as R package
- Latest version: 5.0 (released 05-2023)



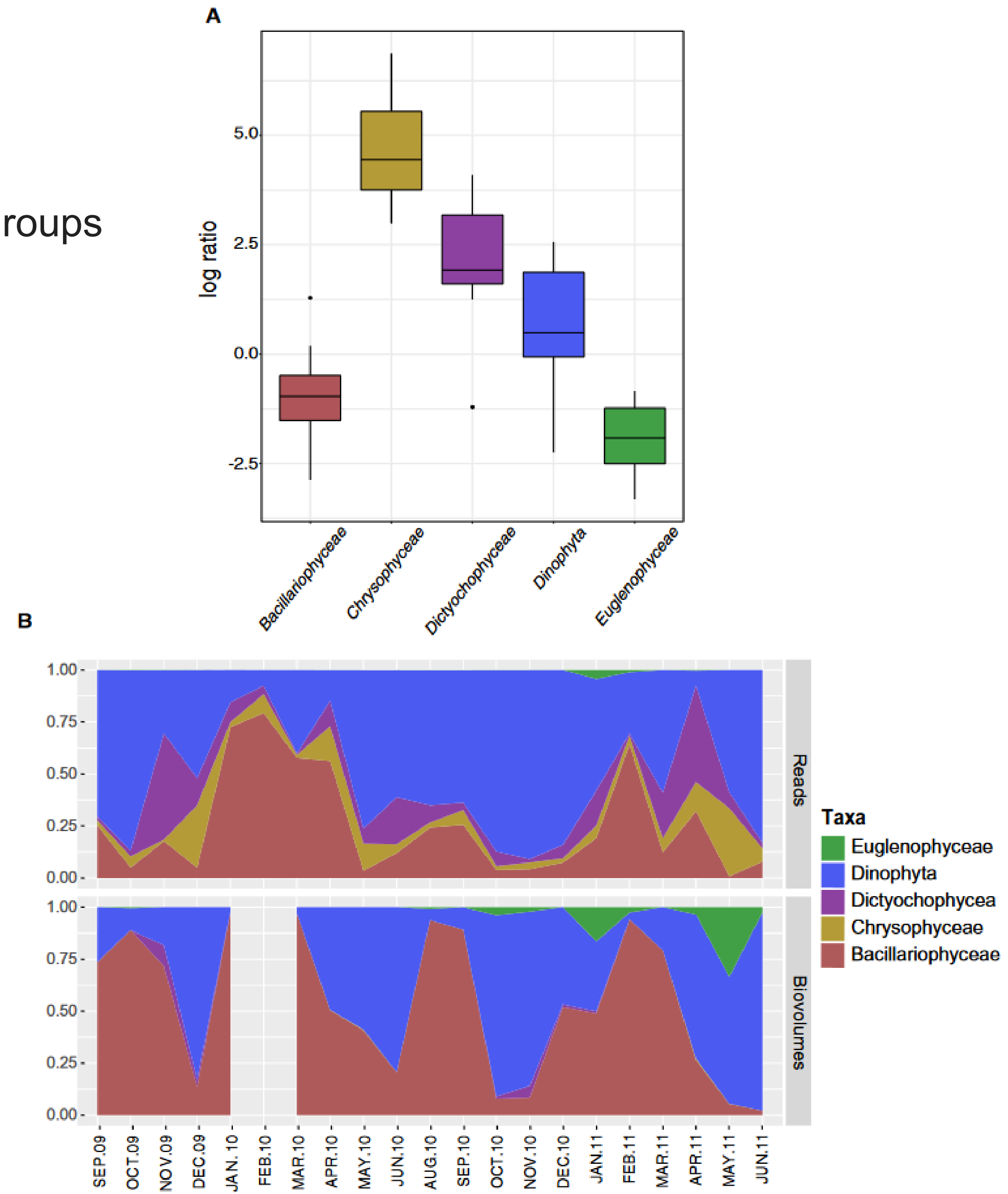


Guillou et al. 2013. The Protist Ribosomal Reference database (PR2): a catalog of unicellular eukaryote Small Sub-Unit rRNA sequences with

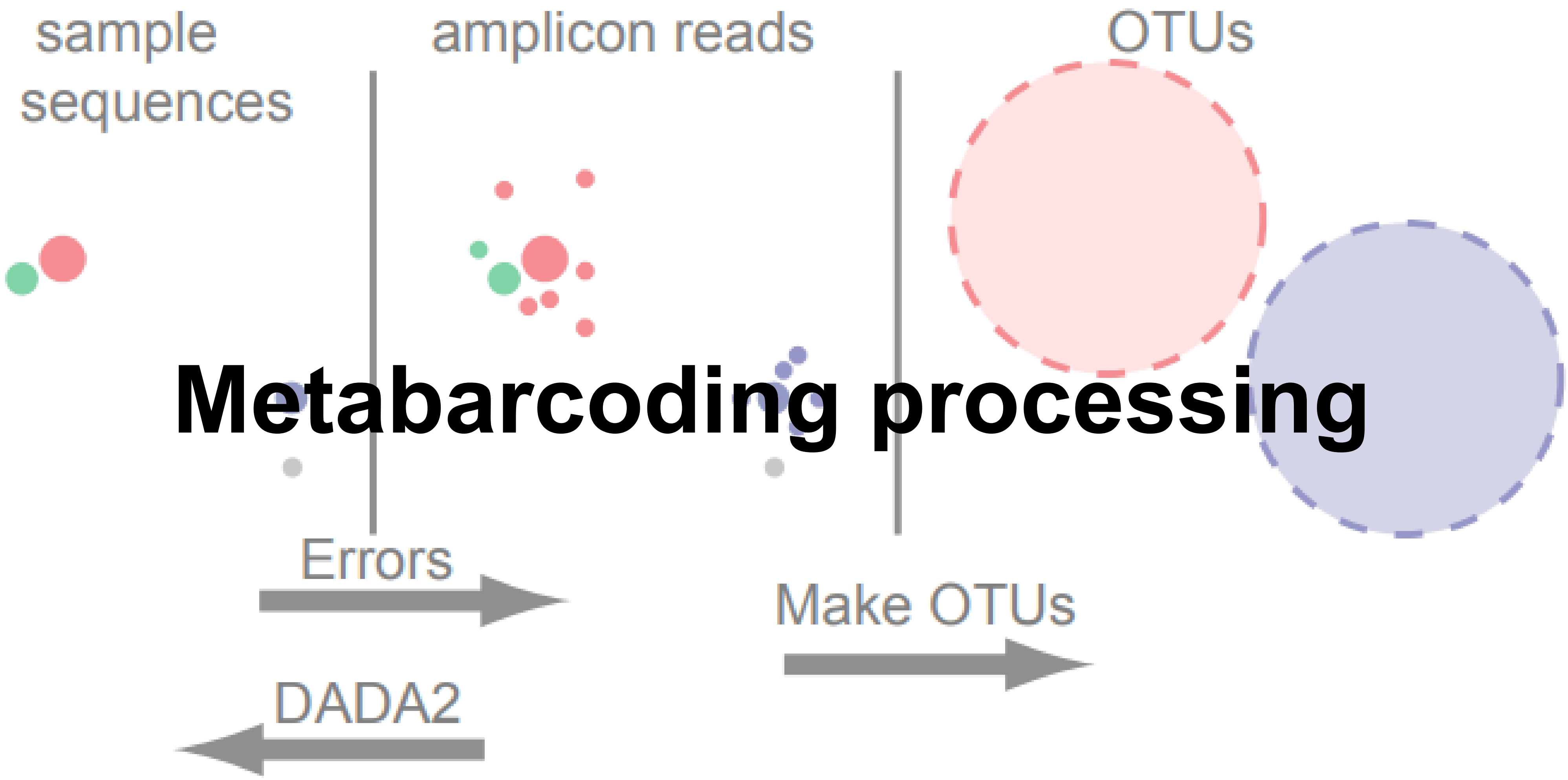
Caveats

- Abundances are **relative** abundances
  - Can be hard to compare different samples
  - Relative abundance depends on abundance of other groups
  - Compare with microscopy counts

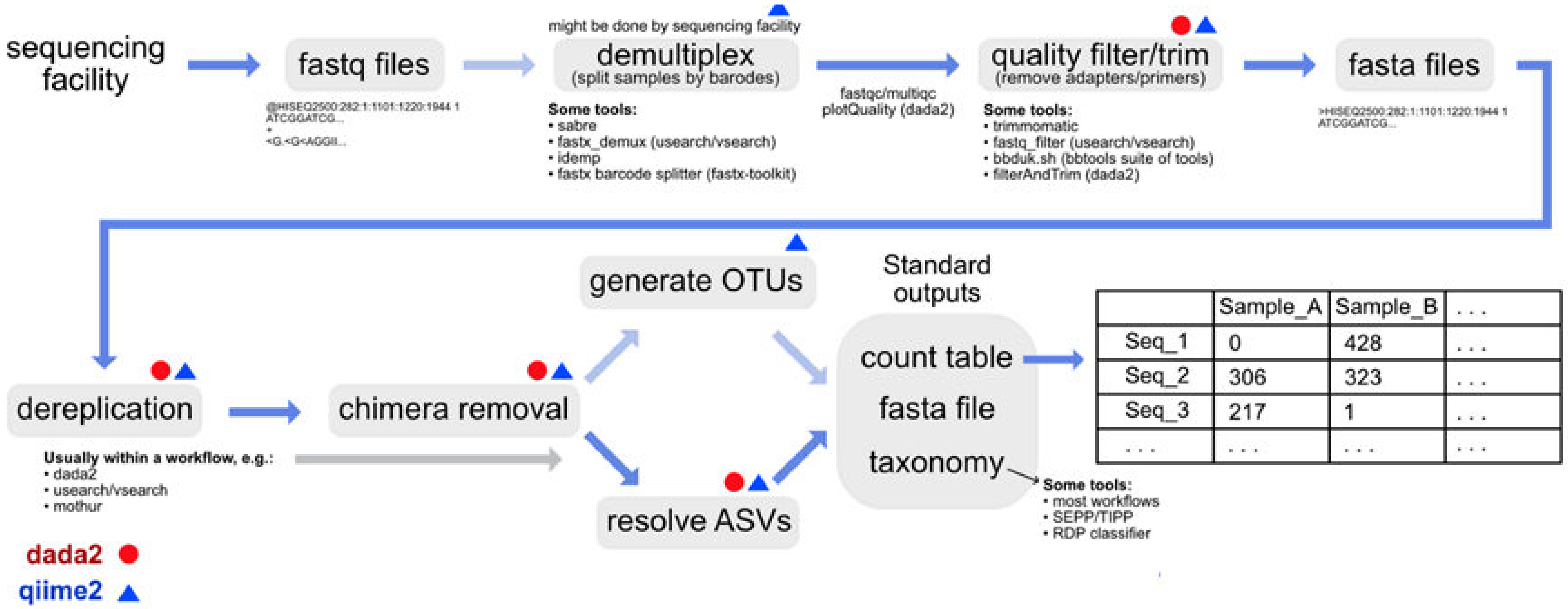
- Biases due to
  - Filtration
  - DNA vs. RNA (total vs. active)
- Depends on reference database
  - Many species do not have reference sequence
  - Need cultures from sampling site
  - Single cell gene sequencing
- Different processing and reference databases
  - Reprocess primary data
    - metaPR2
    - EukBanks



Metabarcoding processing



Metabarcoding pipeline

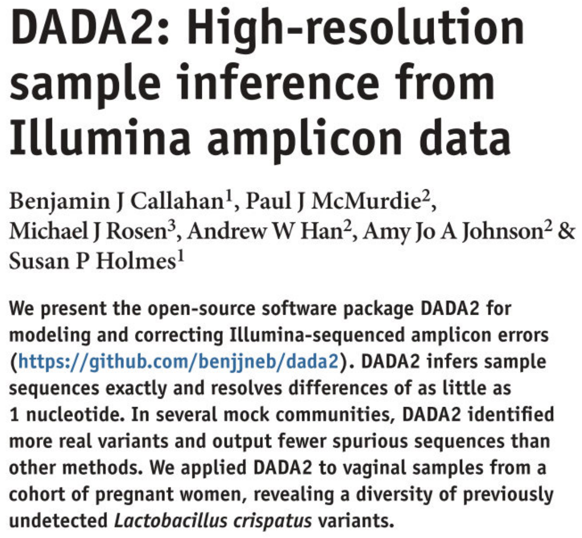


	Sample_A	Sample_B	. . .
Seq_1	0	428	. . .
Seq_2	306	323	. . .
Seq_3	217	1	. . .
. . .	. . .	. . .	. . .

DADA2: Fast and accurate sample inference from amplicon data with single-nucleotide resolution

DADA2

Amplicon Sequencing. **Exactly.** Version 1.18



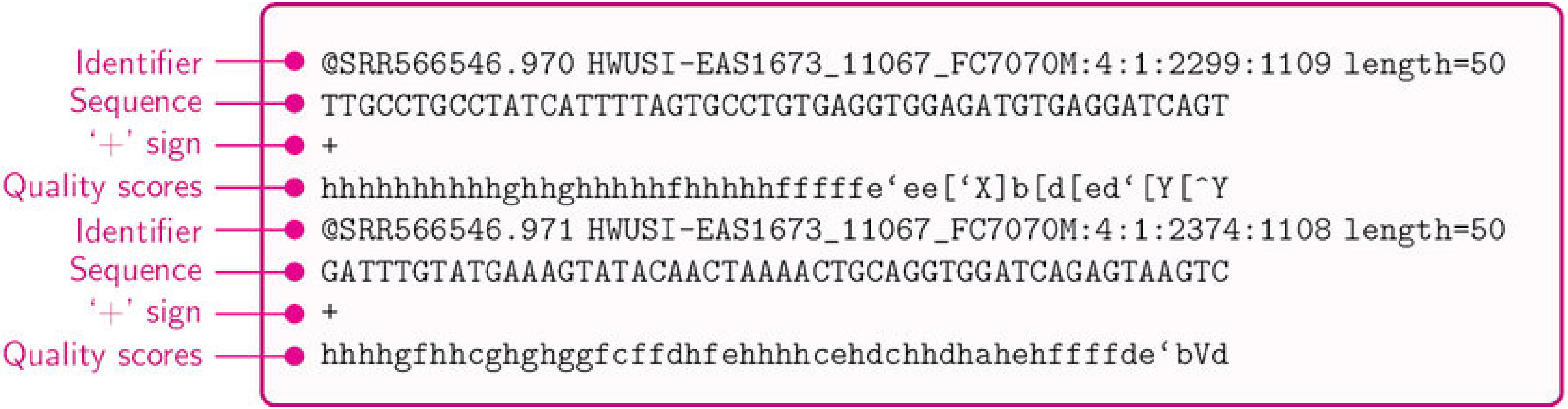
DADA2: High-resolution sample inference from Illumina amplicon data

Benjamin J Callahan[1], Paul J McMurdie[2], Michael J Rosen[3], Andrew W Han[2], Amy Jo A Johnson[2] & Susan P Holmes[1]

We present the open-source software package DADA2 for modeling and correcting Illumina-sequenced amplicon errors (https://github.com/benjjneb/dada2). DADA2 infers sample sequences exactly and resolves differences of as little as 1 nucleotide. In several mock communities, DADA2 identified more real variants and output fewer spurious sequences than other methods. We applied DADA2 to vaginal samples from a cohort of pregnant women, revealing a diversity of previously undetected *Lactobacillus crispatus* variants.

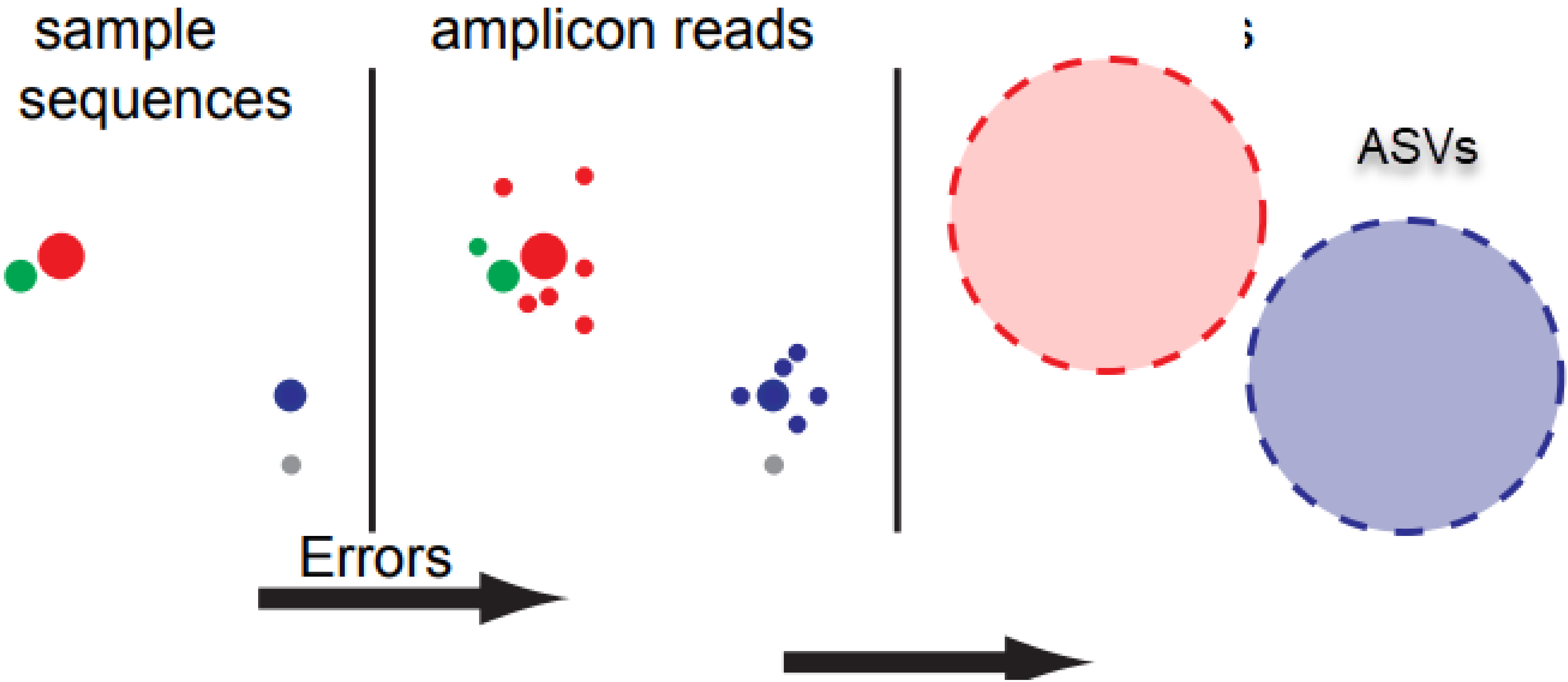
Sequences

Fastq files

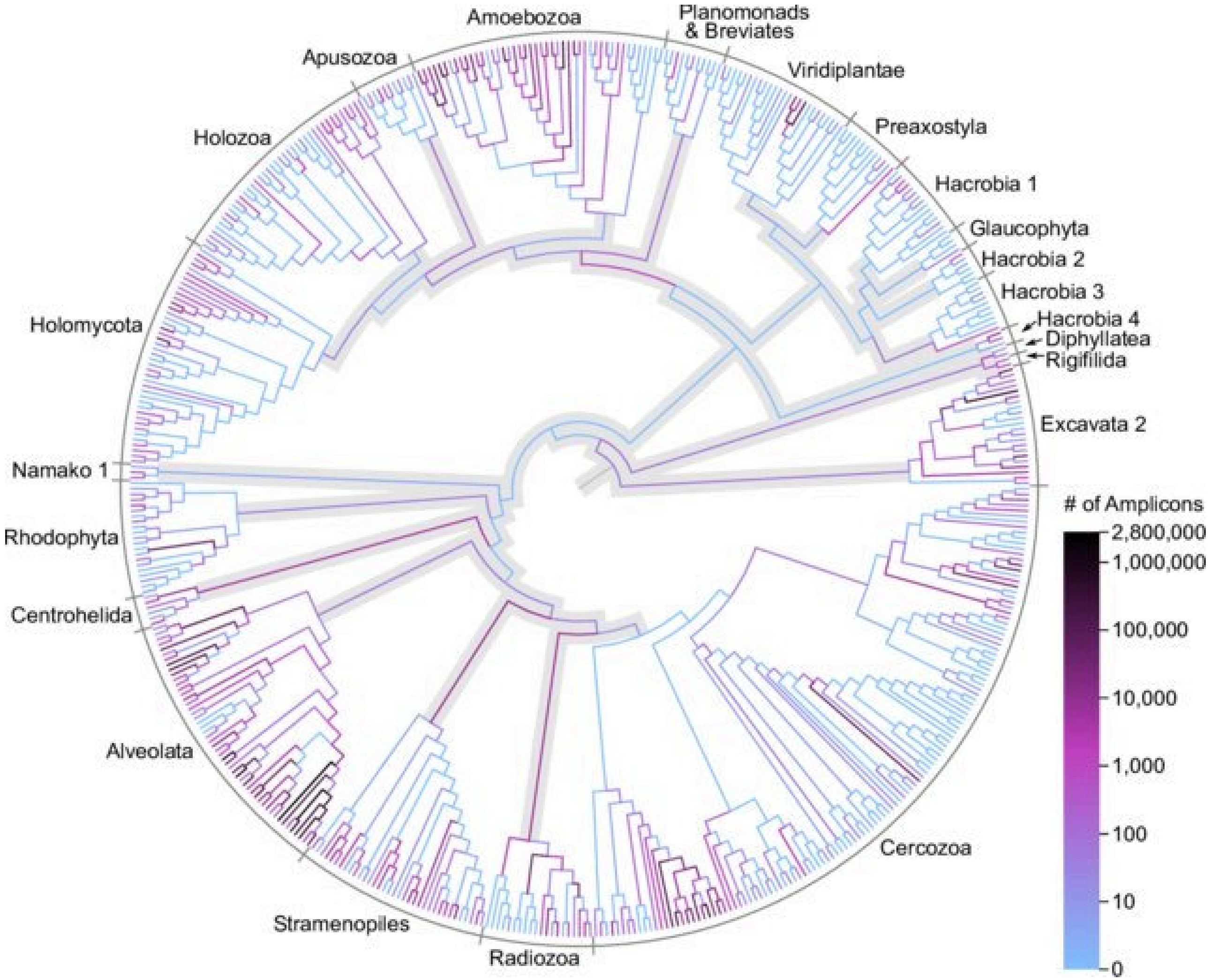


```
@SRR566546.970 HWUSI-EAS1673_11067_FC7070M:4:1:2299:1109 length=50
TTGCCTGCCTATCATTTTAGTGCCTGTGAGGTGGAGATGTGAGGATCAGT
+
hhhhhhhhhhghhghhhhhfhhhhhfffffe‘ee[‘X]b[d[ed‘[Y[^Y
@SRR566546.971 HWUSI-EAS1673_11067_FC7070M:4:1:2374:1108 length=50
GATTTGTATGAAAGTATACAACTAAAACTGCAGGTGGATCAGAGTAAGTC
+
hhhhgfhhcghghggfcffdhfehhhhcehdchhdhahehffffde‘bVd
```

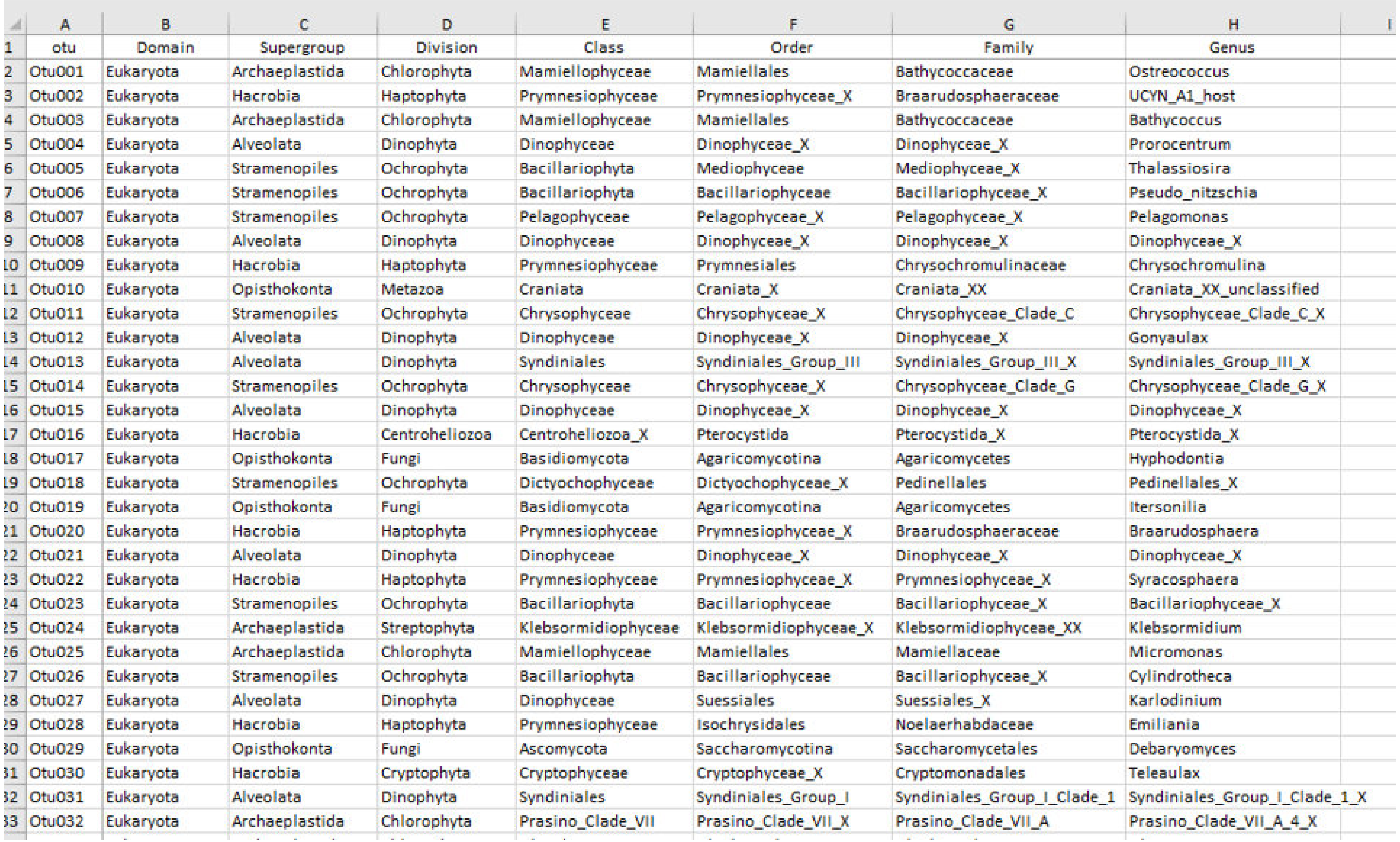
Cluster



Assign

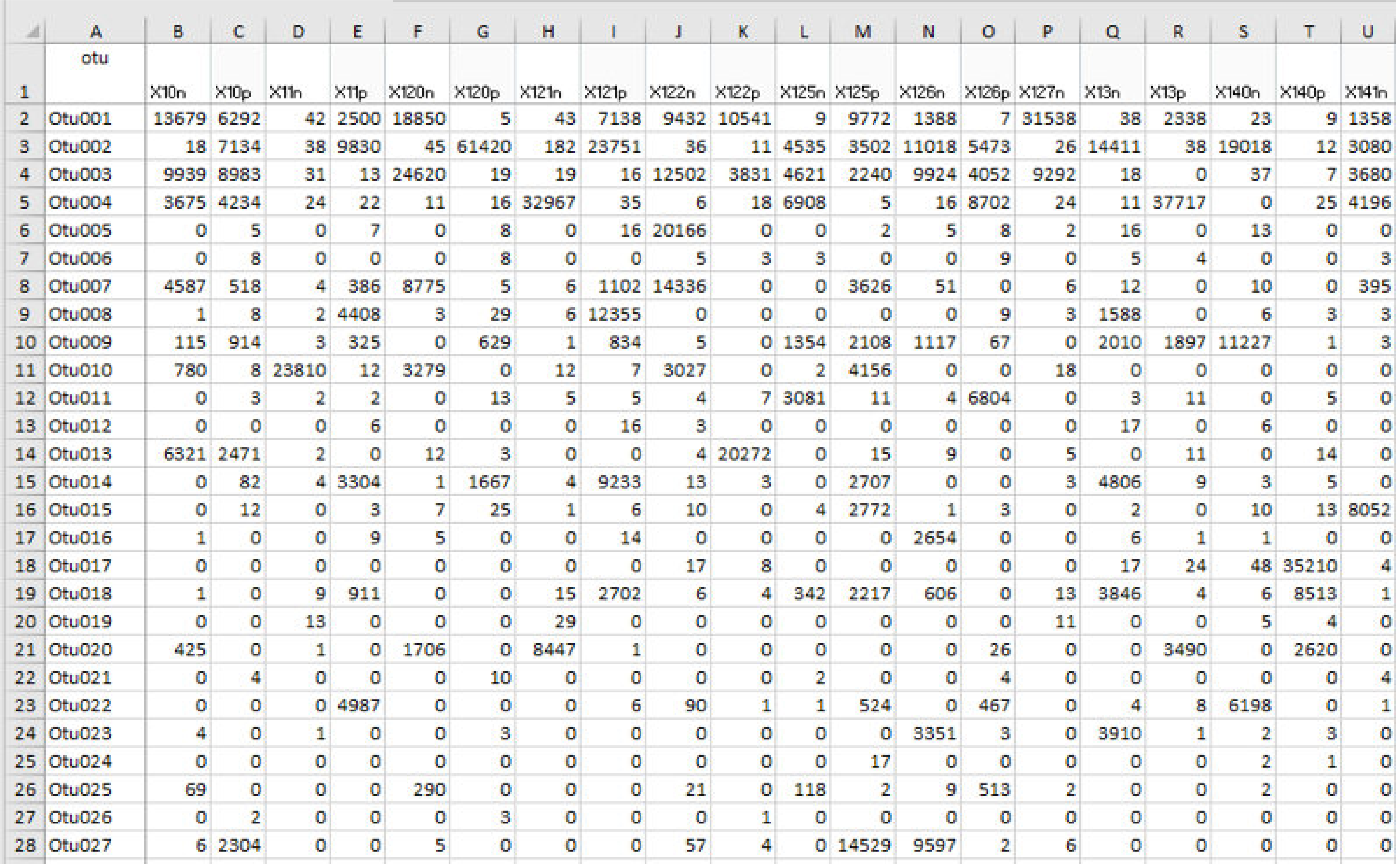

Amoebozoa
Planomonads & Breviates
Apusozoa
Viridiplantae
Preaxostyla
Holozoa
Hacrobia 1
Glaucophyta
Hacrobia 2
Hacrobia 3
Holomycota
Hacrobia 4
Diphyllatea
Rigifilida
Excavata 2
Namako 1
# of Amplicons
Rhodophyta
2,800,000
1,000,000
Centrohelida
100,000
10,000
Alveolata
1,000
100
Cercozoa
10
Stramenopiles
Radiozoa
0

Data tables - ASVs



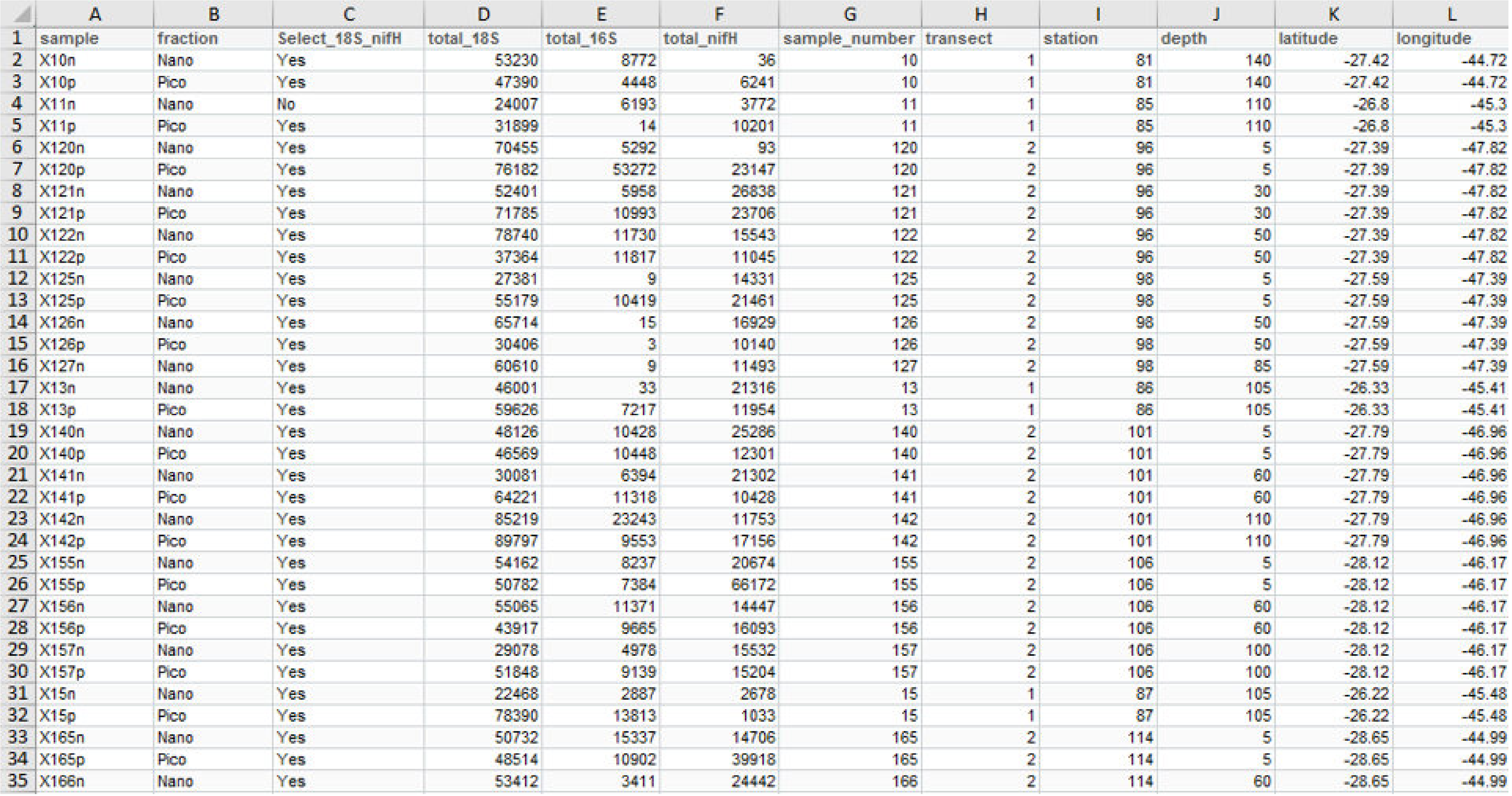
	A	B	C	D	E	F	G	H
1	otu	Domain	Supergroup	Division	Class	Order	Family	Genus
2	Otu001	Eukaryota	Archaeplastida	Chlorophyta	Mamiellophyceae	Mamiellales	Bathycoccaceae	Ostreococcus
3	Otu002	Eukaryota	Hacrobia	Haptophyta	Prymnesiophyceae	Prymnesiophyceae_X	Braarudosphaeraceae	UCYN_A1_host
4	Otu003	Eukaryota	Archaeplastida	Chlorophyta	Mamiellophyceae	Mamiellales	Bathycoccaceae	Bathycoccus
5	Otu004	Eukaryota	Alveolata	Dinophyta	Dinophyceae	Dinophyceae_X	Dinophyceae_X	Prorocentrum
6	Otu005	Eukaryota	Stramenopiles	Ochrophyta	Bacillariophyta	Mediophyceae	Mediophyceae_X	Thalassiosira
7	Otu006	Eukaryota	Stramenopiles	Ochrophyta	Bacillariophyta	Bacillariophyceae	Bacillariophyceae_X	Pseudo_nitzschia
8	Otu007	Eukaryota	Stramenopiles	Ochrophyta	Pelagophyceae	Pelagophyceae_X	Pelagophyceae_X	Pelagomonas
9	Otu008	Eukaryota	Alveolata	Dinophyta	Dinophyceae	Dinophyceae_X	Dinophyceae_X	Dinophyceae_X
10	Otu009	Eukaryota	Hacrobia	Haptophyta	Prymnesiophyceae	Prymnesiales	Chrysochromulinaceae	Chrysochromulina
11	Otu010	Eukaryota	Opisthokonta	Metazoa	Craniata	Craniata_X	Craniata_XX	Craniata_XX_unclassified
12	Otu011	Eukaryota	Stramenopiles	Ochrophyta	Chrysophyceae	Chrysophyceae_X	Chrysophyceae_Clade_C	Chrysophyceae_Clade_C_X
13	Otu012	Eukaryota	Alveolata	Dinophyta	Dinophyceae	Dinophyceae_X	Dinophyceae_X	Gonyaulax
14	Otu013	Eukaryota	Alveolata	Dinophyta	Syndiniales	Syndiniales_Group_III	Syndiniales_Group_III_X	Syndiniales_Group_III_X
15	Otu014	Eukaryota	Stramenopiles	Ochrophyta	Chrysophyceae	Chrysophyceae_X	Chrysophyceae_Clade_G	Chrysophyceae_Clade_G_X
16	Otu015	Eukaryota	Alveolata	Dinophyta	Dinophyceae	Dinophyceae_X	Dinophyceae_X	Dinophyceae_X
17	Otu016	Eukaryota	Hacrobia	Centroheliozoa	Centroheliozoa_X	Pterocystida	Pterocystida_X	Pterocystida_X
18	Otu017	Eukaryota	Opisthokonta	Fungi	Basidiomycota	Agaricomycotina	Agaricomycetes	Hyphodontia
19	Otu018	Eukaryota	Stramenopiles	Ochrophyta	Dictyochophyceae	Dictyochophyceae_X	Pedinellales	Pedinellales_X
20	Otu019	Eukaryota	Opisthokonta	Fungi	Basidiomycota	Agaricomycotina	Agaricomycetes	Itersonilia
21	Otu020	Eukaryota	Hacrobia	Haptophyta	Prymnesiophyceae	Prymnesiophyceae_X	Braarudosphaeraceae	Braarudosphaera
22	Otu021	Eukaryota	Alveolata	Dinophyta	Dinophyceae	Dinophyceae_X	Dinophyceae_X	Dinophyceae_X
23	Otu022	Eukaryota	Hacrobia	Haptophyta	Prymnesiophyceae	Prymnesiophyceae_X	Prymnesiophyceae_X	Syracosphaera
24	Otu023	Eukaryota	Stramenopiles	Ochrophyta	Bacillariophyta	Bacillariophyceae	Bacillariophyceae_X	Bacillariophyceae_X
25	Otu024	Eukaryota	Archaeplastida	Streptophyta	Klebsormidiophyceae	Klebsormidiophyceae_X	Klebsormidiophyceae_XX	Klebsormidium
26	Otu025	Eukaryota	Archaeplastida	Chlorophyta	Mamiellophyceae	Mamiellales	Mamiellaceae	Micromonas
27	Otu026	Eukaryota	Stramenopiles	Ochrophyta	Bacillariophyta	Bacillariophyceae	Bacillariophyceae_X	Cylindrotheca
28	Otu027	Eukaryota	Alveolata	Dinophyta	Dinophyceae	Suessiales	Suessiales_X	Karlodinium
29	Otu028	Eukaryota	Hacrobia	Haptophyta	Prymnesiophyceae	Isochrysidales	Noelaerhabdaceae	Emiliania
30	Otu029	Eukaryota	Opisthokonta	Fungi	Ascomycota	Saccharomycotina	Saccharomycetales	Debaryomyces
31	Otu030	Eukaryota	Hacrobia	Cryptophyta	Cryptophyceae	Cryptophyceae_X	Cryptomonadales	Teleaulax
32	Otu031	Eukaryota	Alveolata	Dinophyta	Syndiniales	Syndiniales_Group_I	Syndiniales_Group_I_Clade_1	Syndiniales_Group_I_Clade_1_X
33	Otu032	Eukaryota	Archaeplastida	Chlorophyta	Prasino_Clade_VII	Prasino_Clade_VII_X	Prasino_Clade_VII_A	Prasino_Clade_VII_A_4_X

Data tables - Abundance



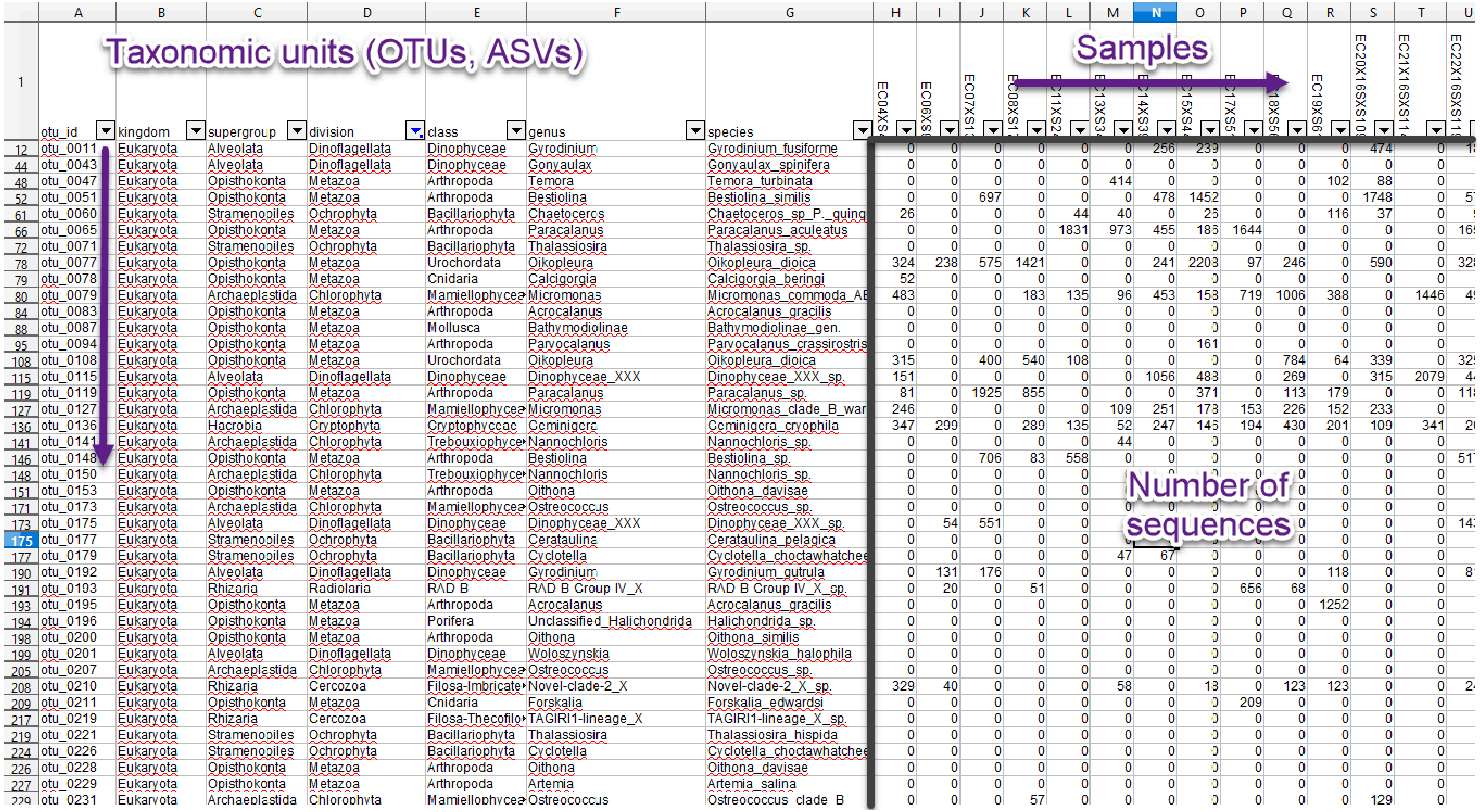
	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U
1	otu	X10n	X10p	X11n	X11p	X120n	X120p	X121n	X121p	X122n	X122p	X125n	X125p	X126n	X126p	X127n	X13n	X13p	X140n	X140p	X141n
2	Otu001	13679	6292	42	2500	18850	5	43	7138	9432	10541	9	9772	1388	7	31538	38	2338	23	9	1358
3	Otu002	18	7134	38	9830	45	61420	182	23751	36	11	4535	3502	11018	5473	26	14411	38	19018	12	3080
4	Otu003	9939	8983	31	13	24620	19	19	16	12502	3831	4621	2240	9924	4052	9292	18	0	37	7	3680
5	Otu004	3675	4234	24	22	11	16	32967	35	6	18	6908	5	16	8702	24	11	37717	0	25	4196
6	Otu005	0	5	0	7	0	8	0	16	20166	0	0	2	5	8	2	16	0	13	0	0
7	Otu006	0	8	0	0	0	8	0	0	5	3	3	0	0	9	0	5	4	0	0	3
8	Otu007	4587	518	4	386	8775	5	6	1102	14336	0	0	3626	51	0	6	12	0	10	0	395
9	Otu008	1	8	2	4408	3	29	6	12355	0	0	0	0	0	9	3	1588	0	6	3	3
10	Otu009	115	914	3	325	0	629	1	834	5	0	1354	2108	1117	67	0	2010	1897	11227	1	3
11	Otu010	780	8	23810	12	3279	0	12	7	3027	0	2	4156	0	0	18	0	0	0	0	0
12	Otu011	0	3	2	2	0	13	5	5	4	7	3081	11	4	6804	0	3	11	0	5	0
13	Otu012	0	0	0	6	0	0	0	16	3	0	0	0	0	0	0	17	0	6	0	0
14	Otu013	6321	2471	2	0	12	3	0	0	4	20272	0	15	9	0	5	0	11	0	14	0
15	Otu014	0	82	4	3304	1	1667	4	9233	13	3	0	2707	0	0	3	4806	9	3	5	0
16	Otu015	0	12	0	3	7	25	1	6	10	0	4	2772	1	3	0	2	0	10	13	8052
17	Otu016	1	0	0	9	5	0	0	14	0	0	0	0	2654	0	0	6	1	1	0	0
18	Otu017	0	0	0	0	0	0	0	0	17	8	0	0	0	0	0	17	24	48	35210	4
19	Otu018	1	0	9	911	0	0	15	2702	6	4	342	2217	606	0	13	3846	4	6	8513	1
20	Otu019	0	0	13	0	0	0	29	0	0	0	0	0	0	0	11	0	0	5	4	0
21	Otu020	425	0	1	0	1706	0	8447	1	0	0	0	0	0	26	0	0	3490	0	2620	0
22	Otu021	0	4	0	0	0	10	0	0	0	0	2	0	0	4	0	0	0	0	0	4
23	Otu022	0	0	0	4987	0	0	0	6	90	1	1	524	0	467	0	4	8	6198	0	1
24	Otu023	4	0	1	0	0	3	0	0	0	0	0	0	3351	3	0	3910	1	2	3	0
25	Otu024	0	0	0	0	0	0	0	0	0	0	0	17	0	0	0	0	0	2	1	0
26	Otu025	69	0	0	0	290	0	0	0	21	0	118	2	9	513	2	0	0	2	0	0
27	Otu026	0	2	0	0	0	3	0	0	0	1	0	0	0	0	0	0	0	0	0	0
28	Otu027	6	2304	0	0	5	0	0	0	57	4	0	14529	9597	2	6	0	0	0	0	0

Data tables - Metadata



	A	B	C	D	E	F	G	H	I	J	K	L
1	sample	fraction	Select_18S_nifH	total_18S	total_16S	total_nifH	sample_number	transect	station	depth	latitude	longitude
2	X10n	Nano	Yes	53230	8772	36	10	1	81	140	-27.42	-44.72
3	X10p	Pico	Yes	47390	4448	6241	10	1	81	140	-27.42	-44.72
4	X11n	Nano	No	24007	6193	3772	11	1	85	110	-26.8	-45.3
5	X11p	Pico	Yes	31899	14	10201	11	1	85	110	-26.8	-45.3
6	X120n	Nano	Yes	70455	5292	93	120	2	96	5	-27.39	-47.82
7	X120p	Pico	Yes	76182	53272	23147	120	2	96	5	-27.39	-47.82
8	X121n	Nano	Yes	52401	5958	26838	121	2	96	30	-27.39	-47.82
9	X121p	Pico	Yes	71785	10993	23706	121	2	96	30	-27.39	-47.82
10	X122n	Nano	Yes	78740	11730	15543	122	2	96	50	-27.39	-47.82
11	X122p	Pico	Yes	37364	11817	11045	122	2	96	50	-27.39	-47.82
12	X125n	Nano	Yes	27381	9	14331	125	2	98	5	-27.59	-47.39
13	X125p	Pico	Yes	55179	10419	21461	125	2	98	5	-27.59	-47.39
14	X126n	Nano	Yes	65714	15	16929	126	2	98	50	-27.59	-47.39
15	X126p	Pico	Yes	30406	3	10140	126	2	98	50	-27.59	-47.39
16	X127n	Nano	Yes	60610	9	11493	127	2	98	85	-27.59	-47.39
17	X13n	Nano	Yes	46001	33	21316	13	1	86	105	-26.33	-45.41
18	X13p	Pico	Yes	59626	7217	11954	13	1	86	105	-26.33	-45.41
19	X140n	Nano	Yes	48126	10428	25286	140	2	101	5	-27.79	-46.96
20	X140p	Pico	Yes	46569	10448	12301	140	2	101	5	-27.79	-46.96
21	X141n	Nano	Yes	30081	6394	21302	141	2	101	60	-27.79	-46.96
22	X141p	Pico	Yes	64221	11318	10428	141	2	101	60	-27.79	-46.96
23	X142n	Nano	Yes	85219	23243	11753	142	2	101	110	-27.79	-46.96
24	X142p	Pico	Yes	89797	9553	17156	142	2	101	110	-27.79	-46.96
25	X155n	Nano	Yes	54162	8237	20674	155	2	106	5	-28.12	-46.17
26	X155p	Pico	Yes	50782	7384	66172	155	2	106	5	-28.12	-46.17
27	X156n	Nano	Yes	55065	11371	14447	156	2	106	60	-28.12	-46.17
28	X156p	Pico	Yes	43917	9665	16093	156	2	106	60	-28.12	-46.17
29	X157n	Nano	Yes	29078	4978	15532	157	2	106	100	-28.12	-46.17
30	X157p	Pico	Yes	51848	9139	15204	157	2	106	100	-28.12	-46.17
31	X15n	Nano	Yes	22468	2887	2678	15	1	87	105	-26.22	-45.48
32	X15p	Pico	Yes	78390	13813	1033	15	1	87	105	-26.22	-45.48
33	X165n	Nano	Yes	50732	15337	14706	165	2	114	5	-28.65	-44.99
34	X165p	Pico	Yes	48514	10902	39918	165	2	114	5	-28.65	-44.99
35	X166n	Nano	Yes	53412	3411	24442	166	2	114	60	-28.65	-44.99

Data tables - Merged



	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U
1	otu_id	kingdom	supergroup	division	class	genus	species	EC04XS	EC06XS	EC07XS1	EC08XS1	EC11XS2	EC13XS3	EC14XS3	EC15XS4	EC17XS5	18XS5	EC19XS6	EC20X16SXS10	EC21X16SXS11	EC22X16SXS11
12	otu_0011	Eukaryota	Alveolata	Dinoflagellata	Dinophyceae	Gyrodinium	Gyrodinium_fusiforme	0	0	0	0	0	0	256	239	0	0	0	474	0	1
44	otu_0043	Eukaryota	Alveolata	Dinoflagellata	Dinophyceae	Gonyaulax	Gonyaulax_spinifera	0	0	0	0	0	0	0	0	0	0	0	0	0	
48	otu_0047	Eukaryota	Opisthokonta	Metazoa	Arthropoda	Temora	Temora_turbinata	0	0	0	0	0	414	0	0	0	0	102	88	0	
52	otu_0051	Eukaryota	Opisthokonta	Metazoa	Arthropoda	Bestiolina	Bestiolina_similis	0	0	697	0	0	0	478	1452	0	0	0	1748	0	5
61	otu_0060	Eukaryota	Stramenopiles	Ochrophyta	Bacillariophyta	Chaetoceros	Chaetoceros_sp_P._quinq	26	0	0	0	44	40	0	26	0	0	116	37	0	
66	otu_0065	Eukaryota	Opisthokonta	Metazoa	Arthropoda	Paracalanus	Paracalanus_aculeatus	0	0	0	0	1831	973	455	186	1644	0	0	0	0	16
72	otu_0071	Eukaryota	Stramenopiles	Ochrophyta	Bacillariophyta	Thalassiosira	Thalassiosira_sp.	0	0	0	0	0	0	0	0	0	0	0	0	0	
78	otu_0077	Eukaryota	Opisthokonta	Metazoa	Urochordata	Oikopleura	Oikopleura_dioica	324	238	575	1421	0	0	241	2208	97	246	0	590	0	32
79	otu_0078	Eukaryota	Opisthokonta	Metazoa	Cnidaria	Calcigorgia	Calcigorgia_beringi	52	0	0	0	0	0	0	0	0	0	0	0	0	
80	otu_0079	Eukaryota	Archaeplastida	Chlorophyta	Mamiellophycea	Micromonas	Micromonas_commoda_A	483	0	0	183	135	96	453	158	719	1006	388	0	1446	4
84	otu_0083	Eukaryota	Opisthokonta	Metazoa	Arthropoda	Acrocalanus	Acrocalanus_gracilis	0	0	0	0	0	0	0	0	0	0	0	0	0	
88	otu_0087	Eukaryota	Opisthokonta	Metazoa	Mollusca	Bathymodiolinae	Bathymodiolinae_gen.	0	0	0	0	0	0	0	0	0	0	0	0	0	
95	otu_0094	Eukaryota	Opisthokonta	Metazoa	Arthropoda	Parvocalanus	Parvocalanus_crassirostris	0	0	0	0	0	0	0	161	0	0	0	0	0	
108	otu_0108	Eukaryota	Opisthokonta	Metazoa	Urochordata	Oikopleura	Oikopleura_dioica	315	0	400	540	108	0	0	0	0	784	64	339	0	32
115	otu_0115	Eukaryota	Alveolata	Dinoflagellata	Dinophyceae	Dinophyceae_XXX	Dinophyceae_XXX_sp.	151	0	0	0	0	0	1056	488	0	269	0	315	2079	4
119	otu_0119	Eukaryota	Opisthokonta	Metazoa	Arthropoda	Paracalanus	Paracalanus_sp.	81	0	1925	855	0	0	0	371	0	113	179	0	0	11
127	otu_0127	Eukaryota	Archaeplastida	Chlorophyta	Mamiellophycea	Micromonas	Micromonas_clade_B_war	246	0	0	0	0	109	251	178	153	226	152	233	0	
136	otu_0136	Eukaryota	Hacrobia	Cryptophyta	Cryptophyceae	Geminigera	Geminigera_cryophila	347	299	0	289	135	52	247	146	194	430	201	109	341	2
141	otu_0141	Eukaryota	Archaeplastida	Chlorophyta	Trebouxiophyce	Nannochloris	Nannochloris_sp.	0	0	0	0	0	44	0	0	0	0	0	0	0	
146	otu_0148	Eukaryota	Opisthokonta	Metazoa	Arthropoda	Bestiolina	Bestiolina_sp.	0	0	706	83	558	0	0	0	0	0	0	0	0	51
148	otu_0150	Eukaryota	Archaeplastida	Chlorophyta	Trebouxiophyce	Nannochloris	Nannochloris_sp.	0	0	0	0	0	[illegible]	0	0	0	0	0	0	0	
151	otu_0153	Eukaryota	Opisthokonta	Metazoa	Arthropoda	Oithona	Oithona_davisae	0	0	0	0	0	[illegible]	[illegible]	[illegible]	[illegible]	0	0	0	0	
171	otu_0173	Eukaryota	Archaeplastida	Chlorophyta	Mamiellophycea	Ostreococcus	Ostreococcus_sp.	0	0	0	0	0	0	0	0	0	0	0	0	0	
173	otu_0175	Eukaryota	Alveolata	Dinoflagellata	Dinophyceae	Dinophyceae_XXX	Dinophyceae_XXX_sp.	0	54	551	0	0	[illegible]	[illegible]	[illegible]	[illegible]	0	0	0	0	14
175	otu_0177	Eukaryota	Stramenopiles	Ochrophyta	Bacillariophyta	Cerataulina	Cerataulina_pelagica	0	0	0	0	0	0	[illegible]	0	0	0	0	0	0	
177	otu_0179	Eukaryota	Stramenopiles	Ochrophyta	Bacillariophyta	Cyclotella	Cyclotella_choctawhatchee	0	0	0	0	0	47	67	0	0	0	0	0	0	
190	otu_0192	Eukaryota	Alveolata	Dinoflagellata	Dinophyceae	Gyrodinium	Gyrodinium_gutrula	0	131	176	0	0	0	0	0	0	0	118	0	0	8
191	otu_0193	Eukaryota	Rhizaria	Radiolaria	RAD-B	RAD-B-Group-IV_X	RAD-B-Group-IV_X_sp.	0	20	0	51	0	0	0	0	656	68	0	0	0	
193	otu_0195	Eukaryota	Opisthokonta	Metazoa	Arthropoda	Acrocalanus	Acrocalanus_gracilis	0	0	0	0	0	0	0	0	0	0	1252	0	0	
194	otu_0196	Eukaryota	Opisthokonta	Metazoa	Porifera	Unclassified_Halichondrida	Halichondrida_sp.	0	0	0	0	0	0	0	0	0	0	0	0	0	
198	otu_0200	Eukaryota	Opisthokonta	Metazoa	Arthropoda	Oithona	Oithona_similis	0	0	0	0	0	0	0	0	0	0	0	0	0	
199	otu_0201	Eukaryota	Alveolata	Dinoflagellata	Dinophyceae	Woloszynskia	Woloszynskia_halophila	0	0	0	0	0	0	0	0	0	0	0	0	0	
205	otu_0207	Eukaryota	Archaeplastida	Chlorophyta	Mamiellophycea	Ostreococcus	Ostreococcus_sp.	0	0	0	0	0	0	0	0	0	0	0	0	0	
208	otu_0210	Eukaryota	Rhizaria	Cercozoa	Filosa-Imbricate	Novel-clade-2_X	Novel-clade-2_X_sp.	329	40	0	0	0	58	0	18	0	123	123	0	0	2
209	otu_0211	Eukaryota	Opisthokonta	Metazoa	Cnidaria	Forskalia	Forskalia_edwardsi	0	0	0	0	0	0	0	0	209	0	0	0	0	
217	otu_0219	Eukaryota	Rhizaria	Cercozoa	Filosa-Thecofilo	TAGIRI1-lineage_X	TAGIRI1-lineage_X_sp.	0	0	0	0	0	0	0	0	0	0	0	0	0	
219	otu_0221	Eukaryota	Stramenopiles	Ochrophyta	Bacillariophyta	Thalassiosira	Thalassiosira_hispida	0	0	0	0	0	0	0	0	0	0	0	0	0	
224	otu_0226	Eukaryota	Stramenopiles	Ochrophyta	Bacillariophyta	Cyclotella	Cyclotella_choctawhatchee	0	0	0	0	0	0	0	0	0	0	0	0	0	
226	otu_0228	Eukaryota	Opisthokonta	Metazoa	Arthropoda	Oithona	Oithona_davisae	0	0	0	0	0	0	0	0	0	0	0	0	0	
227	otu_0229	Eukaryota	Opisthokonta	Metazoa	Arthropoda	Artemia	Artemia_salina	0	0	0	0	0	0	0	0	0	0	0	0	0	
229	otu_0231	Eukaryota	Archaeplastida	Chlorophyta	Mamiellophycea	Ostreococcus	Ostreococcus_clade_B	0	0	0	57	0	0	0	0	0	0	0	129	0	

Long barcodes

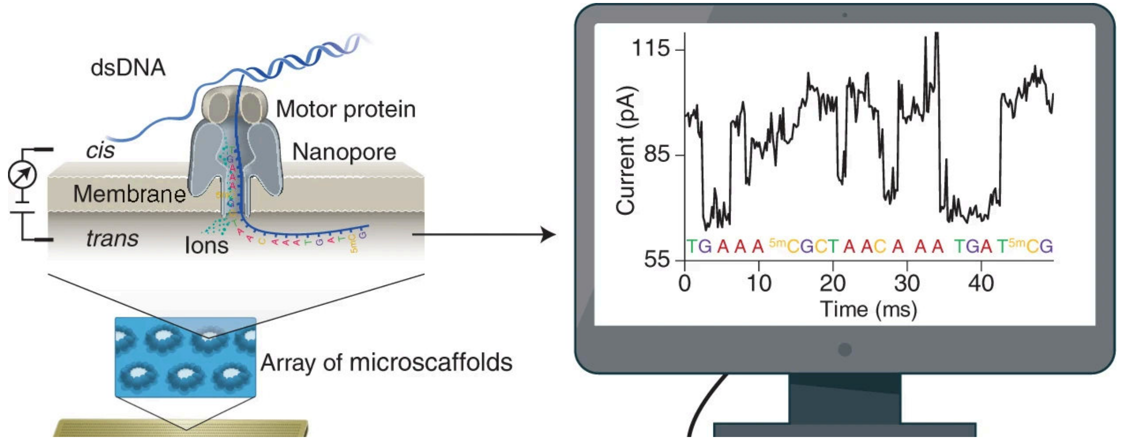
Long read sequencing

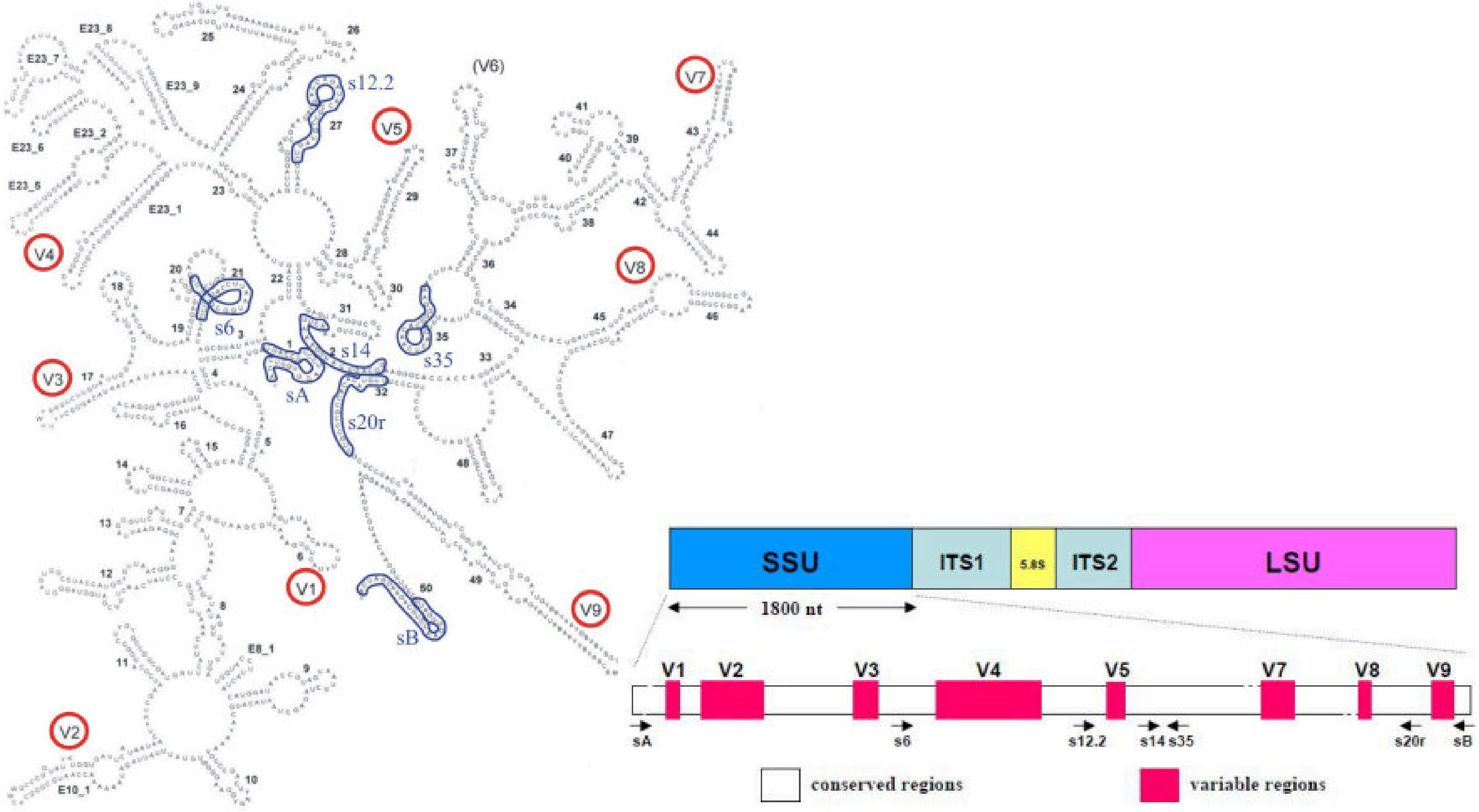
Two technologies

- Nanopore - Cheap but higher error rate
- PacBio - Expensive but better

Longer barcodes

- Full 18S
- Full operon





Processing

- Longer processing time (server)
- Full operon reference missing

Fastq files

Remove primers
cutadapt

Filter
dada2

Make ASVs
dada2

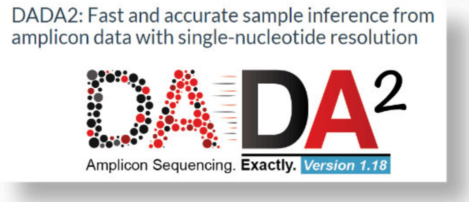
Extract 18S
barrnap

Assign taxonomy
***dada2* - PR2**

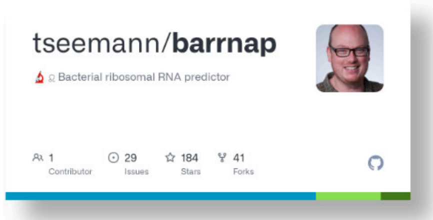
Cutadapt removes adapter sequences from high-throughput sequencing reads

Marcel Martin

Department of Computer Science, TU Dortmund, Germany



DADA2: Fast and accurate sample inference from amplicon data with single-nucleotide resolution

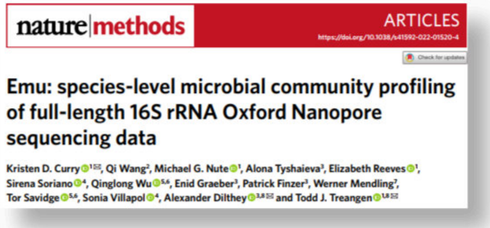


tseemann/**barrnap**

Bacterial ribosomal RNA predictor

1 Contributor
29 Issues
184 Stars
41 Forks

PR2



nature|methods ARTICLES

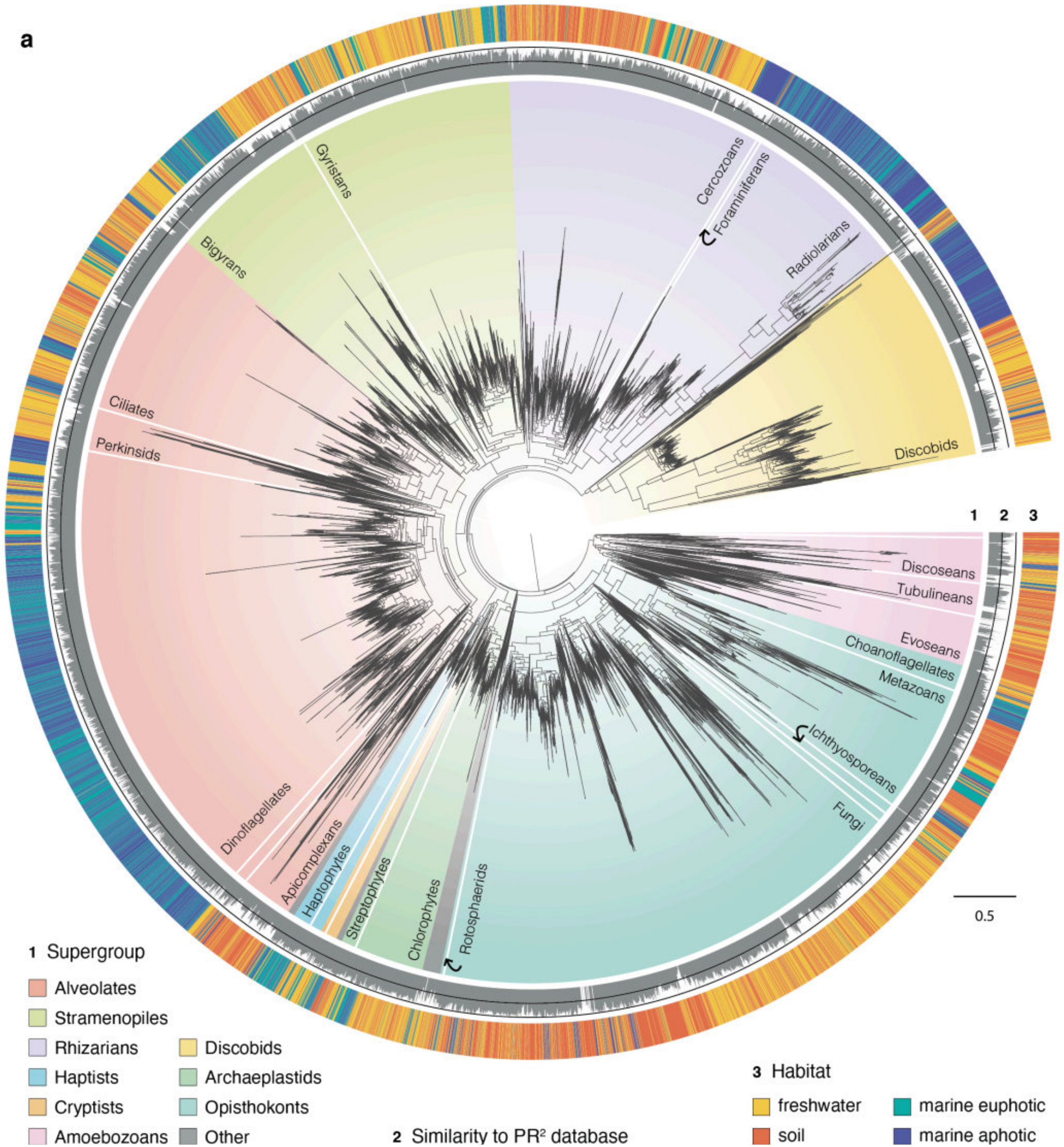
https://doi.org/10.1038/s41592-022-01520-4

Check for updates

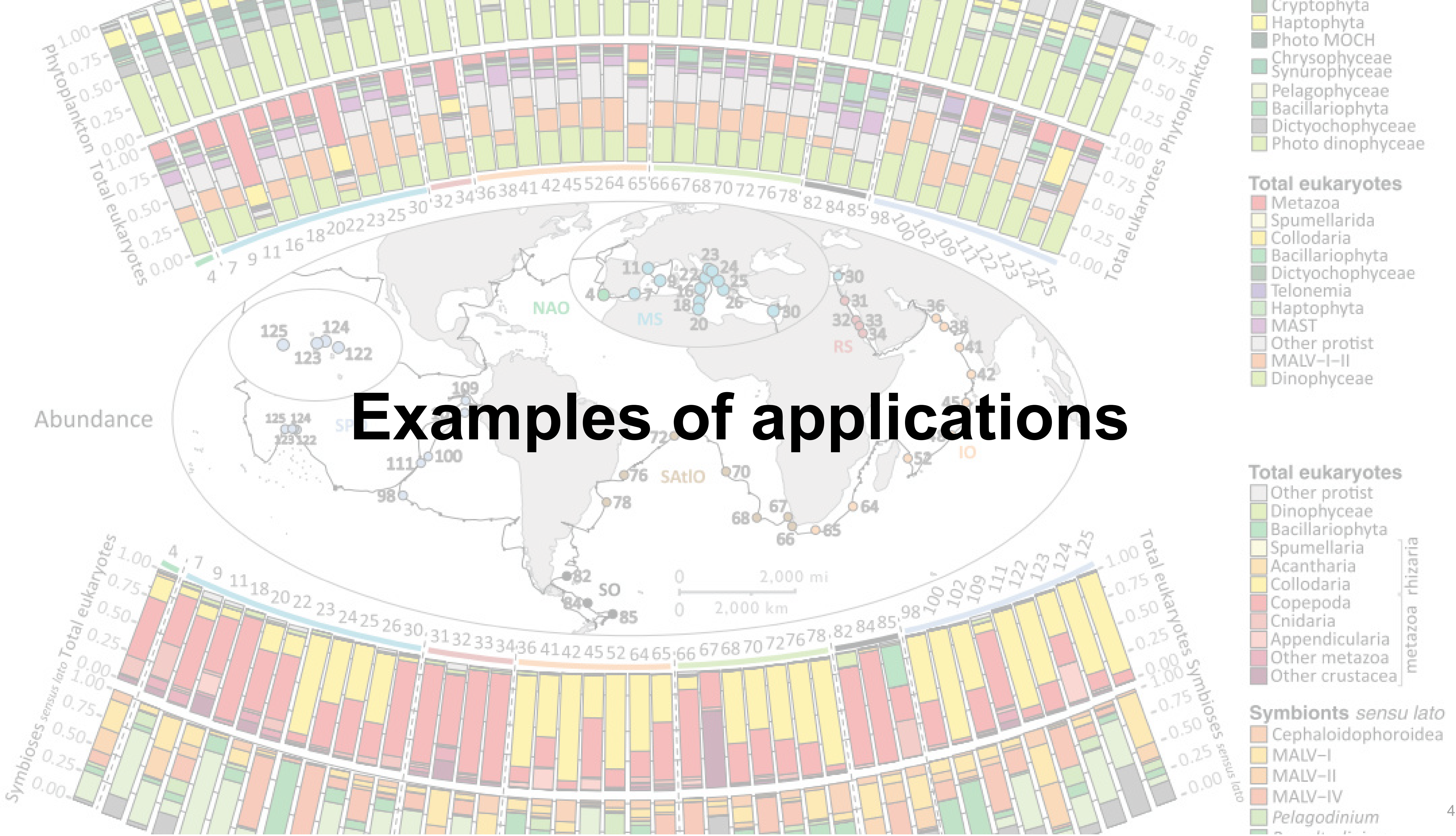
Emu: species-level microbial community profiling of full-length 16S rRNA Oxford Nanopore sequencing data

Kristen D. Curry[1✉], Qi Wang[2], Michael G. Nute[1], Alona Tyshaieva[3], Elizabeth Reeves[1], Sirena Soriano[4], Qinglong Wu[5,6], Enid Graeber[3], Patrick Finzer[3], Werner Mendling[7], Tor Savidge[5,6], Sonia Villapol[4], Alexander Dilthey[3,8✉] and Todd J. Treangen[1,8✉]

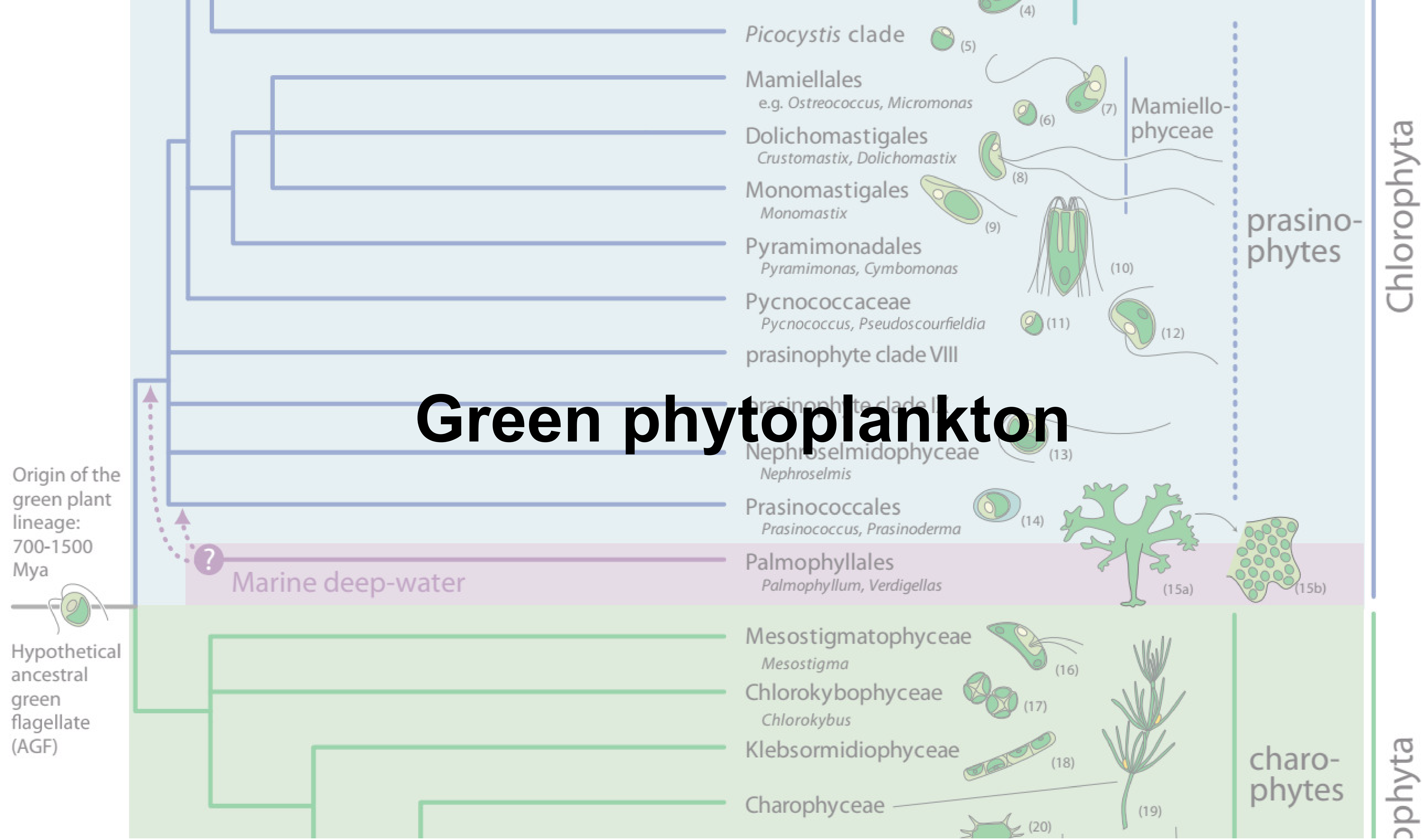
Better Phylogeny



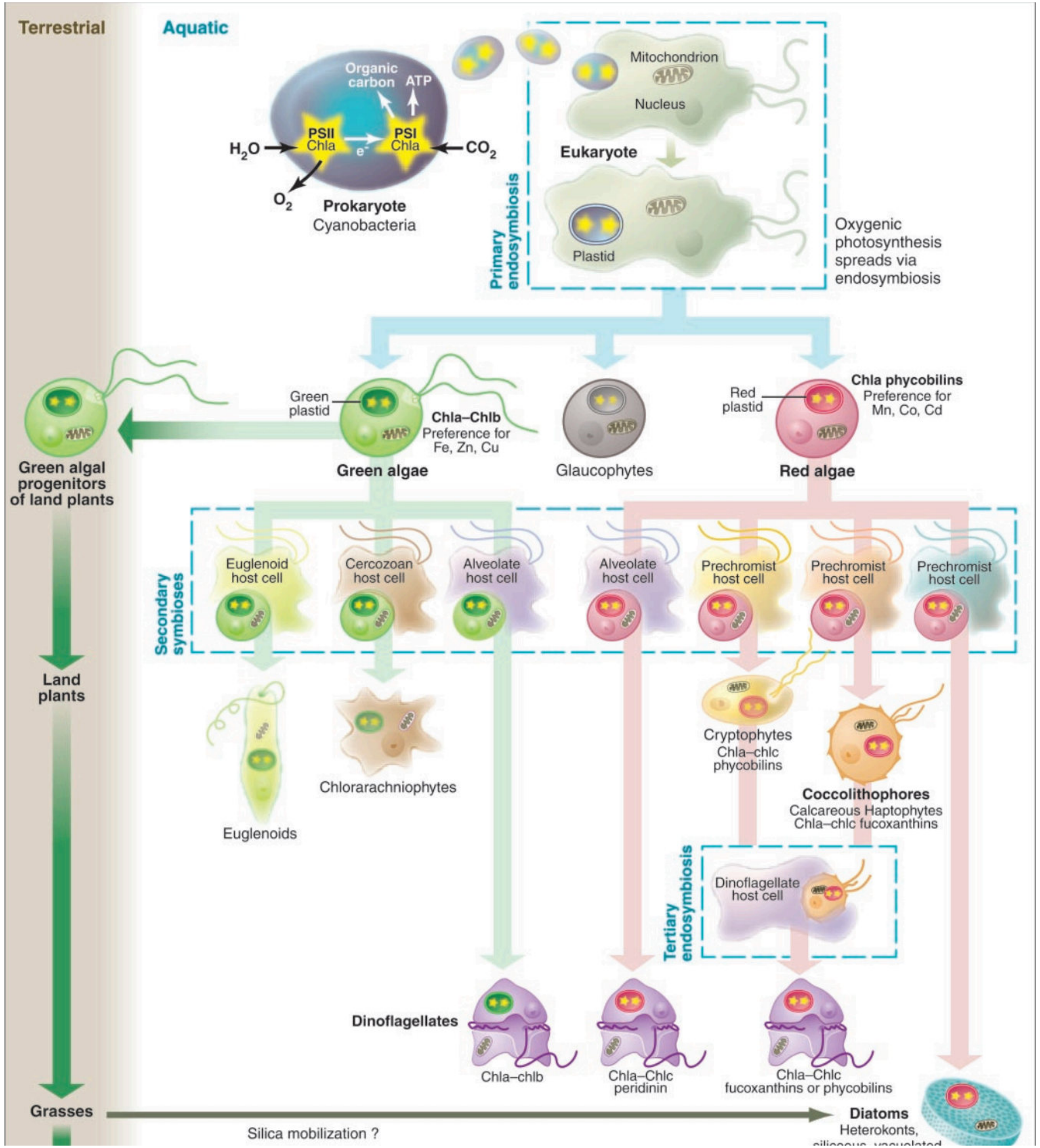
Examples of applications



Green phytoplankton



Green vs. Red lineages



What is the color of the ocean ?

- Why did the green lineage rapidly decline in ecological importance during the early Mesozoic?
- Why have terrestrial photoautotrophs not followed similar trajectories following the end-Permian extinction?

But maybe the question should be: Is the ocean really red ?

Why is the Land Green and the Ocean Red?

Paul G. FALKOWSKI[1, 2], Oscar SCHOFIELD[2], Miriam E. KATZ[1], Bas VAN DE SCHOOTBRUGGE[1, 2], and Andrew H. KNOLL[3]

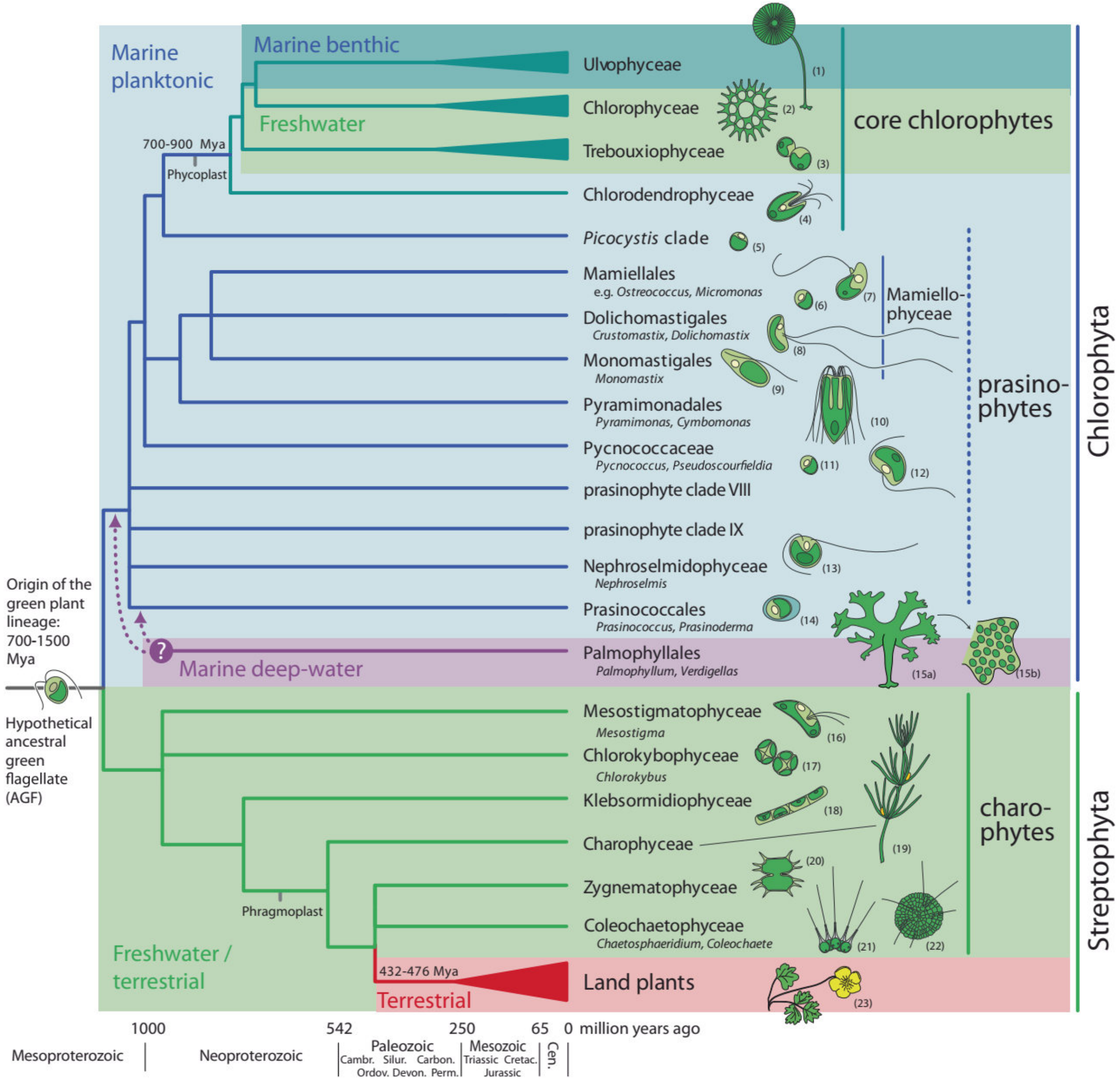
[1] Department of Geological Sciences, Rutgers University, 71 Dudley Road, New Brunswick, New Jersey 08901, USA. falko@imcs.rutgers.edu, mimikatz@rci.rutgers.edu, vandesch@imcs.rutgers.edu

[2] Environmental Biophysics and Molecular Ecology Program, Institute of Marine and Coastal Sciences, Rutgers University, 71 Dudley Road, New Brunswick, New Jersey 08901, USA. oscar@imcs.rutgers.edu

[3] Department of Organismic and Evolutionary Biology, Harvard University, 24 Oxford Street, Cambridge, Massachusetts 02138, USA. aknoll@oeb.harvard.edu

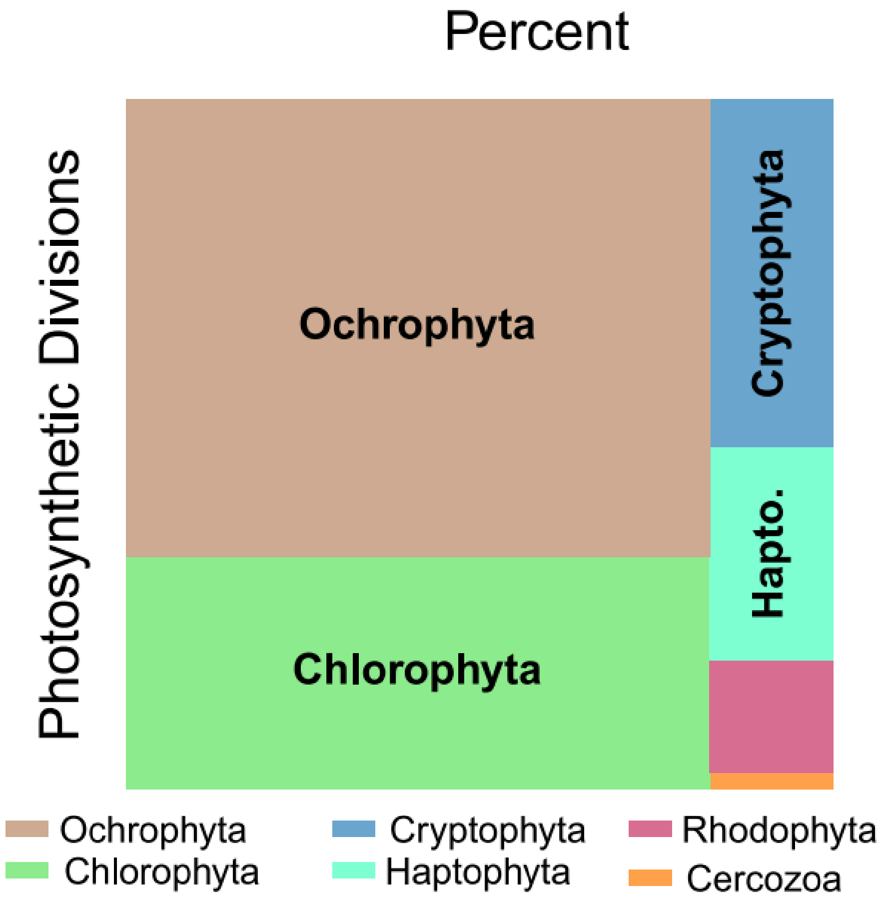
The green lineage

- Streptophyta
    - Land plants
- Chlorophyta
    - Core chlorophytes
    - “Prasinophytes”
        - **Mamiellophyceae**
        - **Chloropicophyceae**

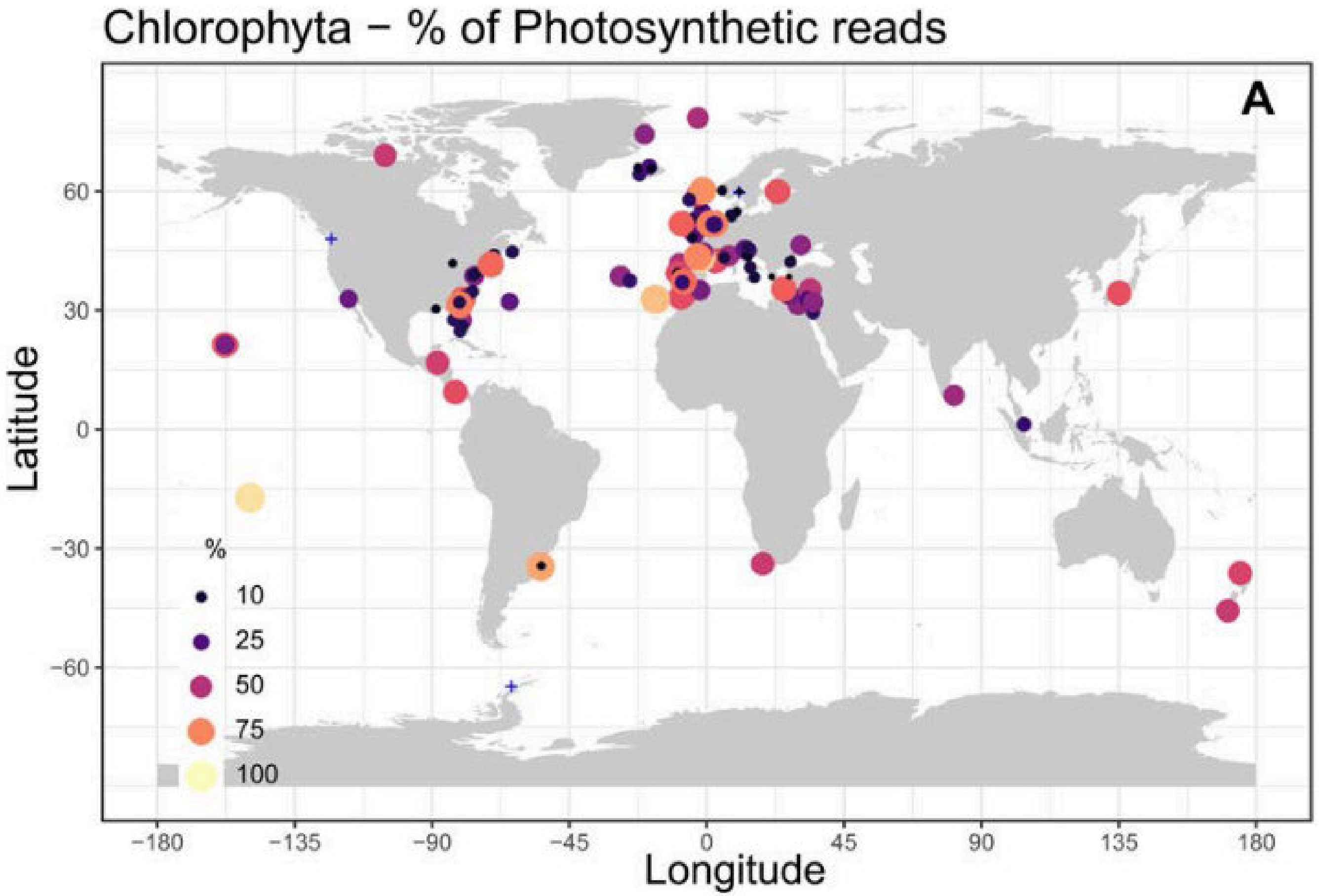


The green lineage represents 25% of coastal phytoplankton

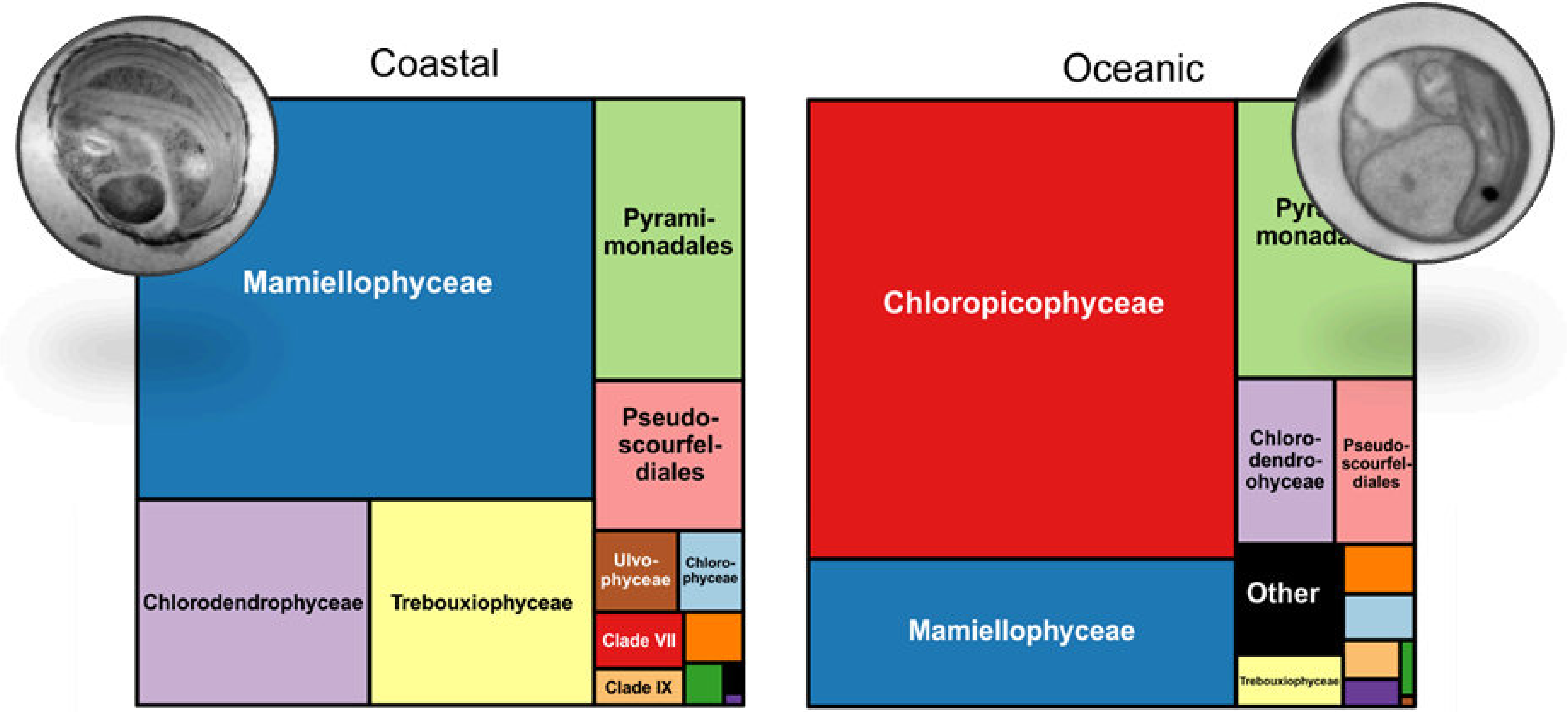
Data from OSD



But up to 94%…

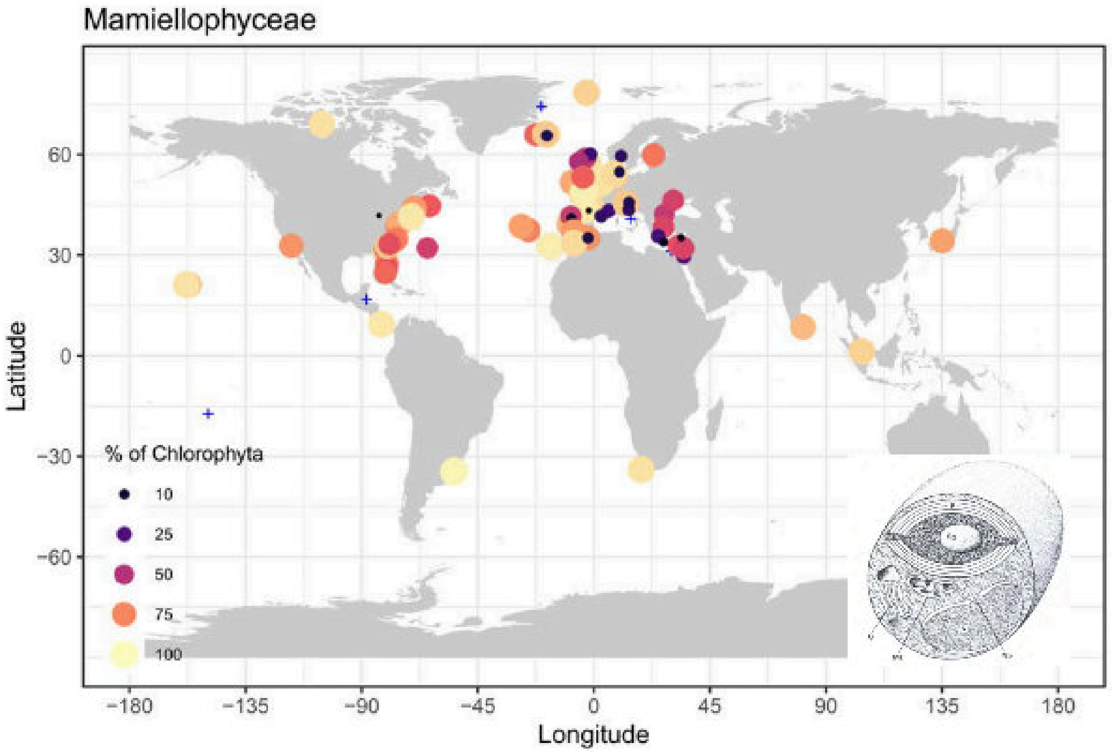


Which are the most important classes ?

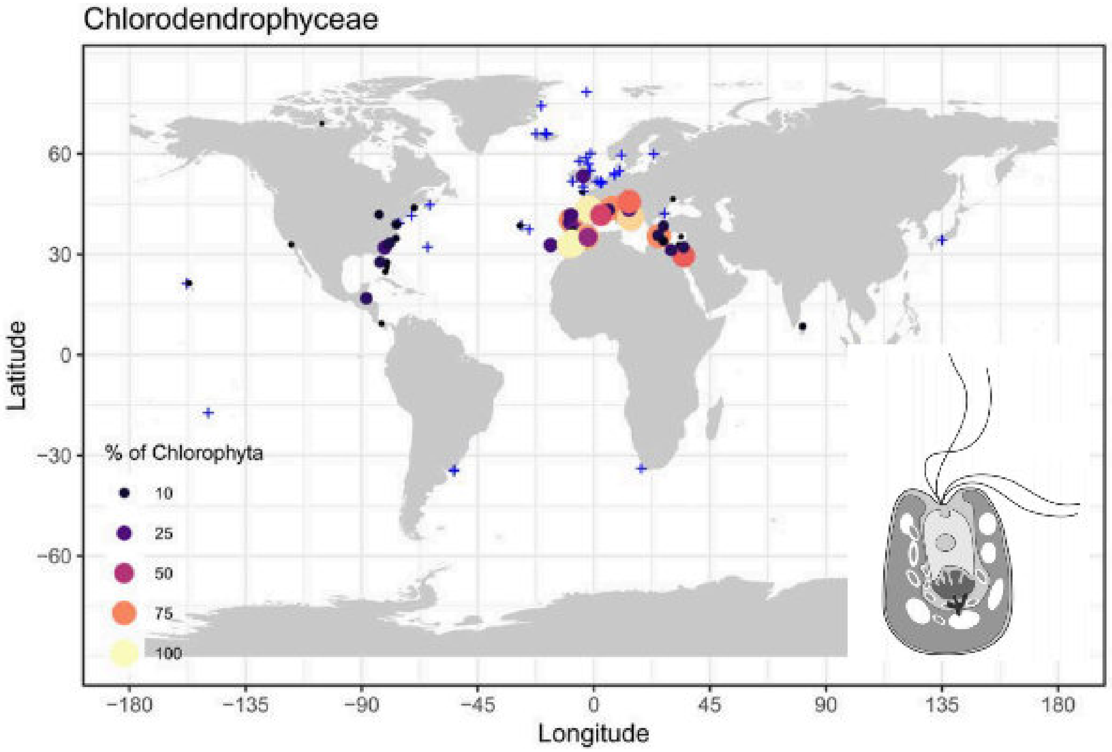


Are all classes uniformly distributed?

Coastal waters (OSD)



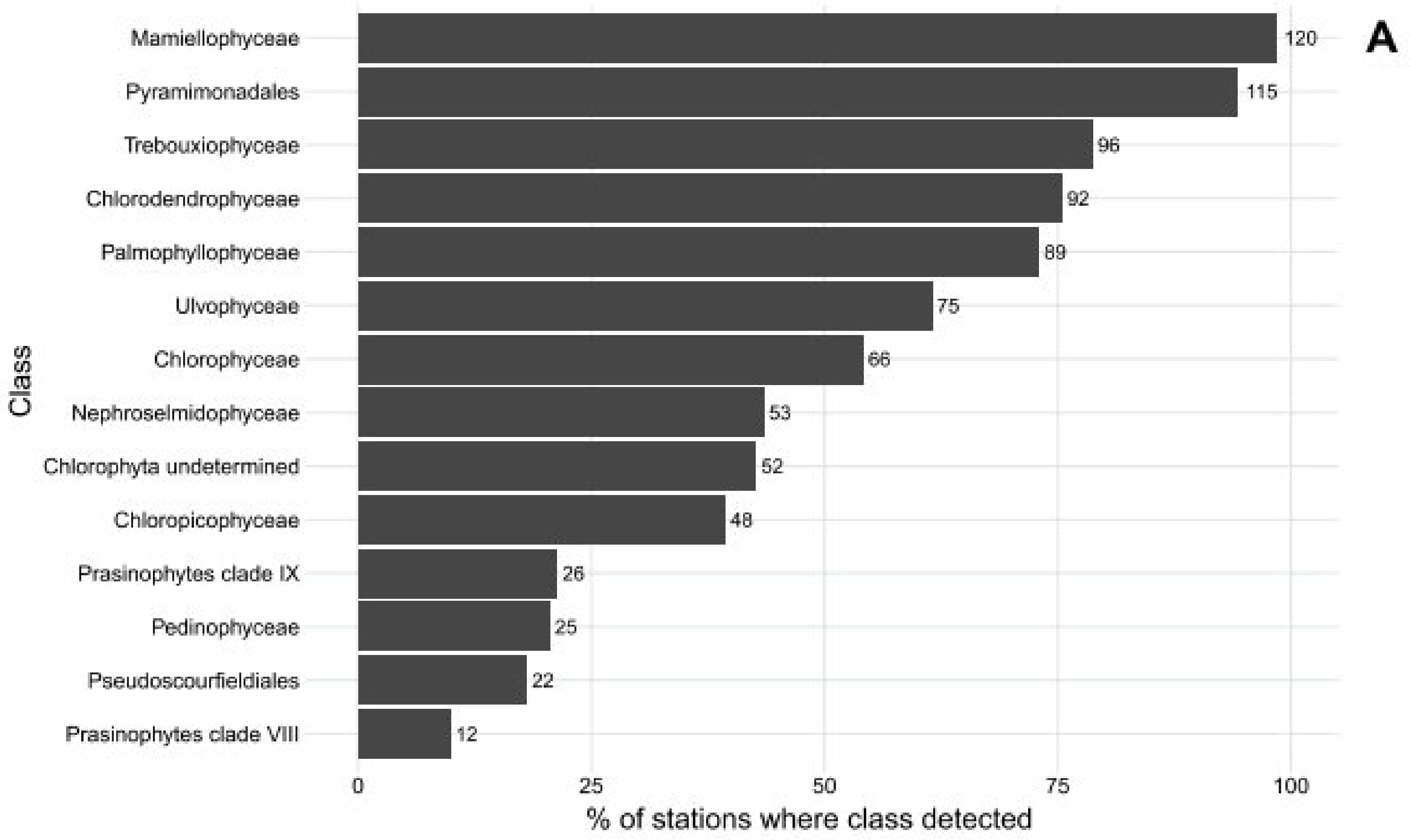
Mamiellophyceae



Chlorodendrophyceae

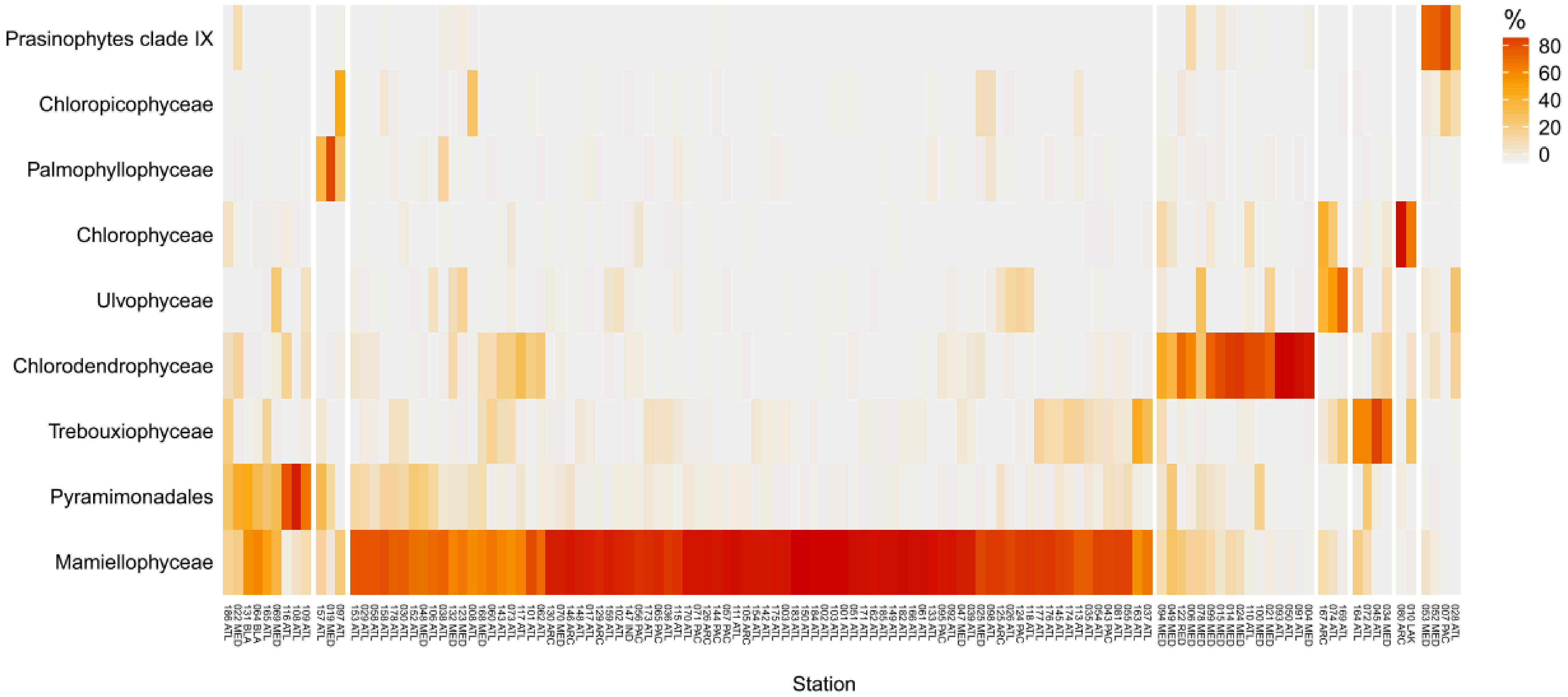
Are all classes equally prevalent?

Coastal waters (OSD)



Community structure?

Coastal waters (OSD)



Micromonas

First picoplankton species described in 1952

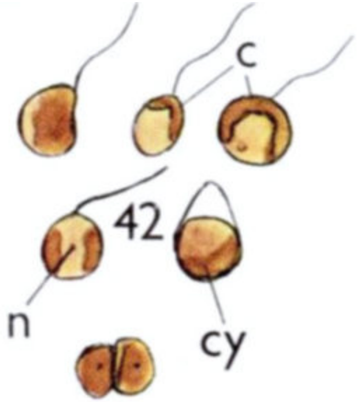
- *M. pusilla* (*Chromulina pusilla*)
- 1.5 µm
- Flagellum

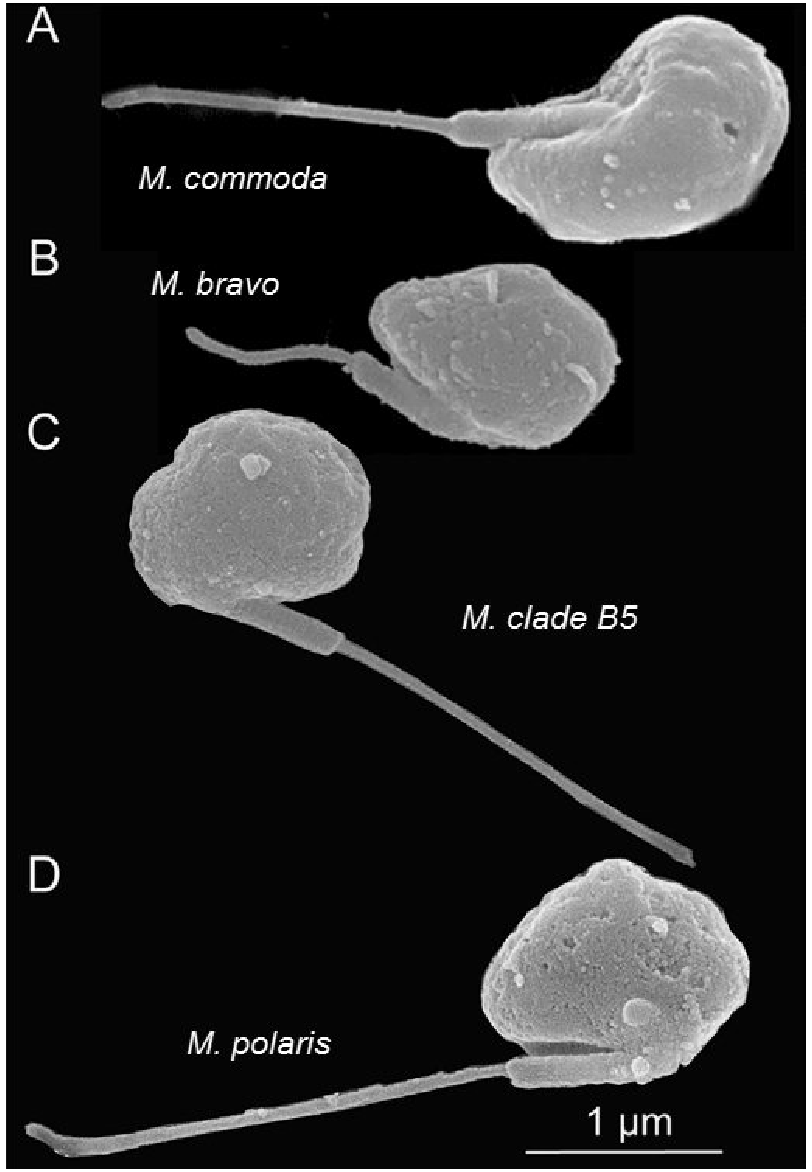
Three more species recently described

- *M. commoda*
- *M. bravo*
- *M. polaris*

Two “candidate” species

- sp. 1
- sp. 2





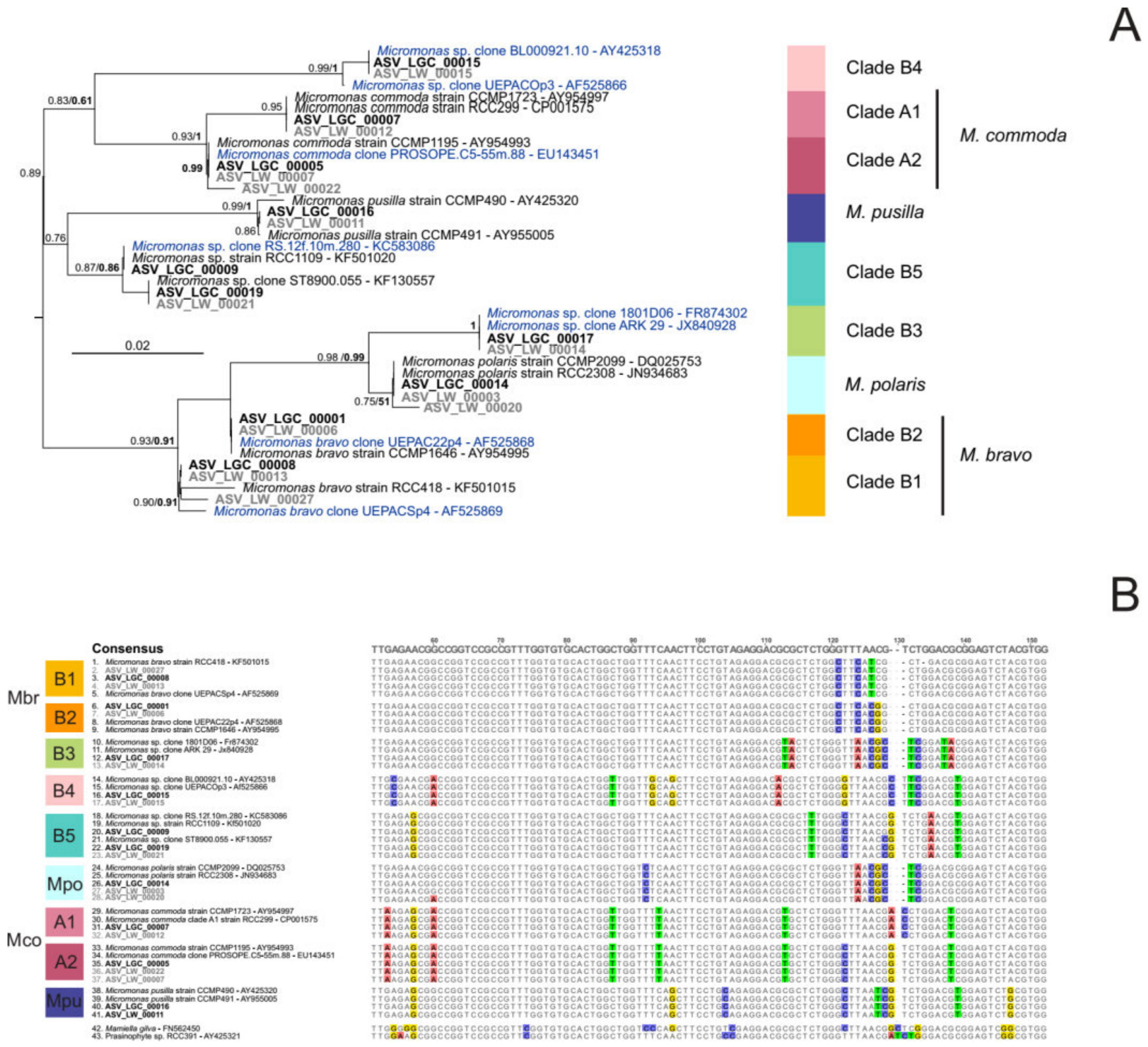
Butcher, R.W. 1952. J. Mar. Biol. Assoc. U.K. 31:175–91.

Micromonas

OSD data set

Nine clades/species

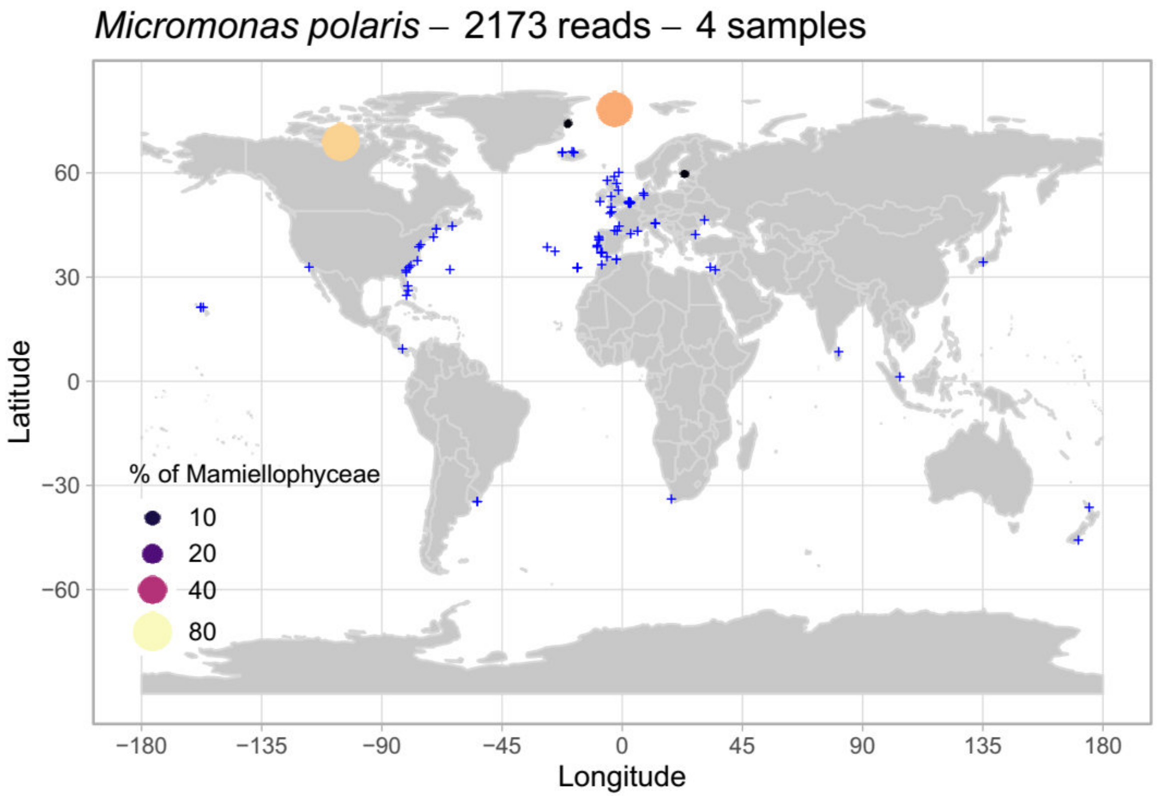
- *M. pusilla*
- *M. commoda*
  - A1-A2
- *M. bravo*
  - B1-B2
- *M. polaris*
- clades B3-B5



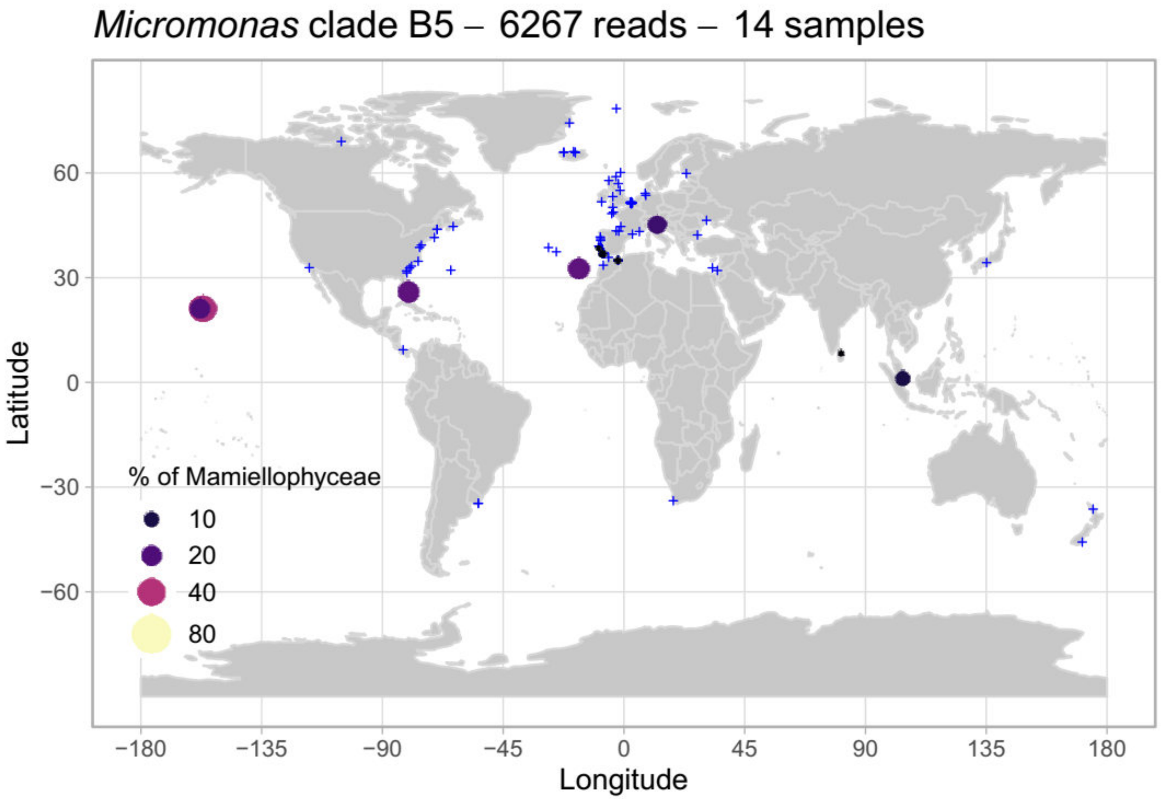
Micromonas

OSD data set

M. polaris only found in polar waters



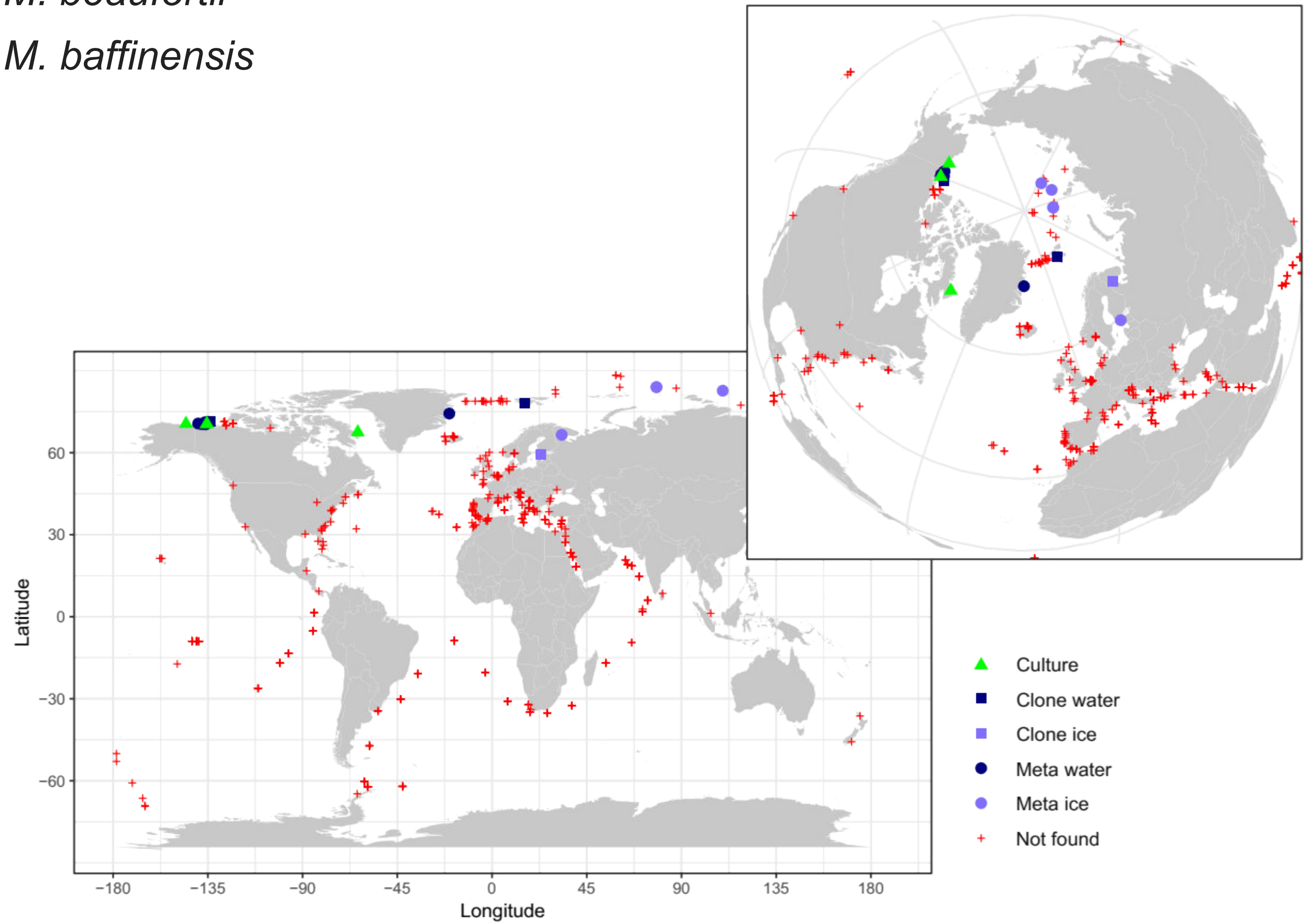
clade B5 only found in tropical waters

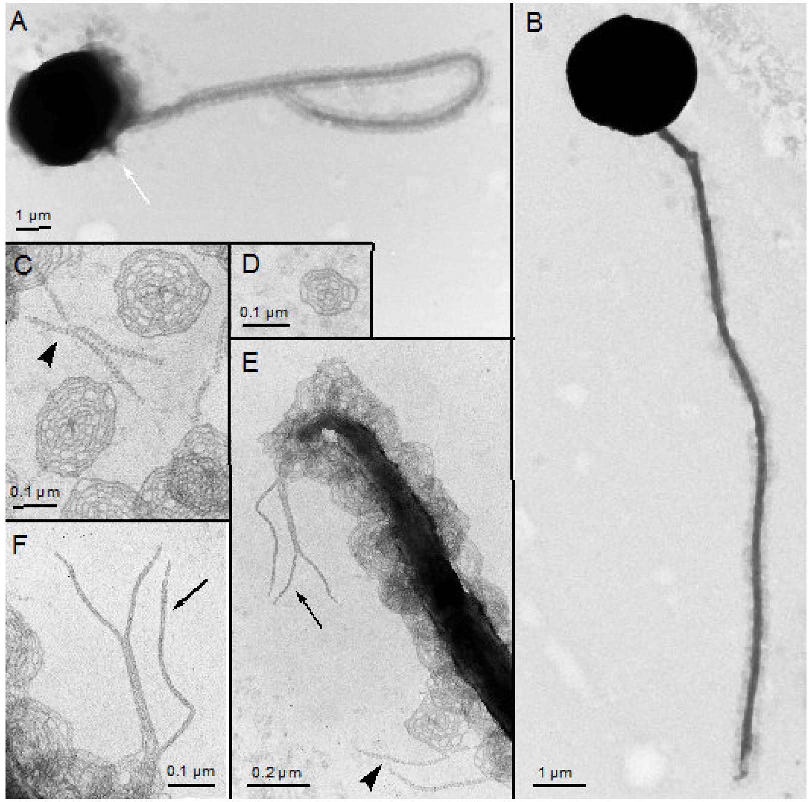


Mantoniella

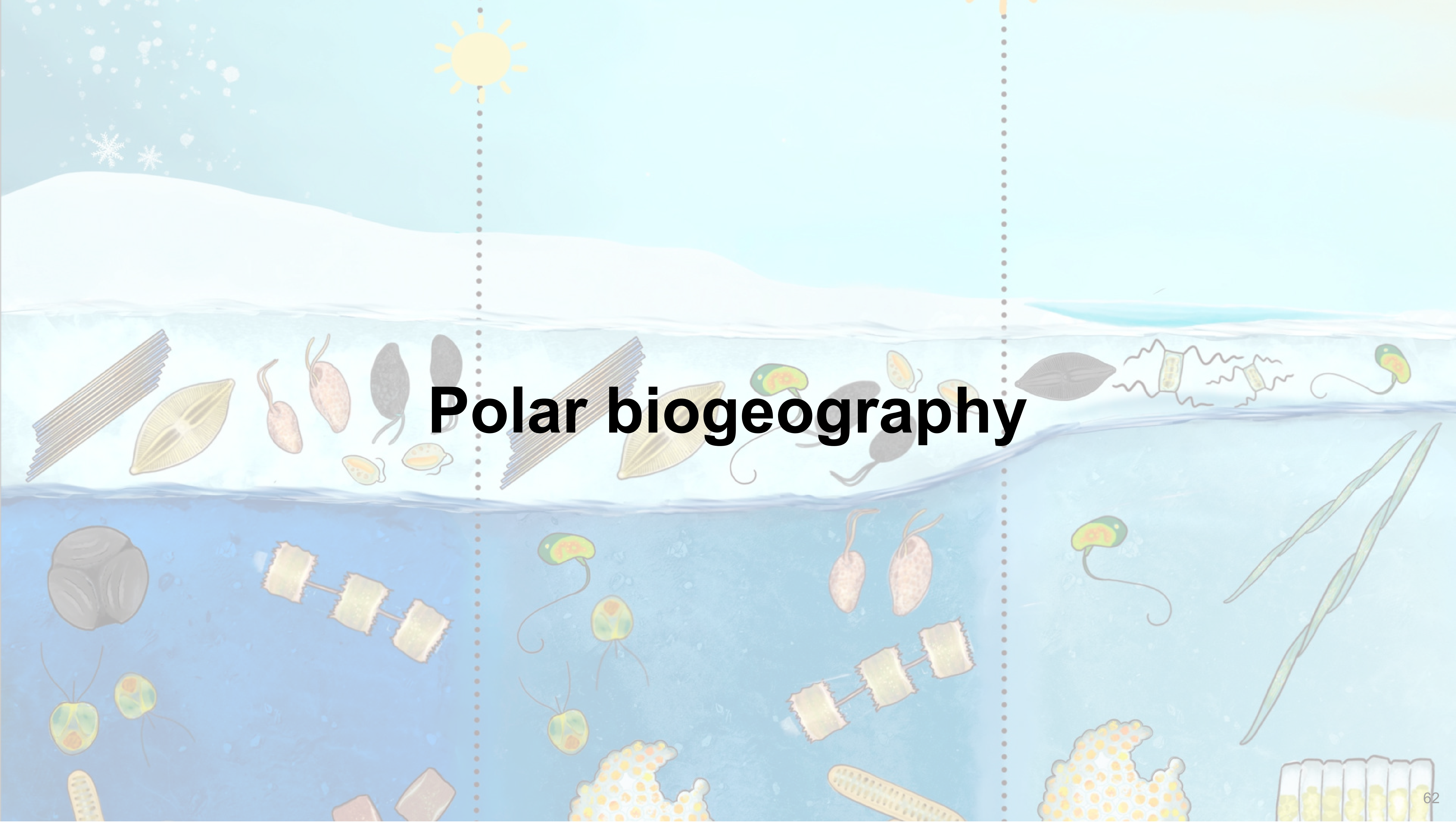
Two new polar species:

- *M. beaufortii*
- *M. baffinensis*





Yau et al.., 2020. *Mantoniella beaufortii* and *Mantoniella baffinensis* sp. nov. (Mamiellales, Mamiellophyceae), two new green algal species from



Polar biogeography

Svalbard diatoms

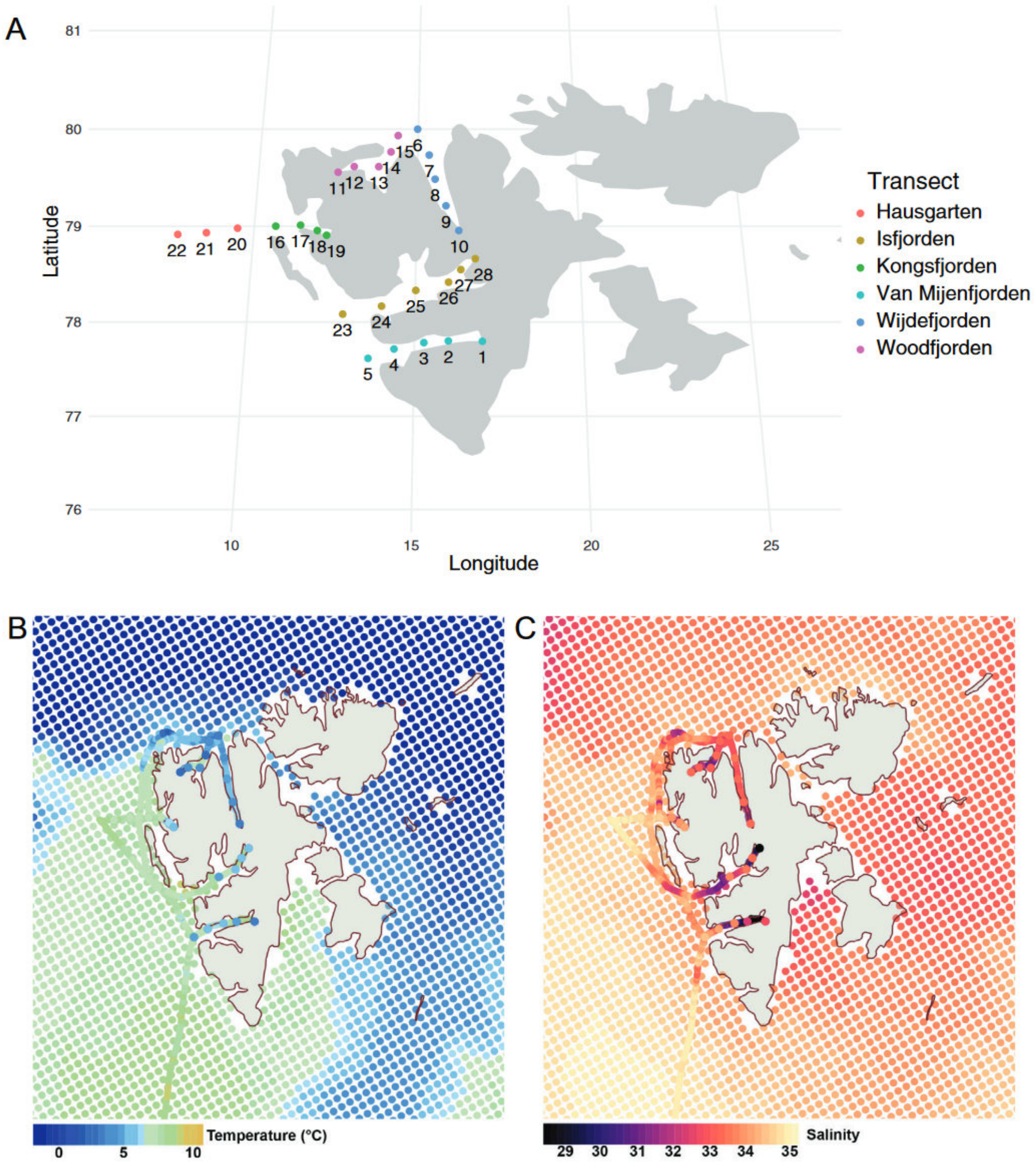
ELEMENTA
Science of the Anthropocene

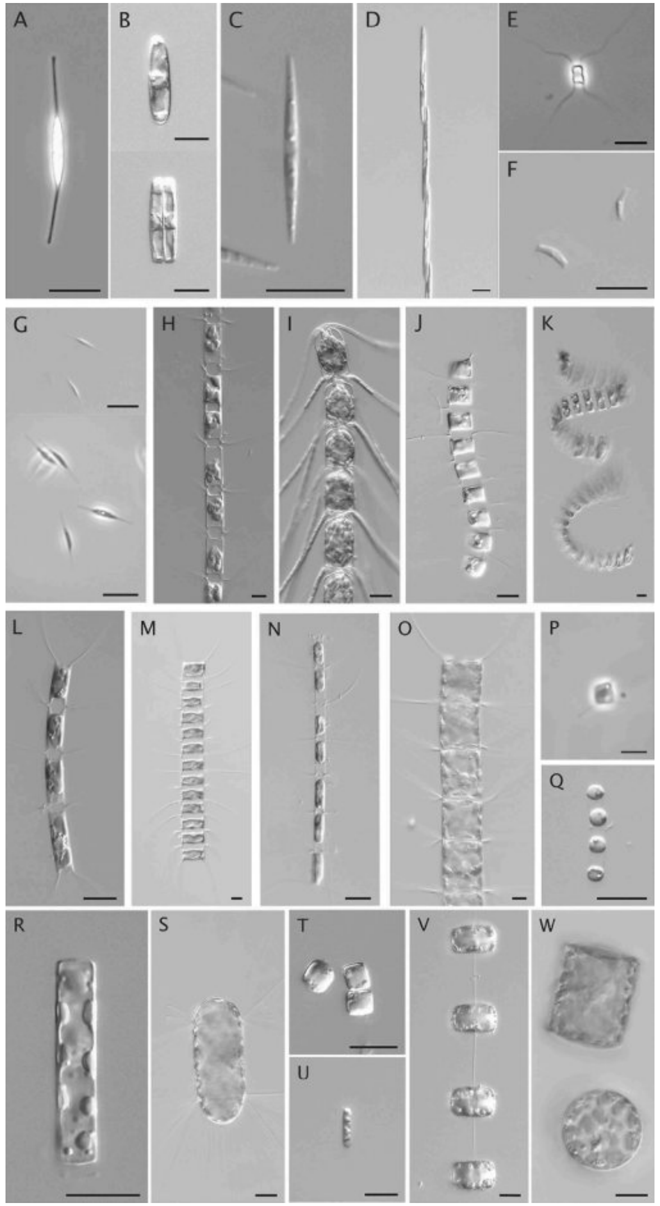
Šupraha, L, et al. 2022. Diversity and biogeography of planktonic diatoms in Svalbard fjords. *Elem Sci Anth*, 10: 1.
DOI: https://doi.org/10.1525/elementa.2021.00117

RESEARCH ARTICLE

Diversity and biogeography of planktonic diatoms in Svalbard fjords: The role of dispersal and Arctic endemism in phytoplankton community structuring

Luka Šupraha[1,*], Kerstin Klemm[2], Sandra Gran-Stadniczeñko[1], Cora Hörstmann[2], Daniel Vaulot[3], Bente Edvardsen[1], and Uwe John[2,4,*]

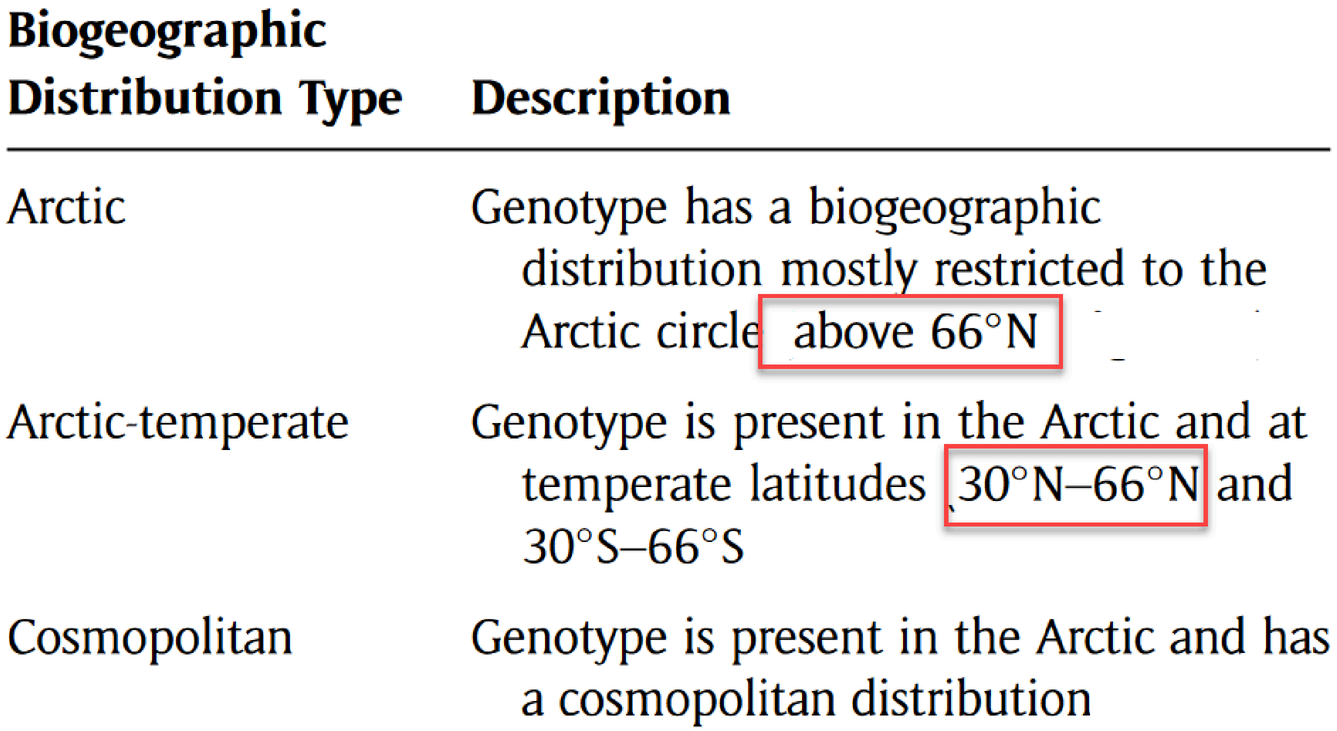




Šupraha, L. et al., 2022. Diversity and biogeography of planktonic diatoms in Svalbard fjords: the role of dispersal and Arctic endemism in

Svalbard diatoms

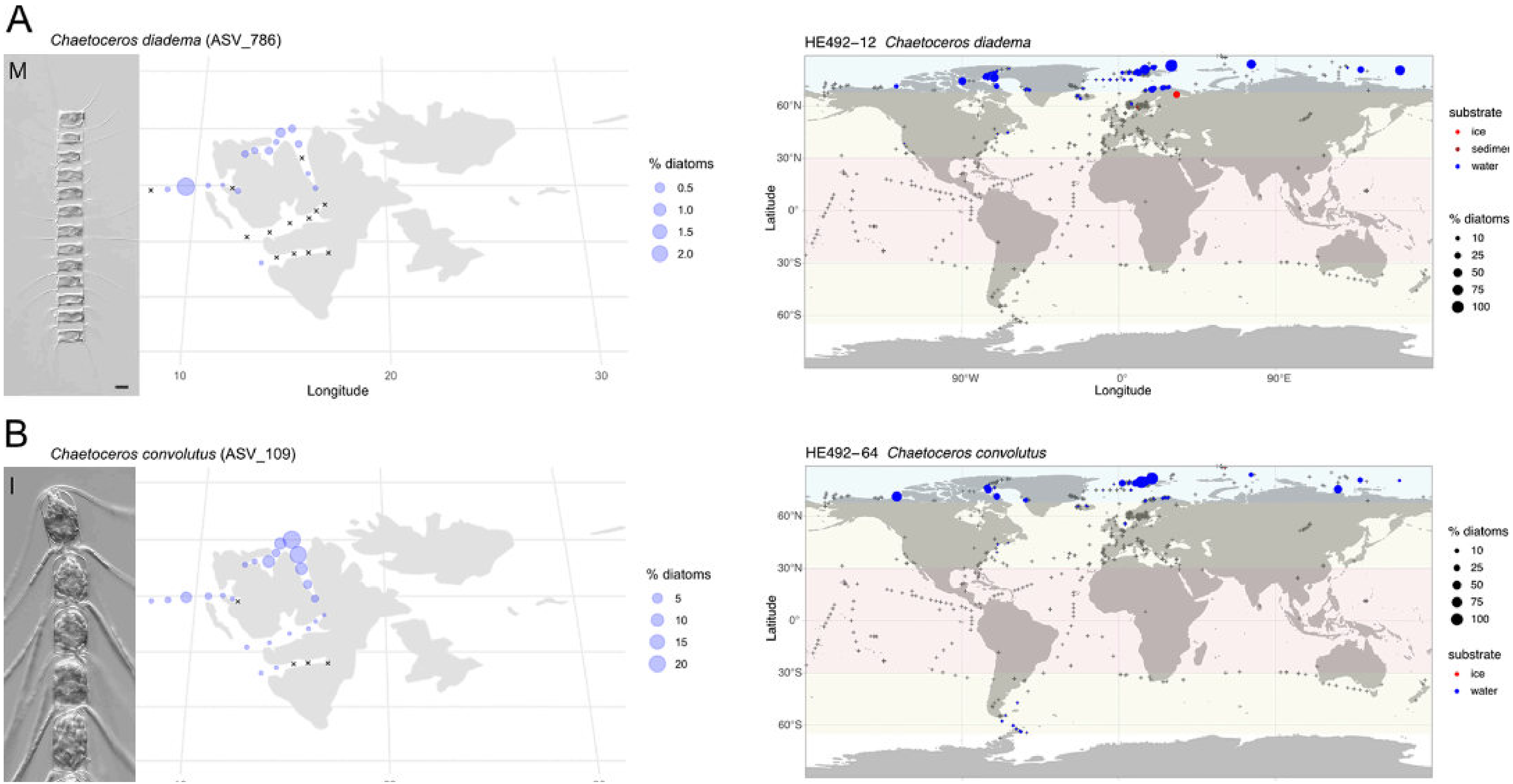
Biogeographic distribution types



Biogeographic Distribution Type	Description
Arctic	Genotype has a biogeographic distribution mostly restricted to the Arctic circle above 66°N
Arctic-temperate	Genotype is present in the Arctic and at temperate latitudes 30°N–66°N and 30°S–66°S
Cosmopolitan	Genotype is present in the Arctic and has a cosmopolitan distribution

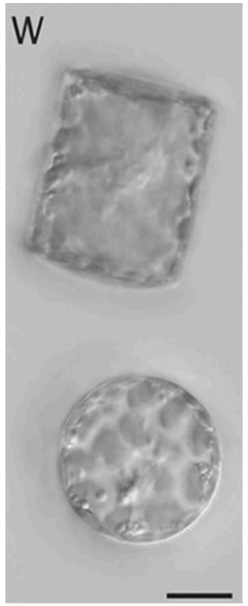
Svalbard diatoms

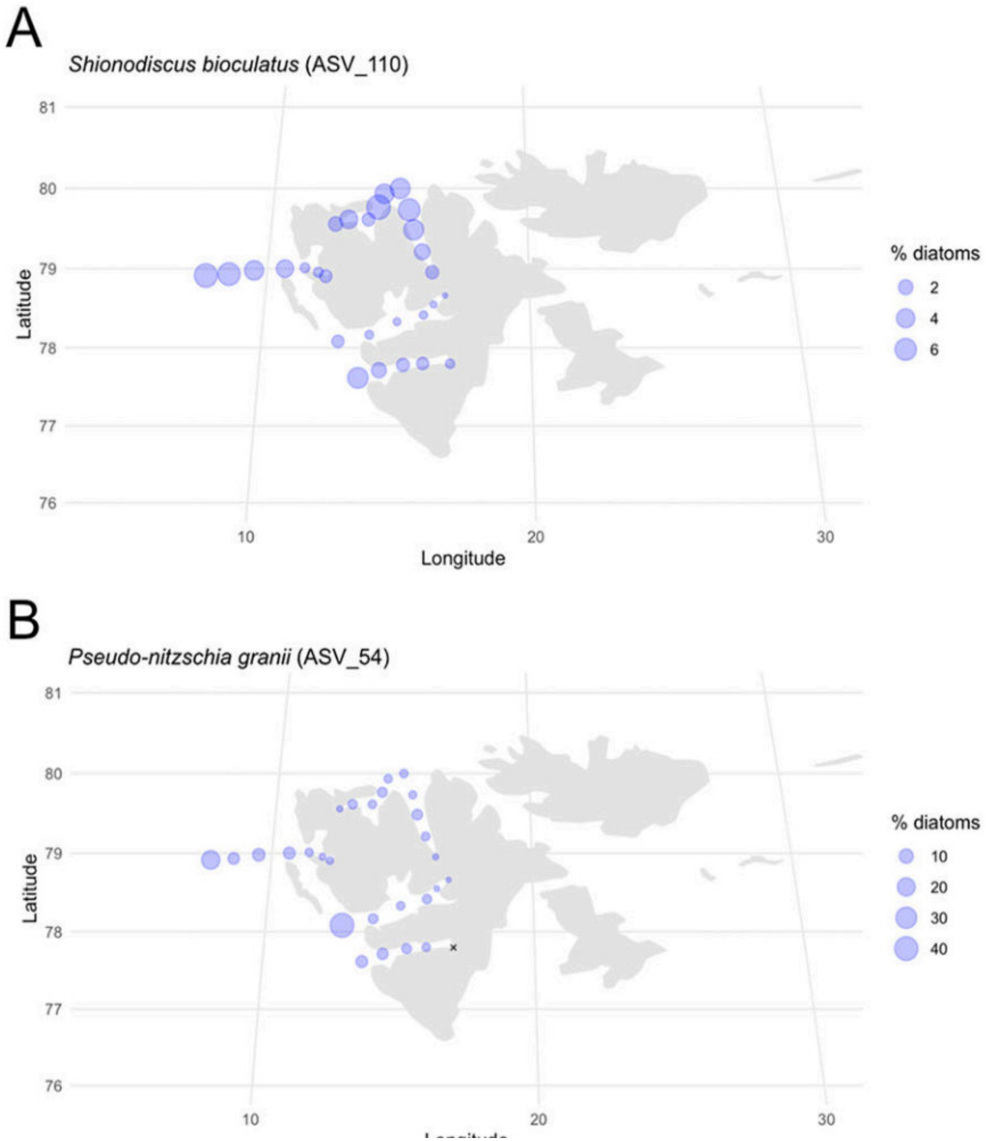
Polar genotypes

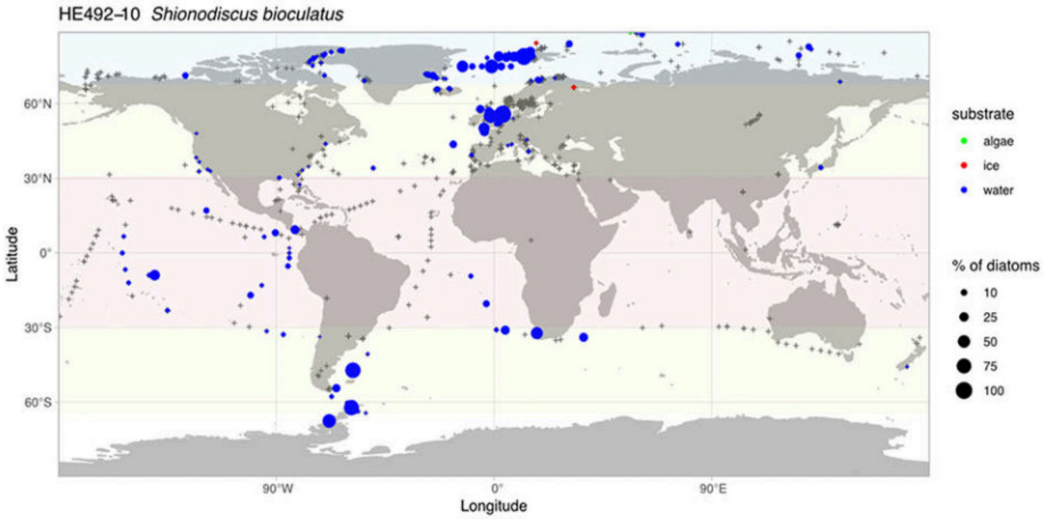


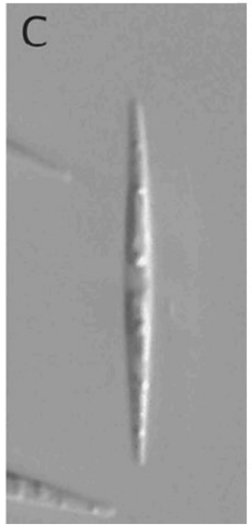
Svalbard diatoms

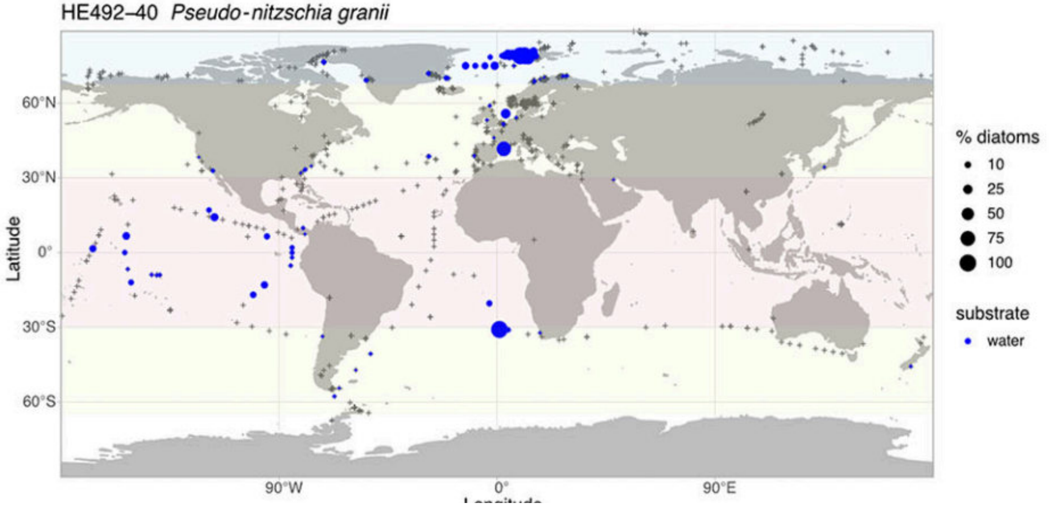
Cosmopolitan genotypes





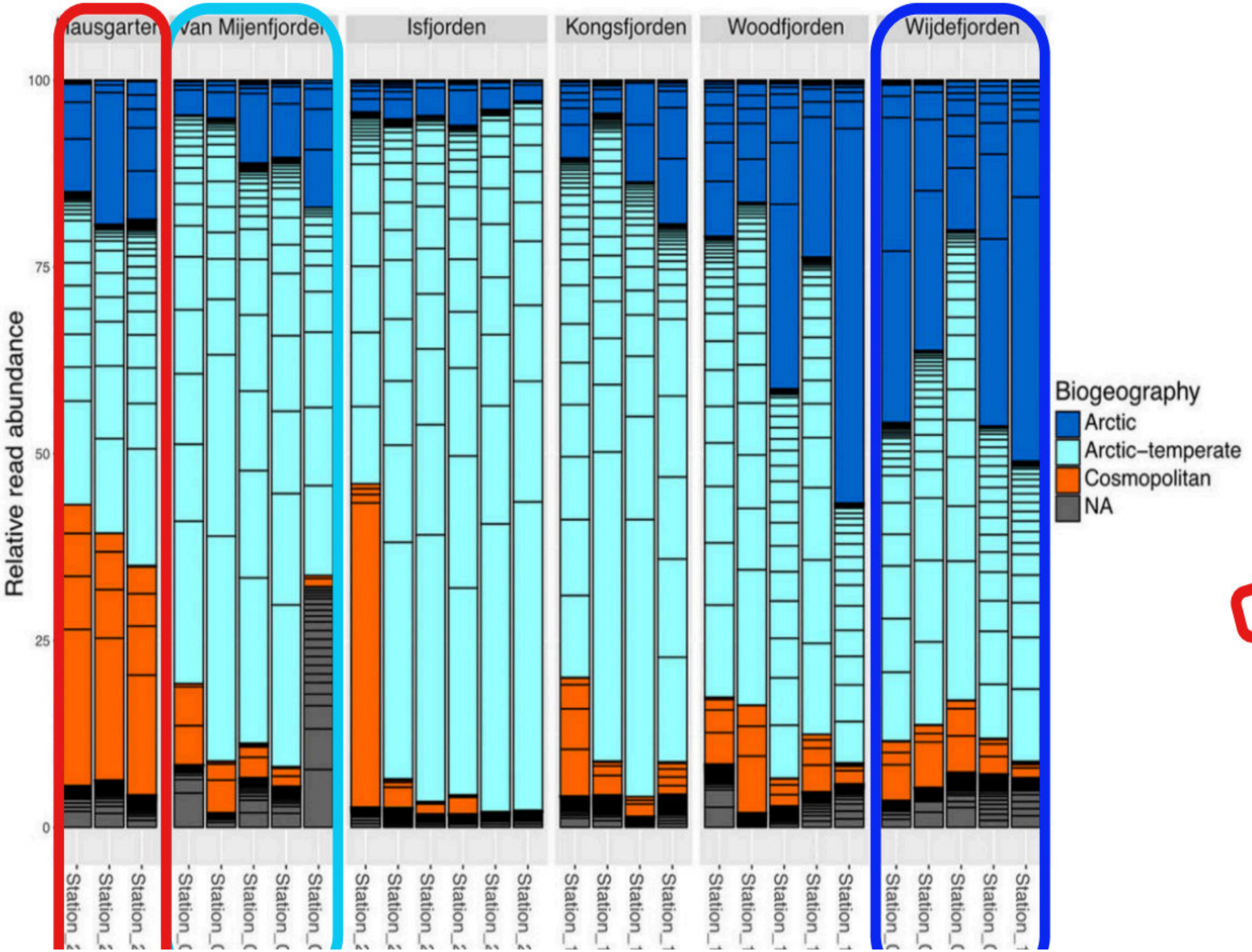


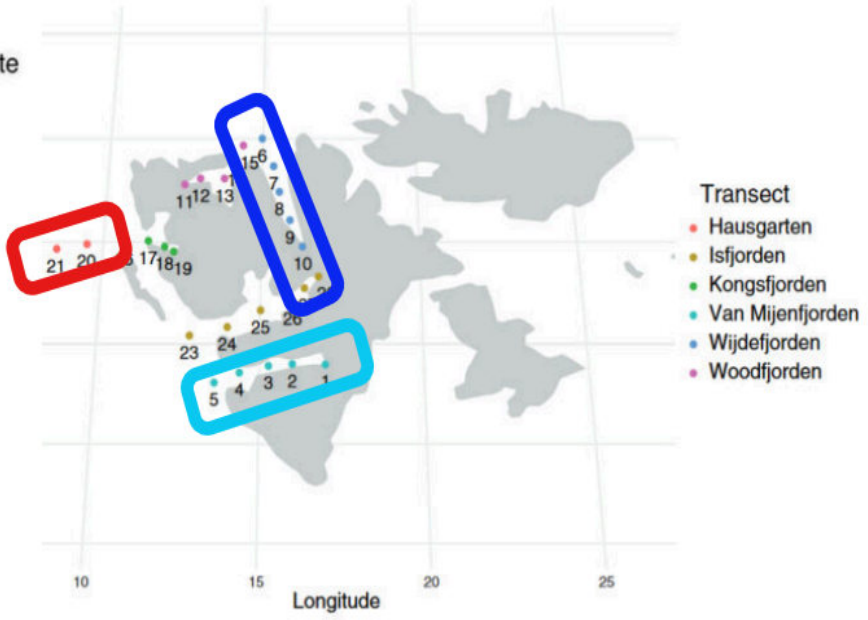




Svalbard diatoms

Distribution in fjords





Under ice bloom

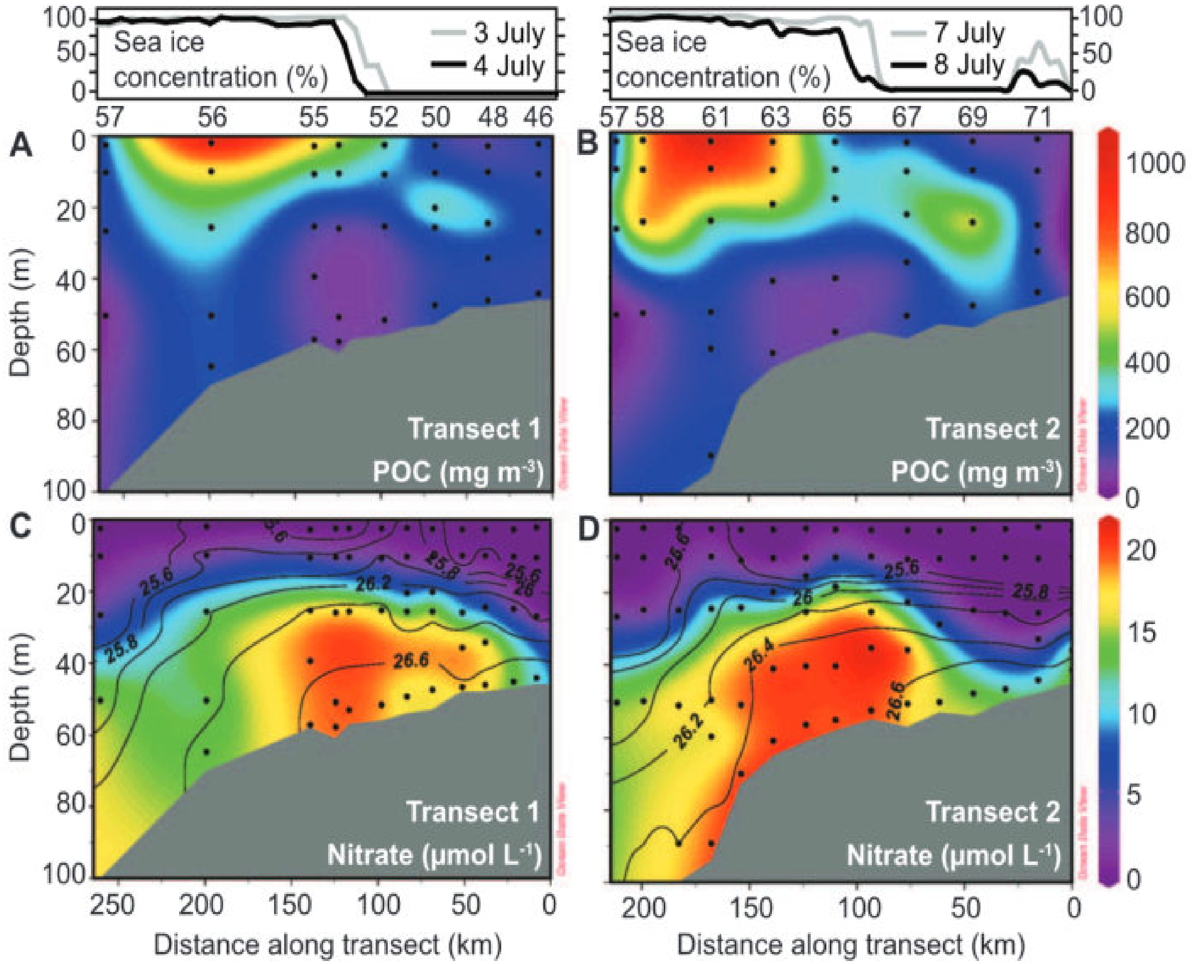
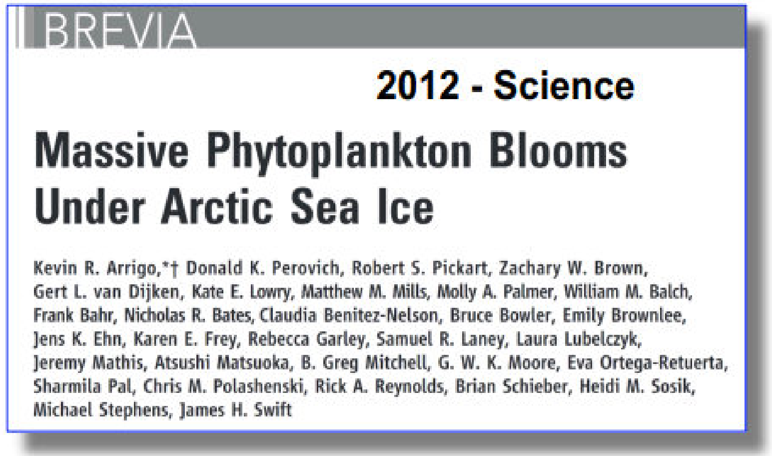


Fig. 1. Under-ice phytoplankton bloom observed during ICESCAPE 2011. (**A**) Particulate organic carbon (POC) and (**C**) nitrate from transect 1. (**B**) POC and (**D**) nitrate from transect 2. Sea ice concentrations and station numbers are shown above (A) and (B); black dots represent sampling depths; black

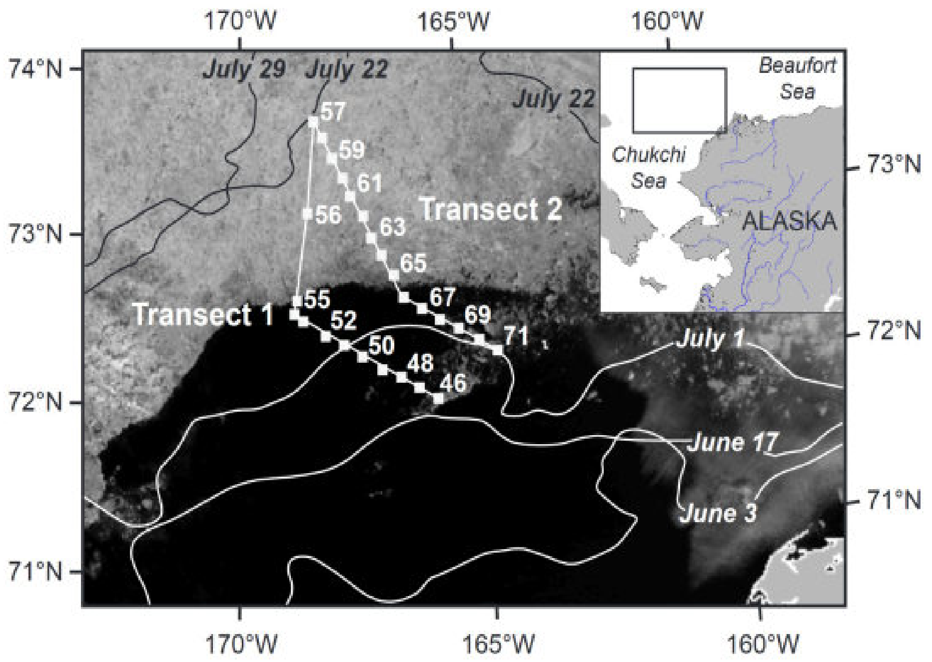


BREVIA

2012 - Science

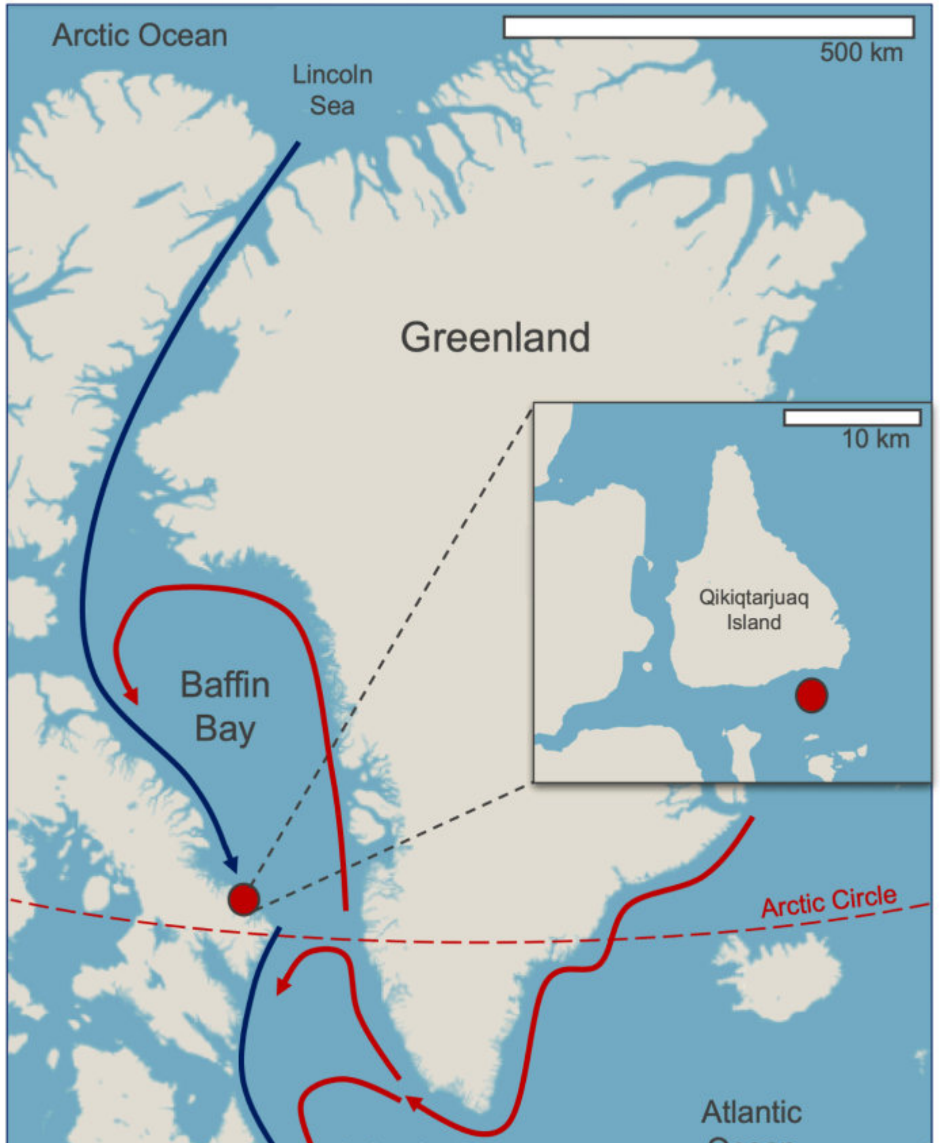
Massive Phytoplankton Blooms Under Arctic Sea Ice

Kevin R. Arrigo,*† Donald K. Perovich, Robert S. Pickart, Zachary W. Brown, Gert L. van Dijken, Kate E. Lowry, Matthew M. Mills, Molly A. Palmer, William M. Balch, Frank Bahr, Nicholas R. Bates, Claudia Benitez-Nelson, Bruce Bowler, Emily Brownlee, Jens K. Ehn, Karen E. Frey, Rebecca Garley, Samuel R. Laney, Laura Lubelczyk, Jeremy Mathis, Atsushi Matsuoka, B. Greg Mitchell, G. W. K. Moore, Eva Ortega-Retuerta, Sharmila Pal, Chris M. Polashenski, Rick A. Reynolds, Brian Schieber, Heidi M. Sosik, Michael Stephens, James H. Swift



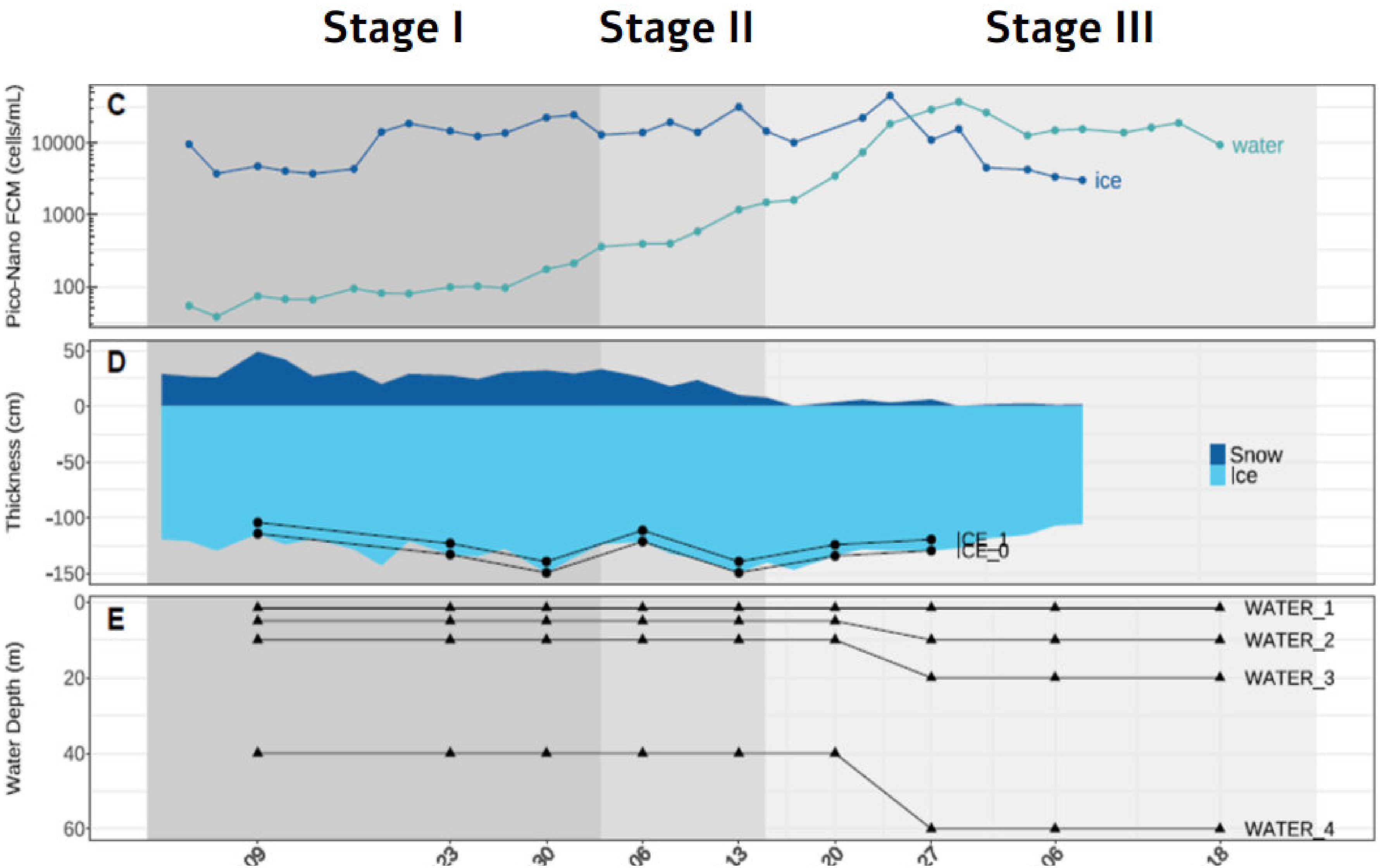
Under ice bloom

Green edge - 2016



Under ice bloom

Green edge - 2016



Ice

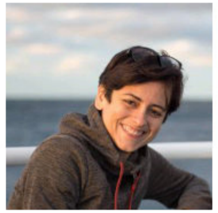
Water

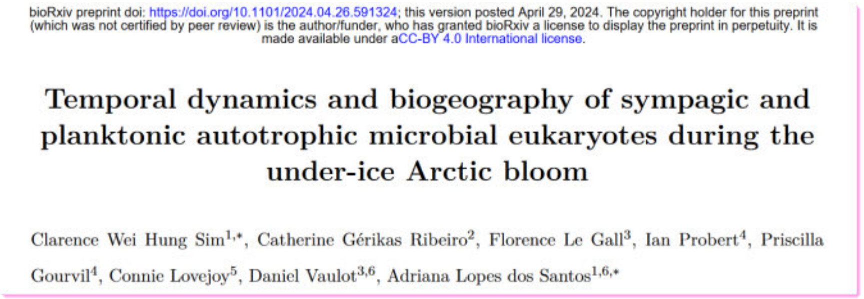
Under ice bloom

Metabarcoding

Clarence Sim – PhD candidate

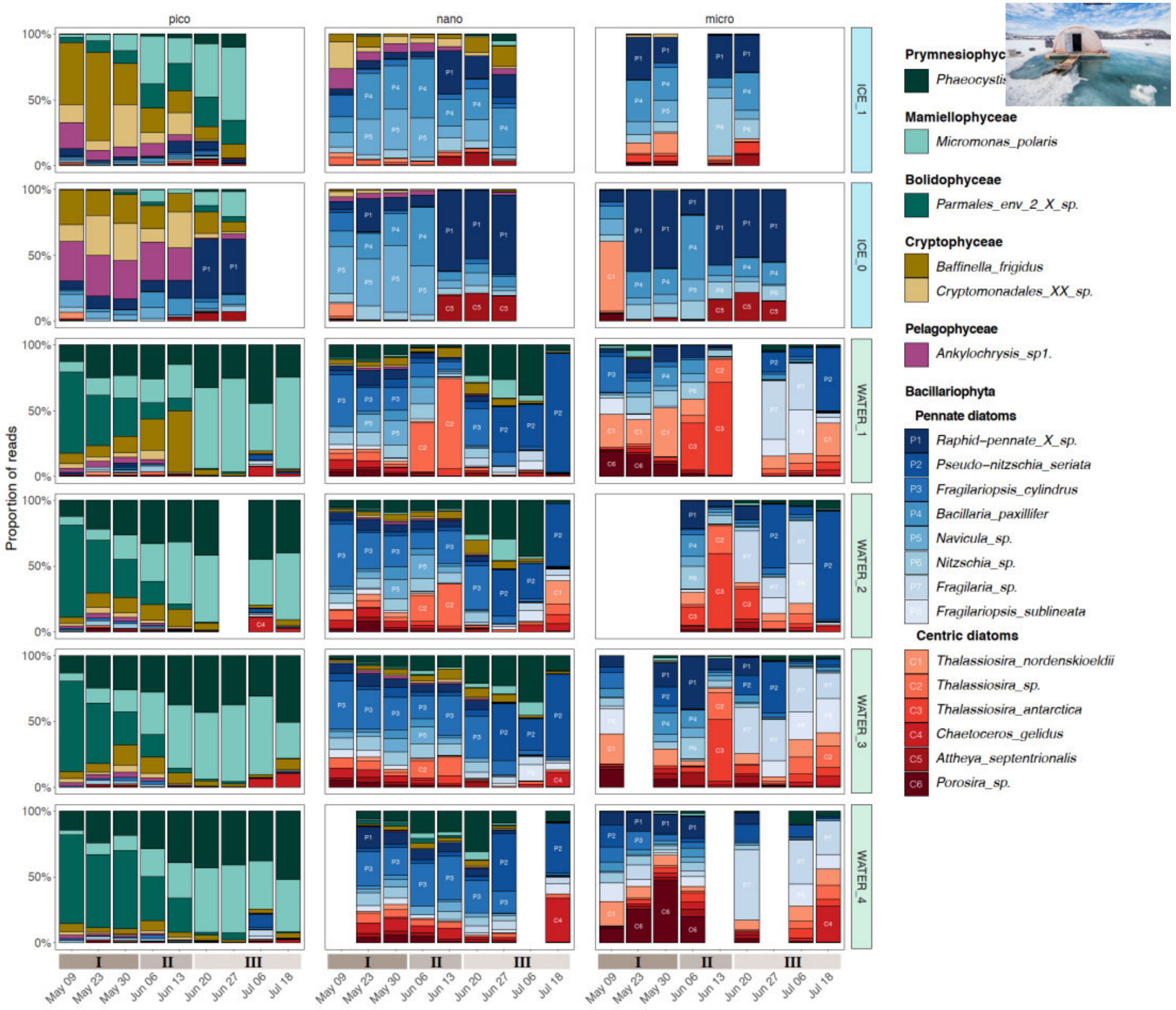
Nanyang Technological University





Temporal dynamics and biogeography of sympagic and planktonic autotrophic microbial eukaryotes during the under-ice Arctic bloom

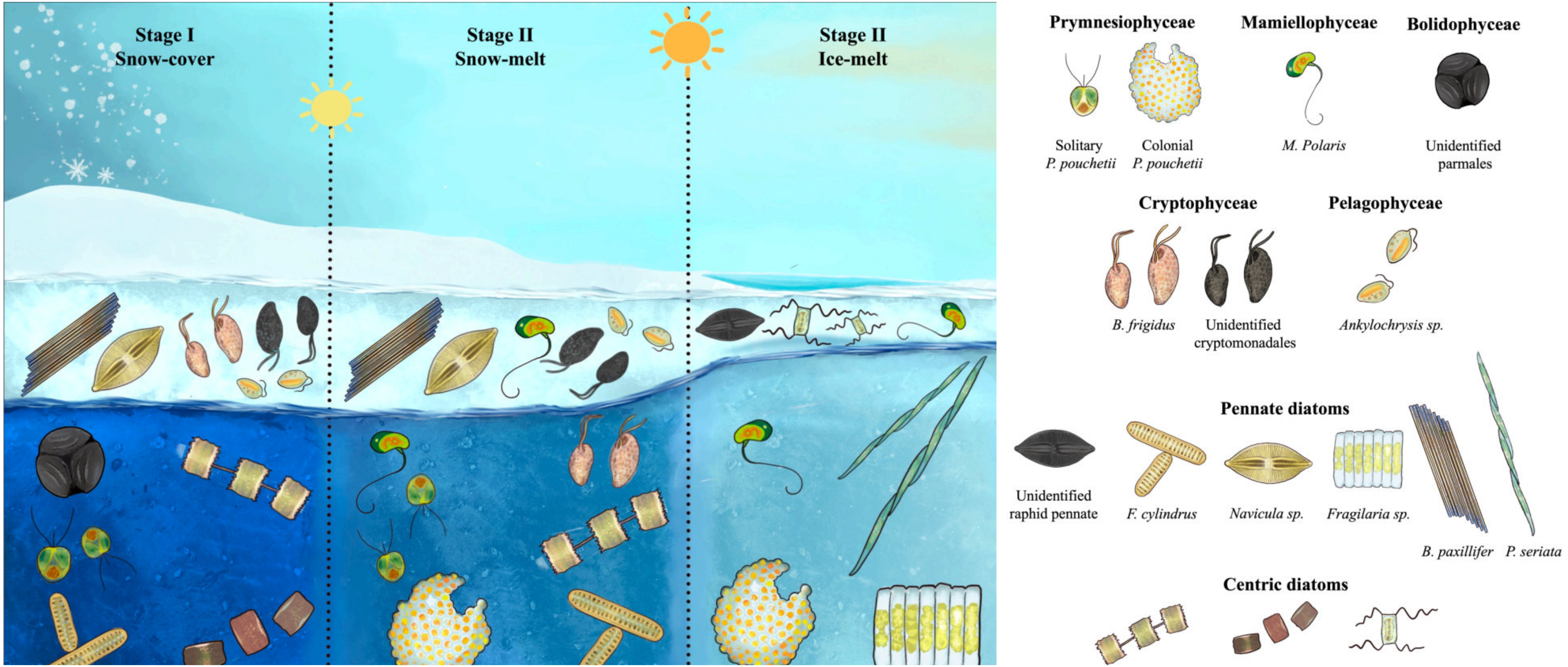
Clarence Wei Hung Sim[1,*], Catherine Gérikas Ribeiro[2], Florence Le Gall[3], Ian Probert[4], Priscilla Gourvil[4], Connie Lovejoy[5], Daniel Vaulot[3,6], Adriana Lopes dos Santos[1,6,*]



Sim et al. 2024. Temporal dynamics and biogeography of sympagic and planktonic autotrophic microbial eukaryotes during the under-ice Arctic

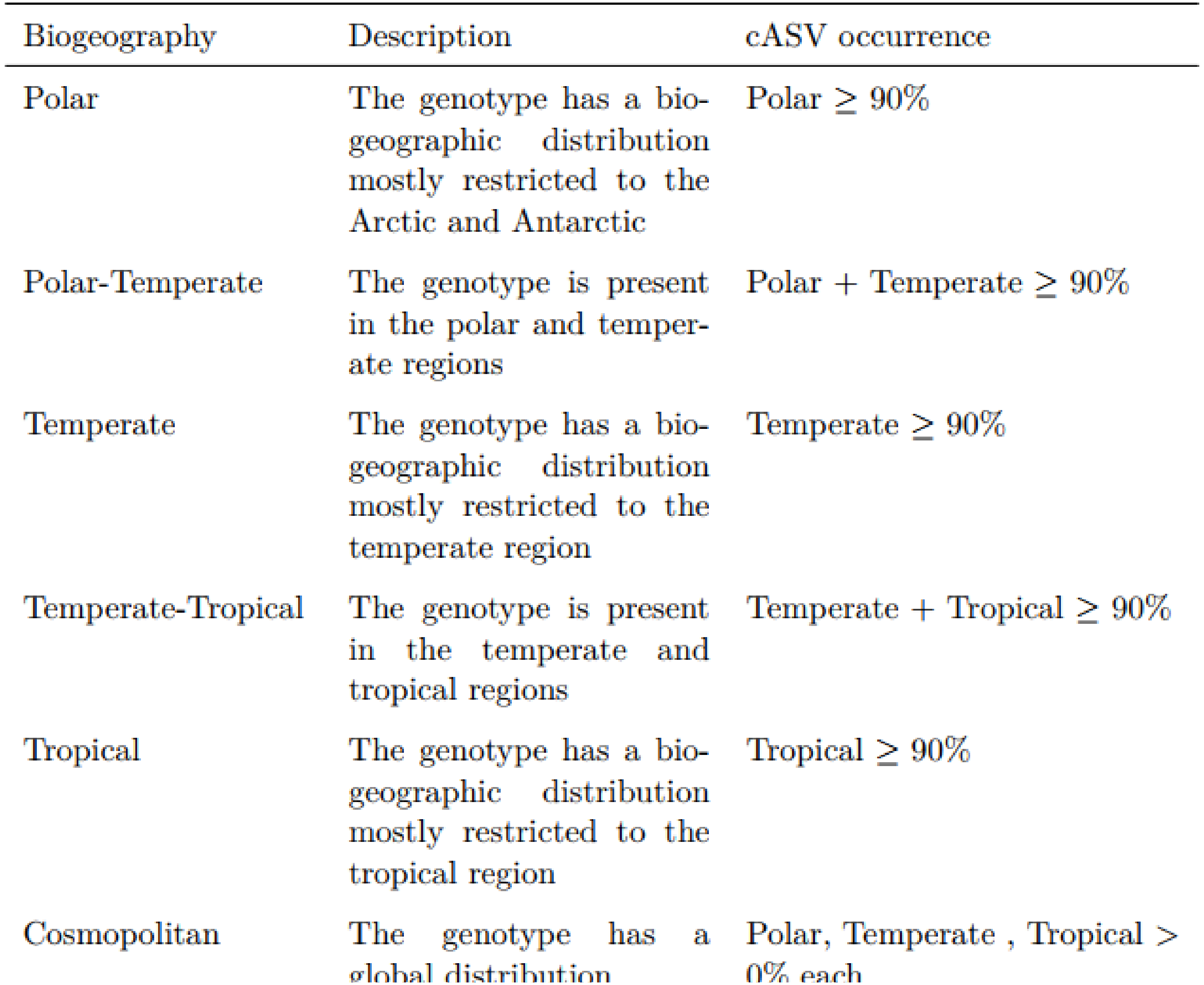
Under ice bloom

Species succession



Under ice bloom

Biogeography - metaPR2

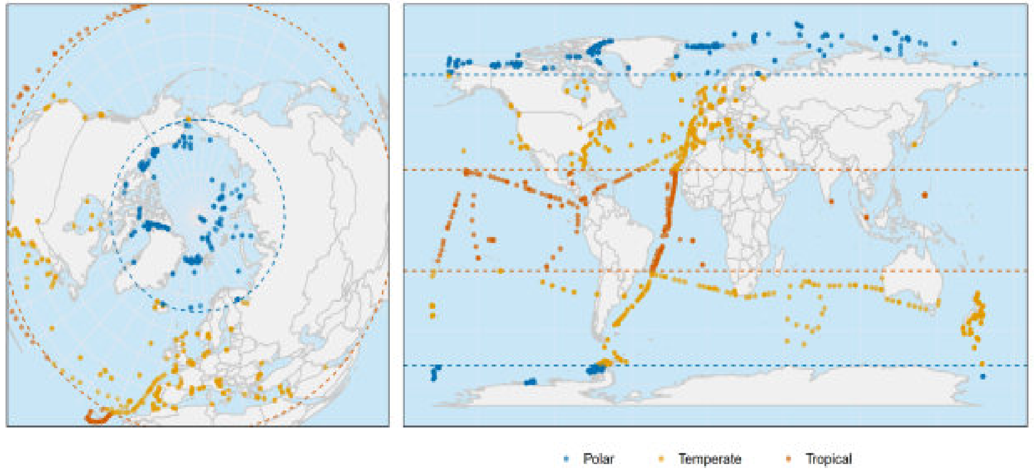


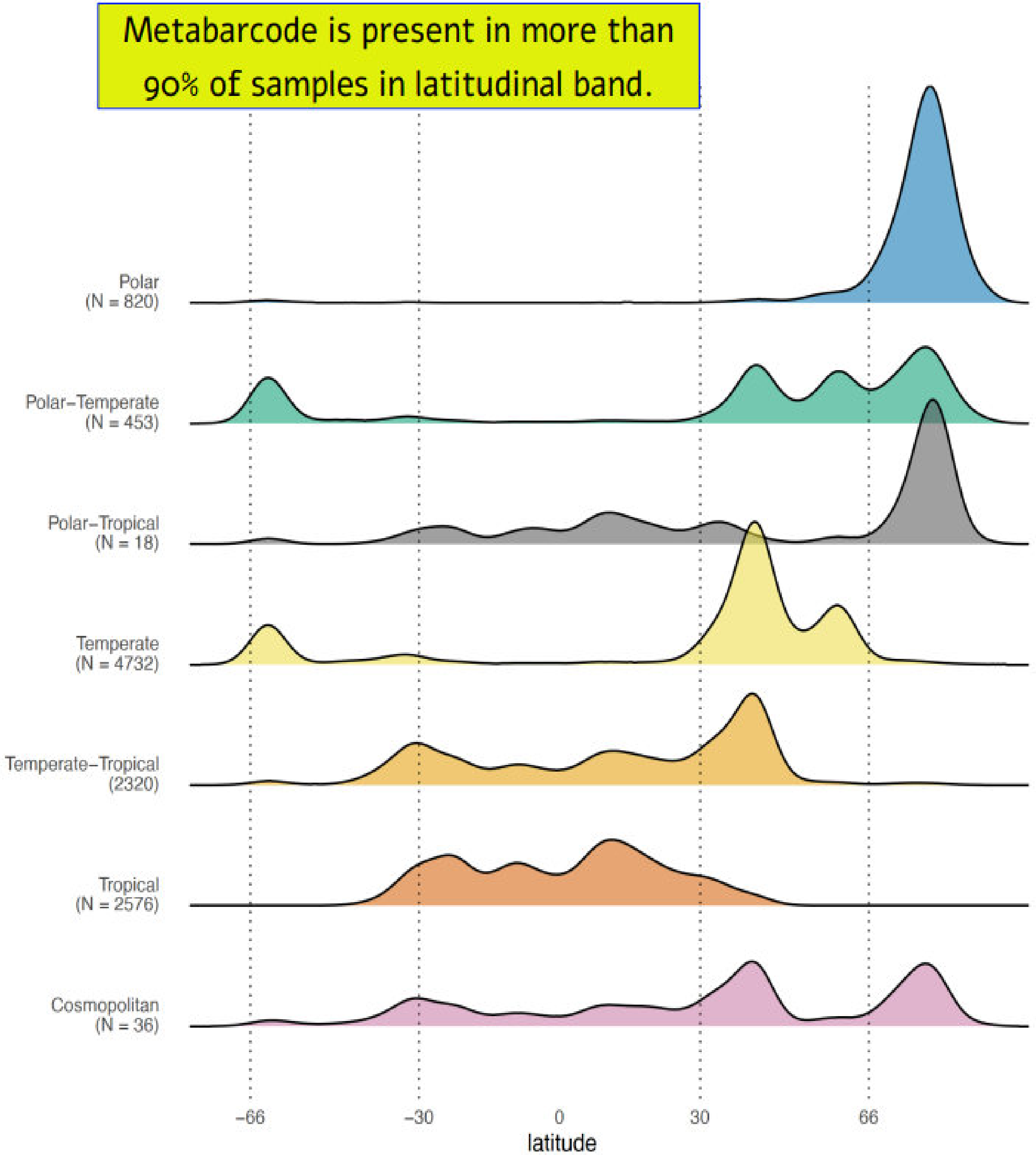
Biogeography	Description	cASV occurrence
Polar	The genotype has a biogeographic distribution mostly restricted to the Arctic and Antarctic	Polar $\geq$ 90%
Polar-Temperate	The genotype is present in the polar and temperate regions	Polar + Temperate $\geq$ 90%
Temperate	The genotype has a biogeographic distribution mostly restricted to the temperate region	Temperate $\geq$ 90%
Temperate-Tropical	The genotype is present in the temperate and tropical regions	Temperate + Tropical $\geq$ 90%
Tropical	The genotype has a biogeographic distribution mostly restricted to the tropical region	Tropical $\geq$ 90%
Cosmopolitan	The genotype has a global distribution	Polar, Temperate , Tropical > 0% each

Under ice bloom

Biogeography - global ASVs

- Polar 820
- Polar-temperate 453
- Polar- tropical 18
- Temperate 4732
- Temperate-tropical 2320
- Tropical 2576
- Cosmopolitan 36

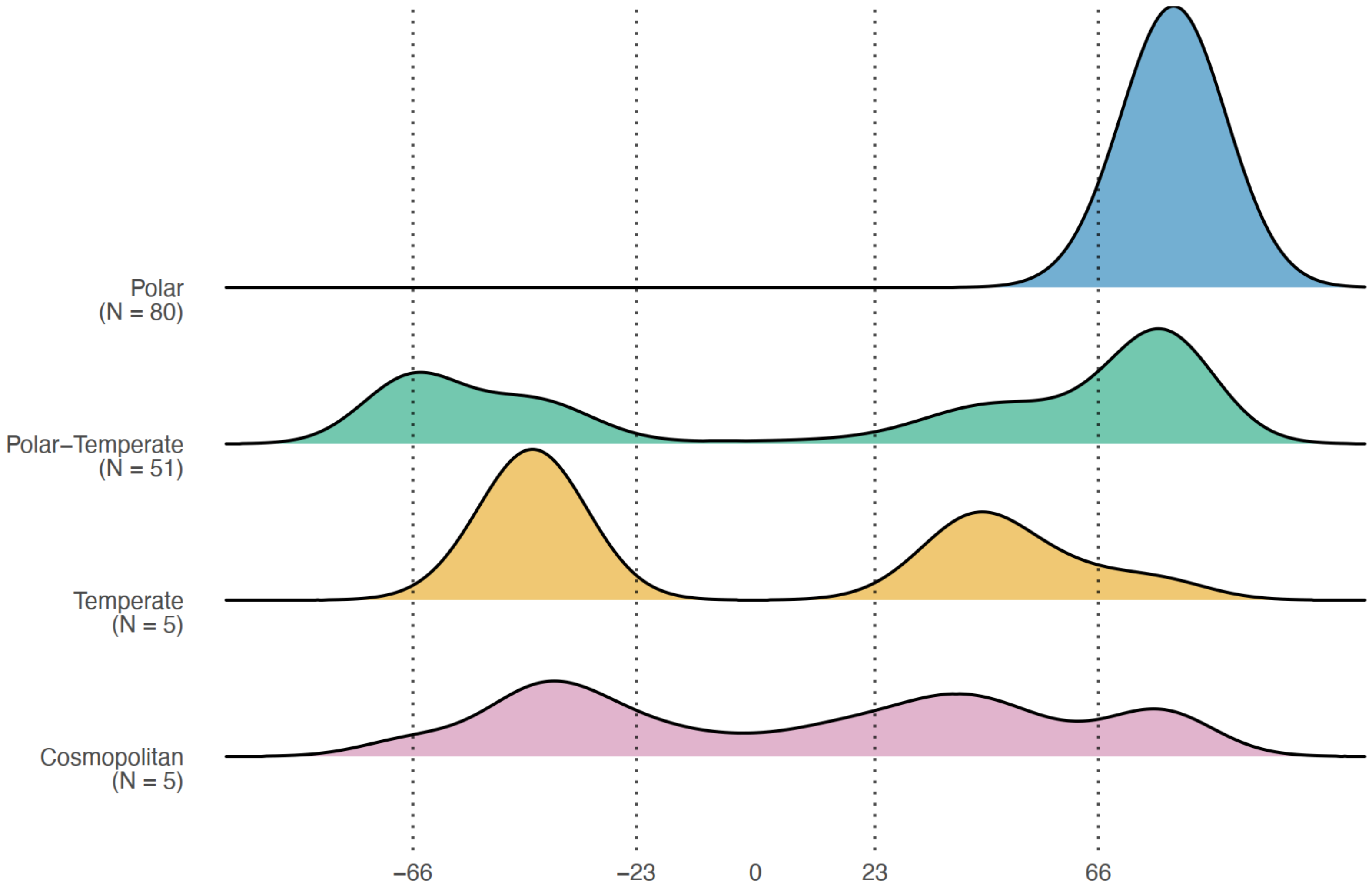



Metabarcode is present in more than 90% of samples in latitudinal band.

Under ice bloom

Biogeography - Green Edge ASVs

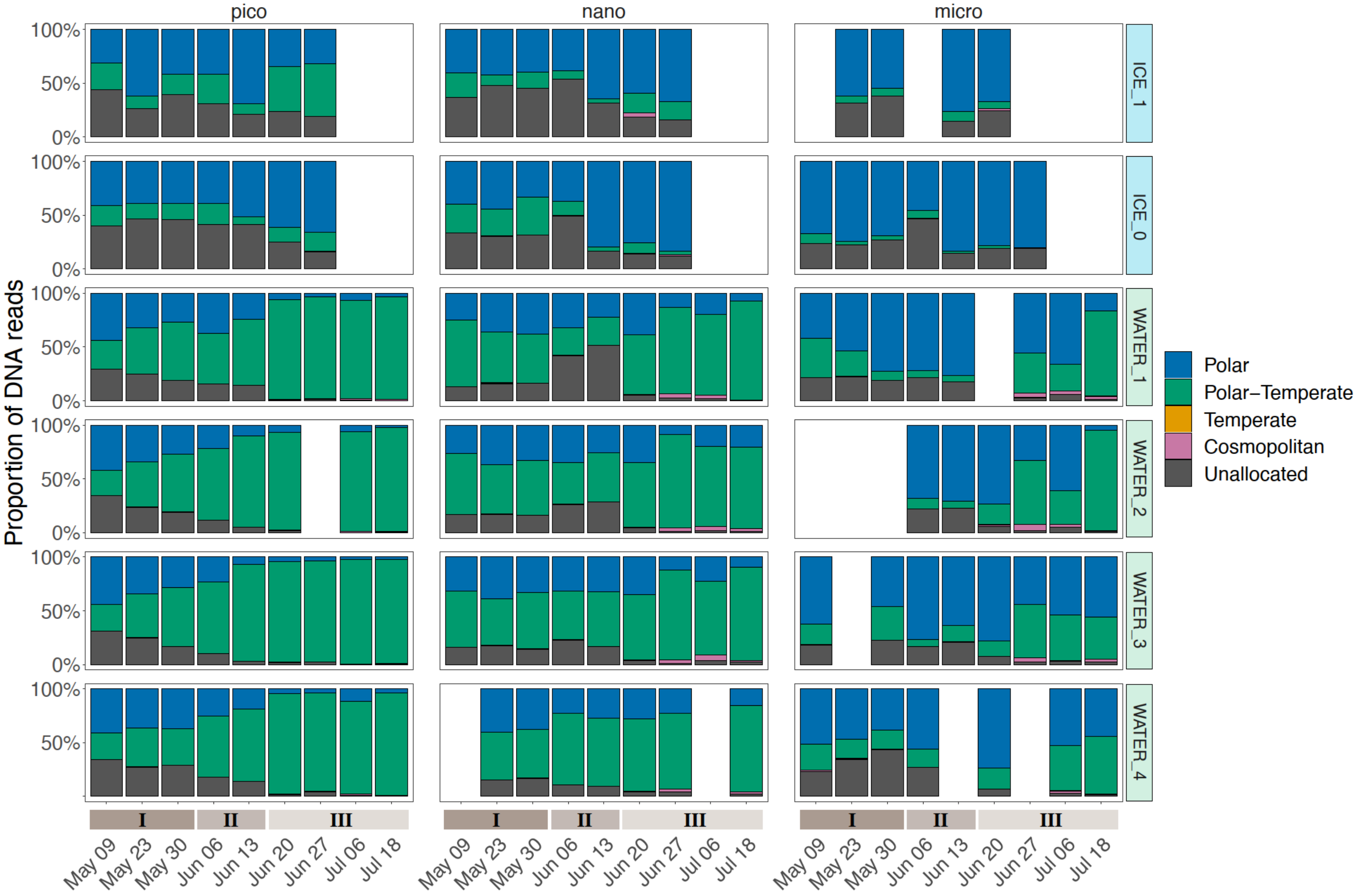
- **Polar 80**
- **Polar-temperate 51**
- Polar- tropical
- Temperate 5
- Temperate-tropical
- Tropical
- Cosmopolitan 5



Under ice bloom

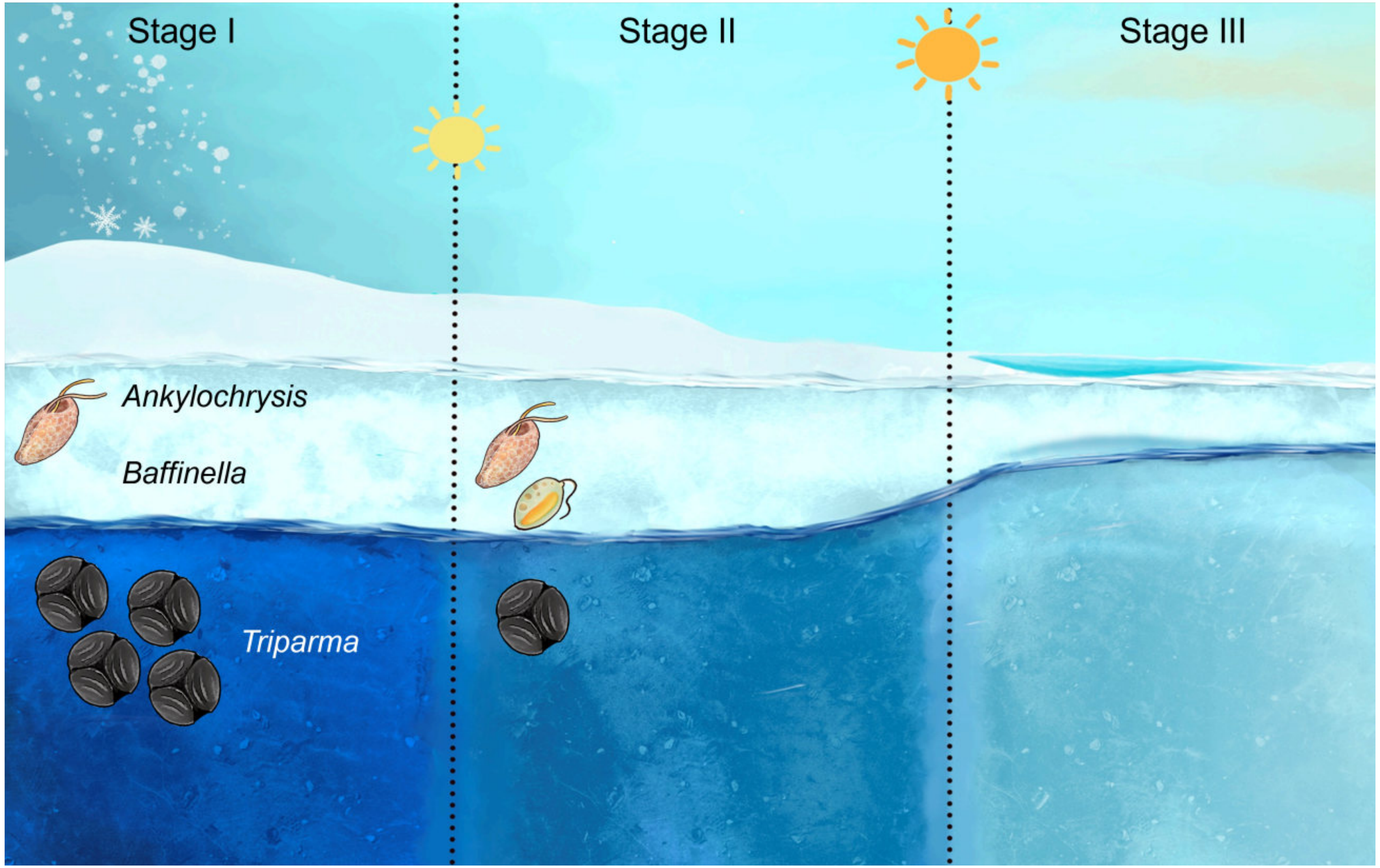
Community change

- Ice: more polar species
- Water: increase of polar-temperate in stage III
- Nano and micro communities are more stable



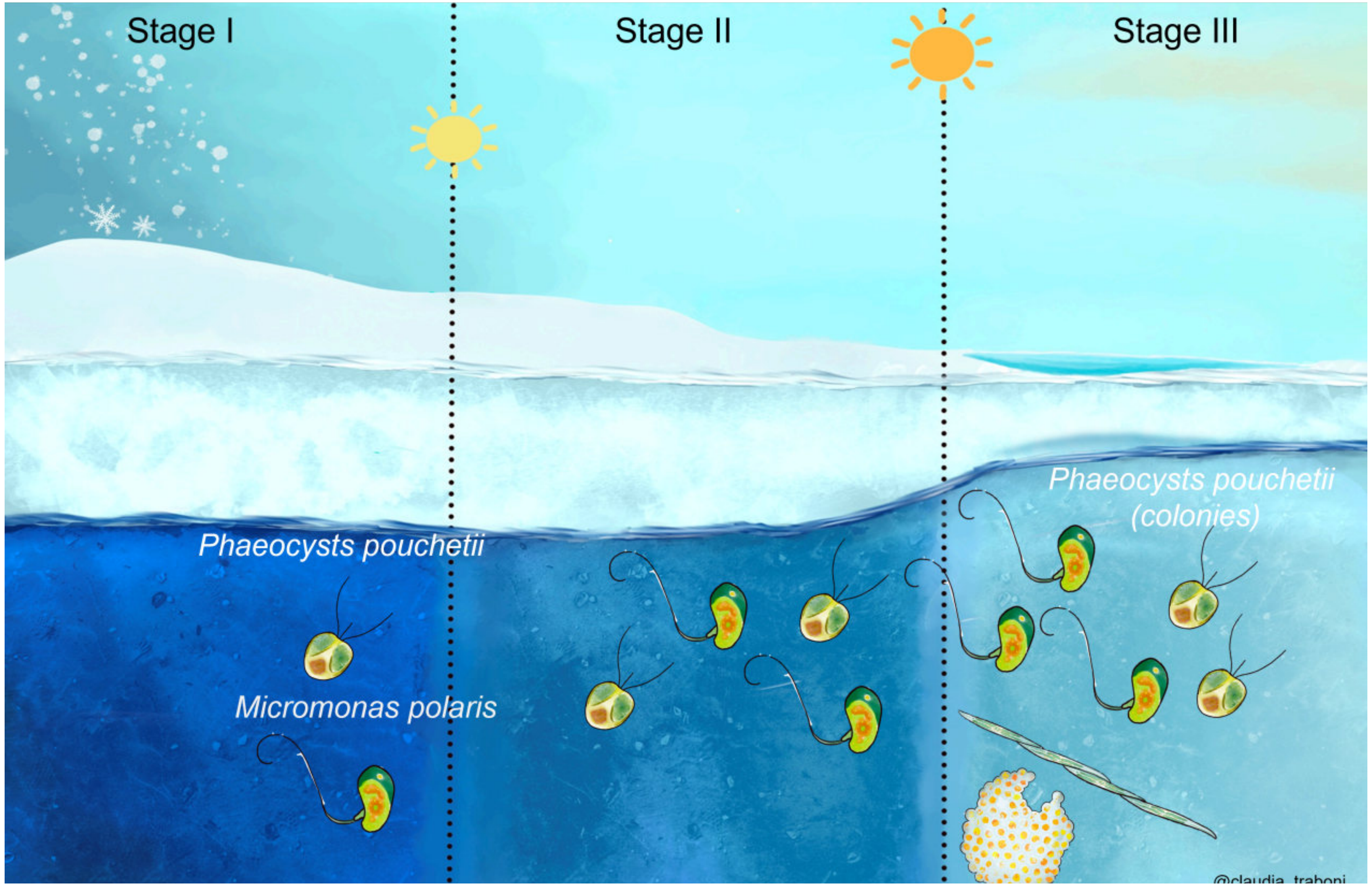
Under ice bloom

Early stage - Polar species



Under ice bloom

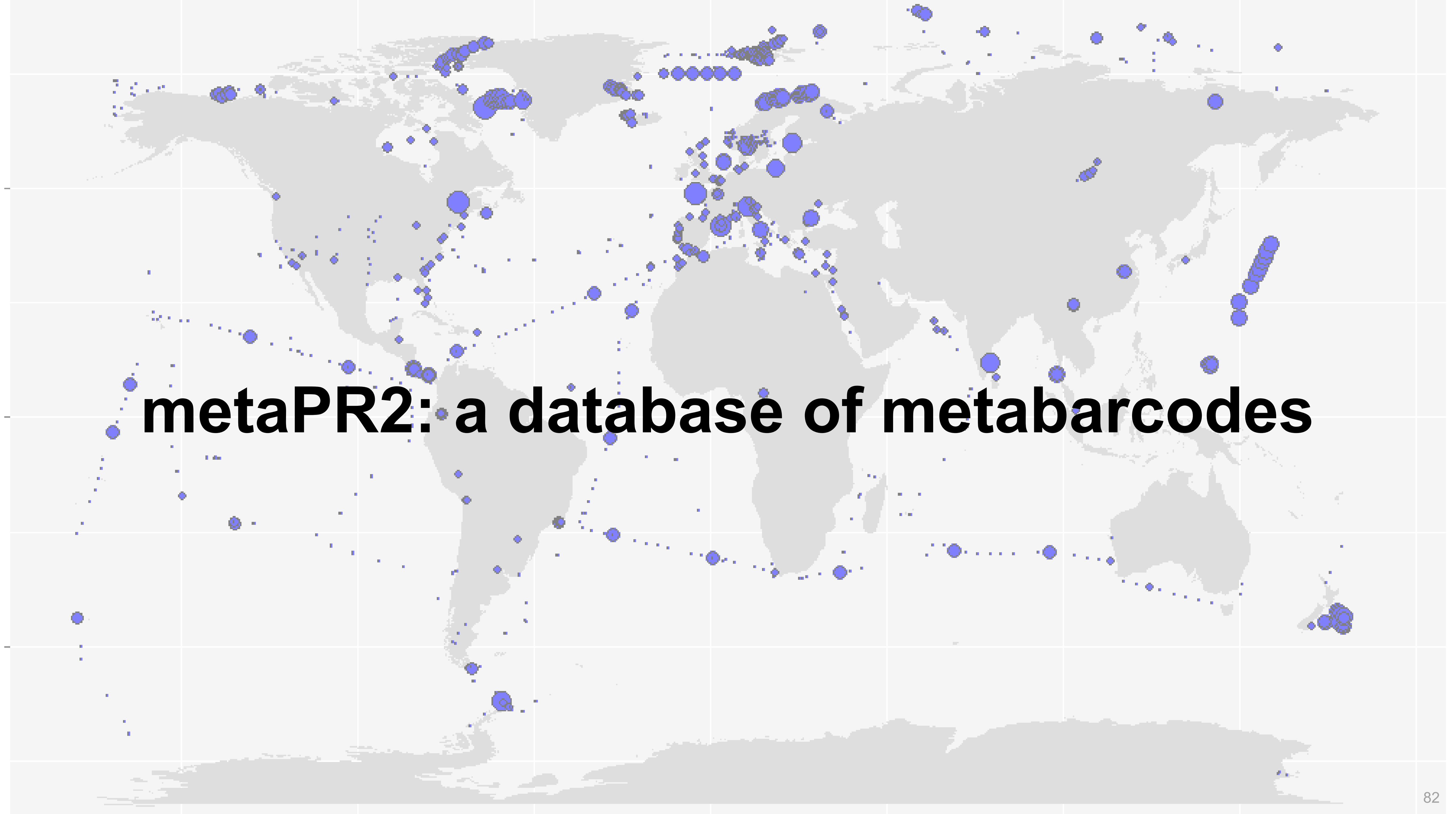
Late stage - Polar-temperate species



Take home messages

- Metabarcoding is a very powerful tool to asses community composition
    - Most used regions are 18S V4 and V9
- Example of use
    - Assess environmental nich of species
    - Biogeography
- Future direction
    - Validation by comparison with microscopy/automated image
    - Quantification with internal standard
    - rRNA operon metabarcoding

Questions ?



metaPR2: a database of metabarcodes

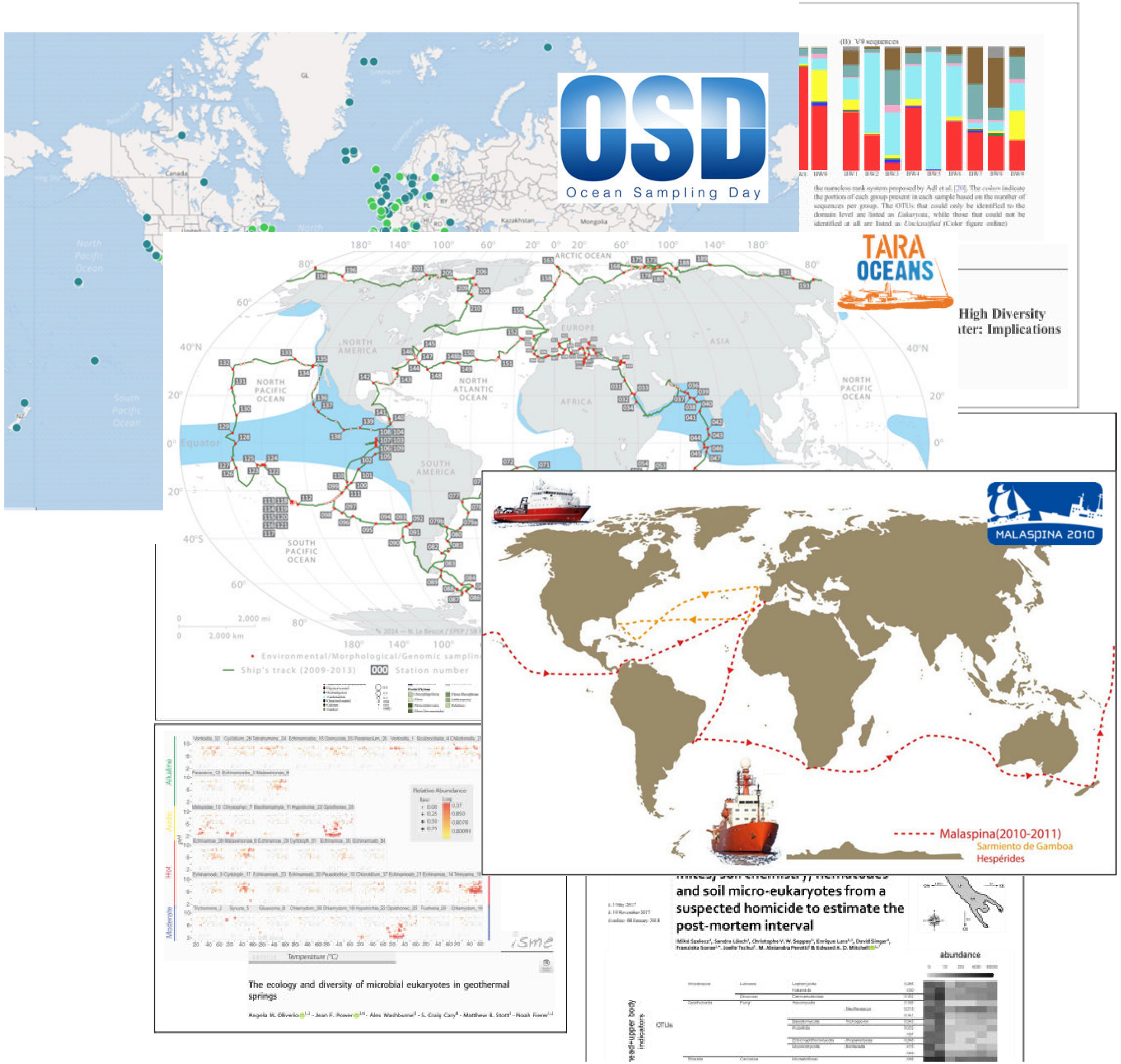
Many metabarcoding studies for eukaryotes

But hard to use…

- Processed with different pipelines
- Different primers
- Different levels of similarity
- Different reference databases
- Metadata lacking

Large datasets

- Ocean Sampling Day
- Tara Oceans
- Malaspina



The ecology and diversity of microbial eukaryotes in geothermal springs

and soil micro-eukaryotes from a suspected homicide to estimate the post-mortem interval

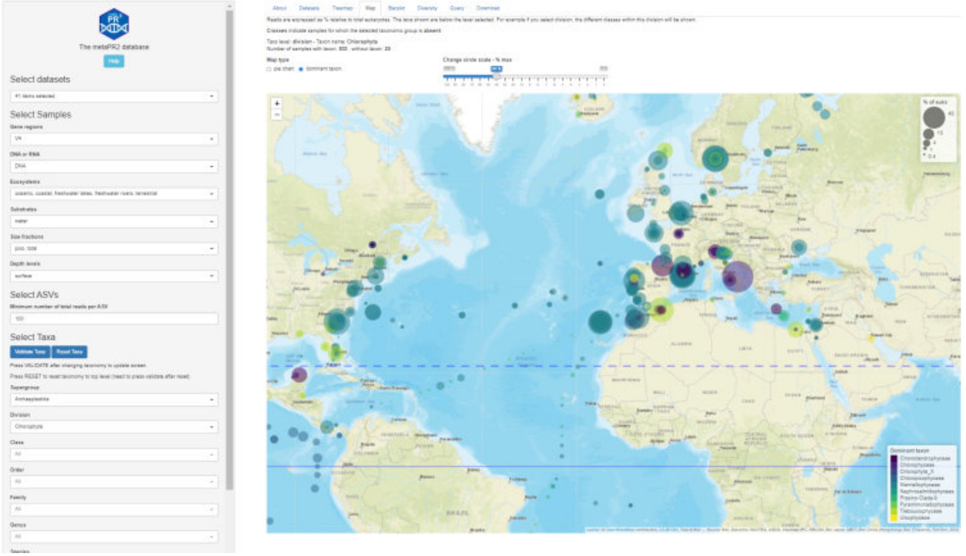
metaPR2 a database of metabarcodes

Reprocess public data

- Download Genbank (SRA) data
  - Raw sequences
  - Metadata
- Reprocess
  - Amplicon Sequence Variant (*dada2*)
  - Merge ASVs with same sequence
- Store in MySQL database
- Develop under R
- Web interface and R package
- https://app.metapr2.org

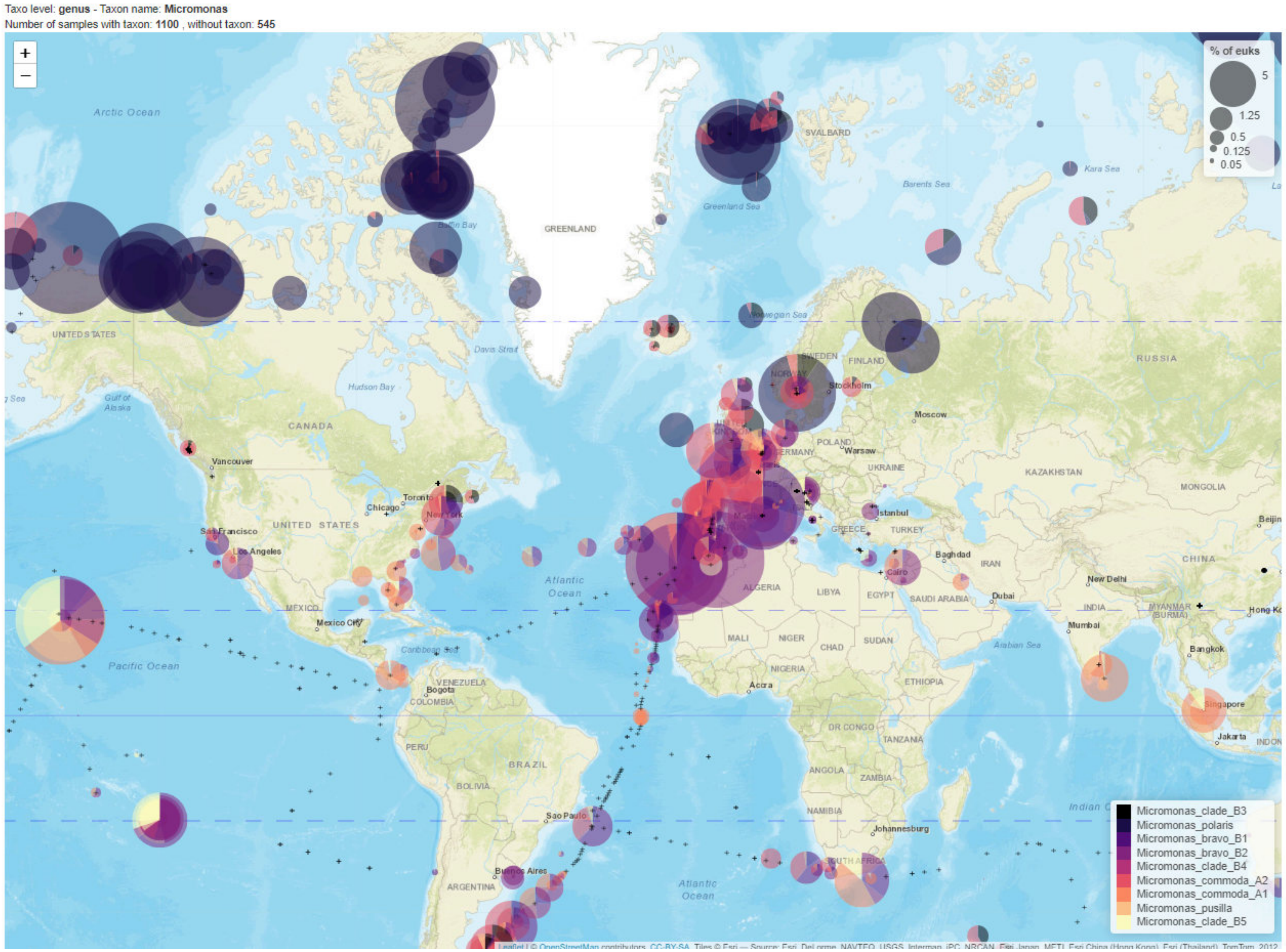
Current status

- Version 2.1
- Datasets: 59
- Samples: 6,202
- Barcodes (ASVs): 93,127



Vaulot, D. et al. metaPR2: a database of eukaryotic 18S rRNA metabarcodes with an emphasis on protists. Molecular Ecology Resources 22,

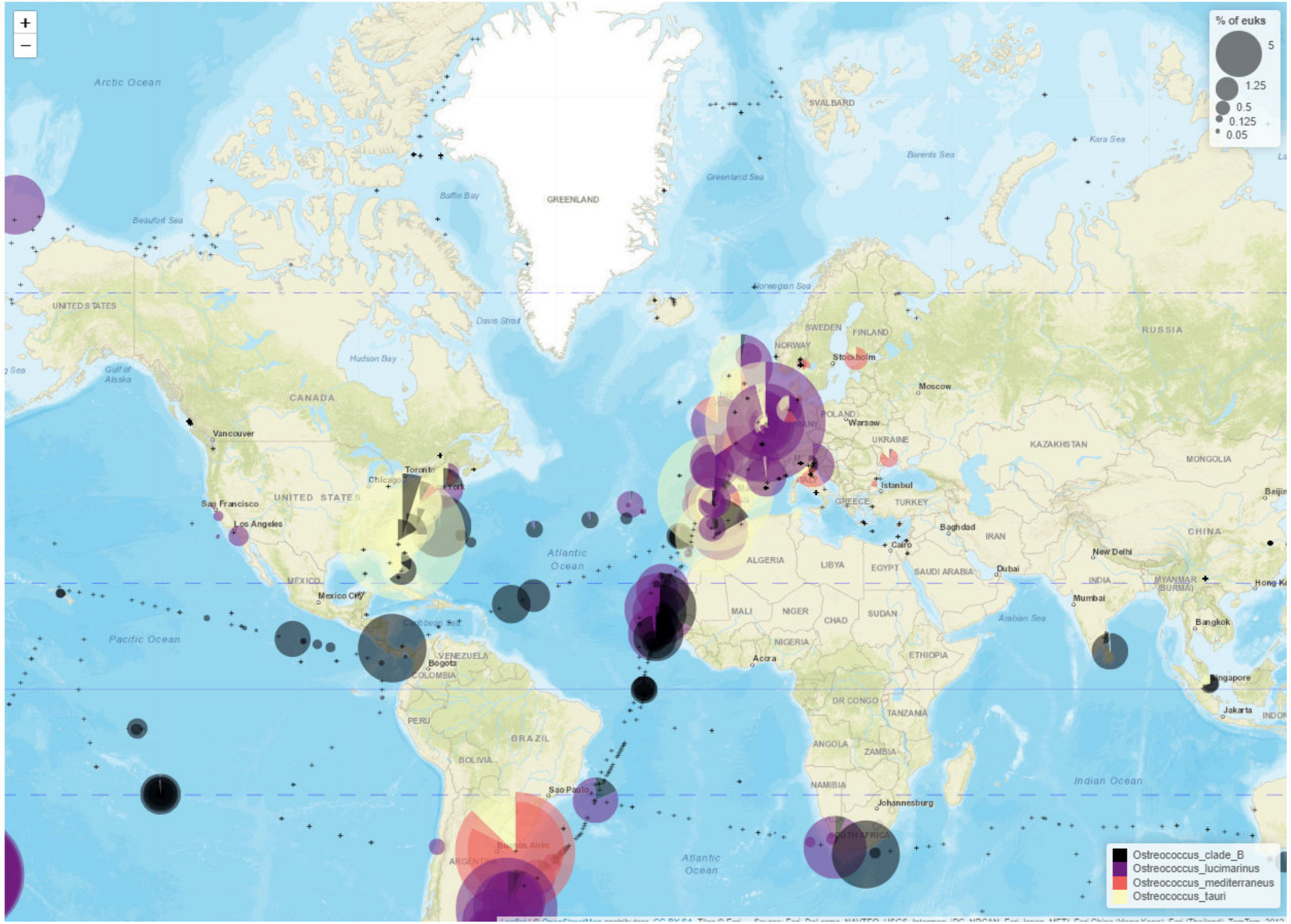
Example: Biogeography of *Micromonas*



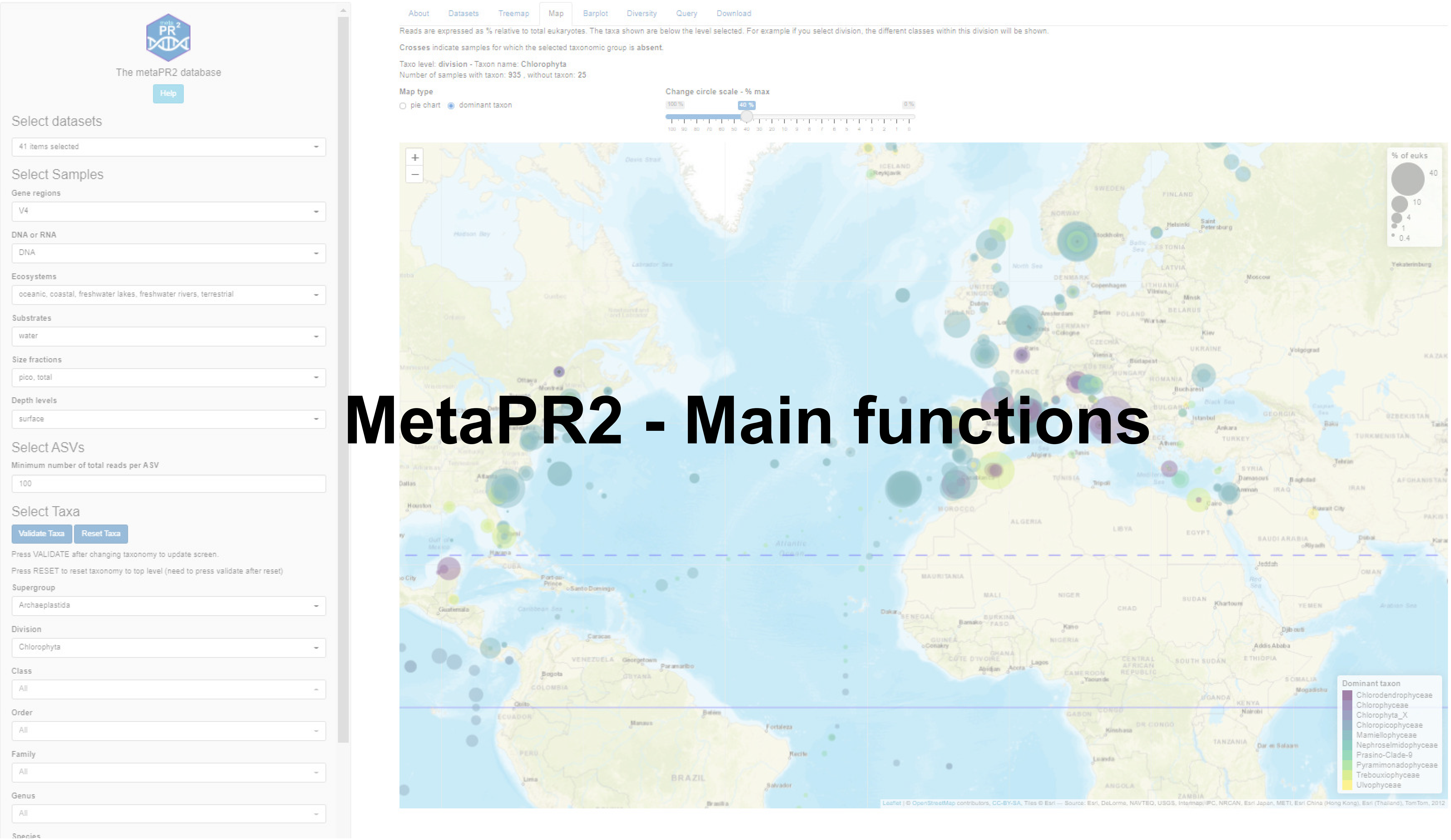
Taxo level: **genus** - Taxon name: **Micromonas**
Number of samples with taxon: **1100** , without taxon: **545**

Example: Biogeography of *Ostreococcus*

Taxo level: **genus** - Taxon name: **Ostreococcus**
Number of samples with taxon: **695** , without taxon: **950**



MetaPR2 - Main functions



The metaPR2 database

Help

Select datasets

41 items selected

Select Samples

Gene regions: V4

DNA or RNA: DNA

Ecosystems: oceanic, coastal, freshwater lakes, freshwater rivers, terrestrial

Substrates: water

Size fractions: pico, total

Depth levels: surface

Select ASVs

Minimum number of total reads per ASV: 100

Select Taxa

Validate Taxa | Reset Taxa

Press VALIDATE after changing taxonomy to update screen.

Press RESET to reset taxonomy to top level (need to press validate after reset)

Supergroup: Archaeplastida

Division: Chlorophyta

Class: All

Order: All

Family: All

Genus: All

Species

About | Datasets | Treemap | Map | Barplot | Diversity | Query | Download

Reads are expressed as % relative to total eukaryotes. The taxa shown are below the level selected. For example if you select division, the different classes within this division will be shown.

Crosses indicate samples for which the selected taxonomic group is absent.

Taxo level: **division** - Taxon name: **Chlorophyta**
Number of samples with taxon: 935 , without taxon: 25

Map type
○ pie chart ● dominant taxon

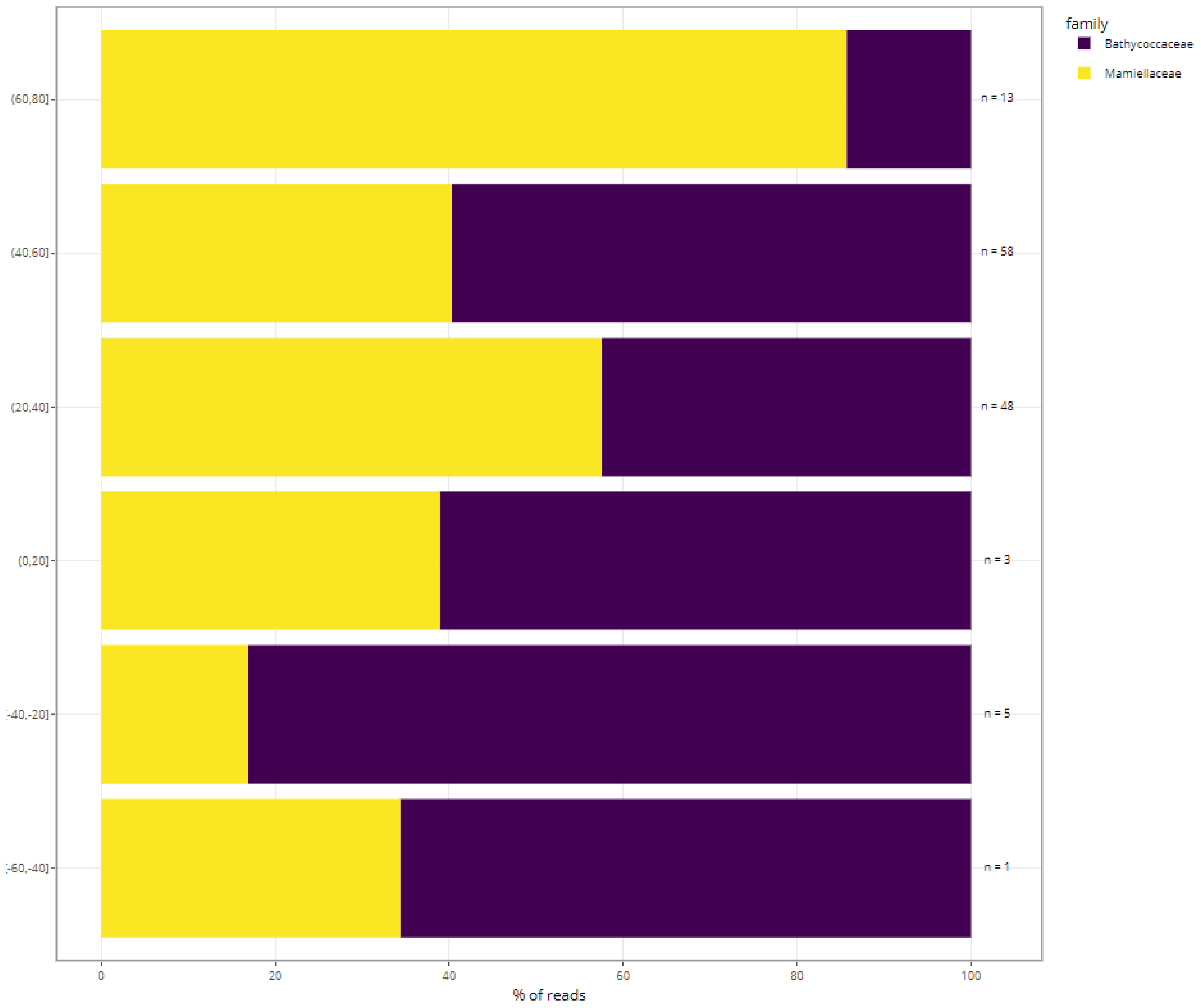
Change circle scale - % max: 40 %

MetaPR2 - Taxonomy

Nine levels:

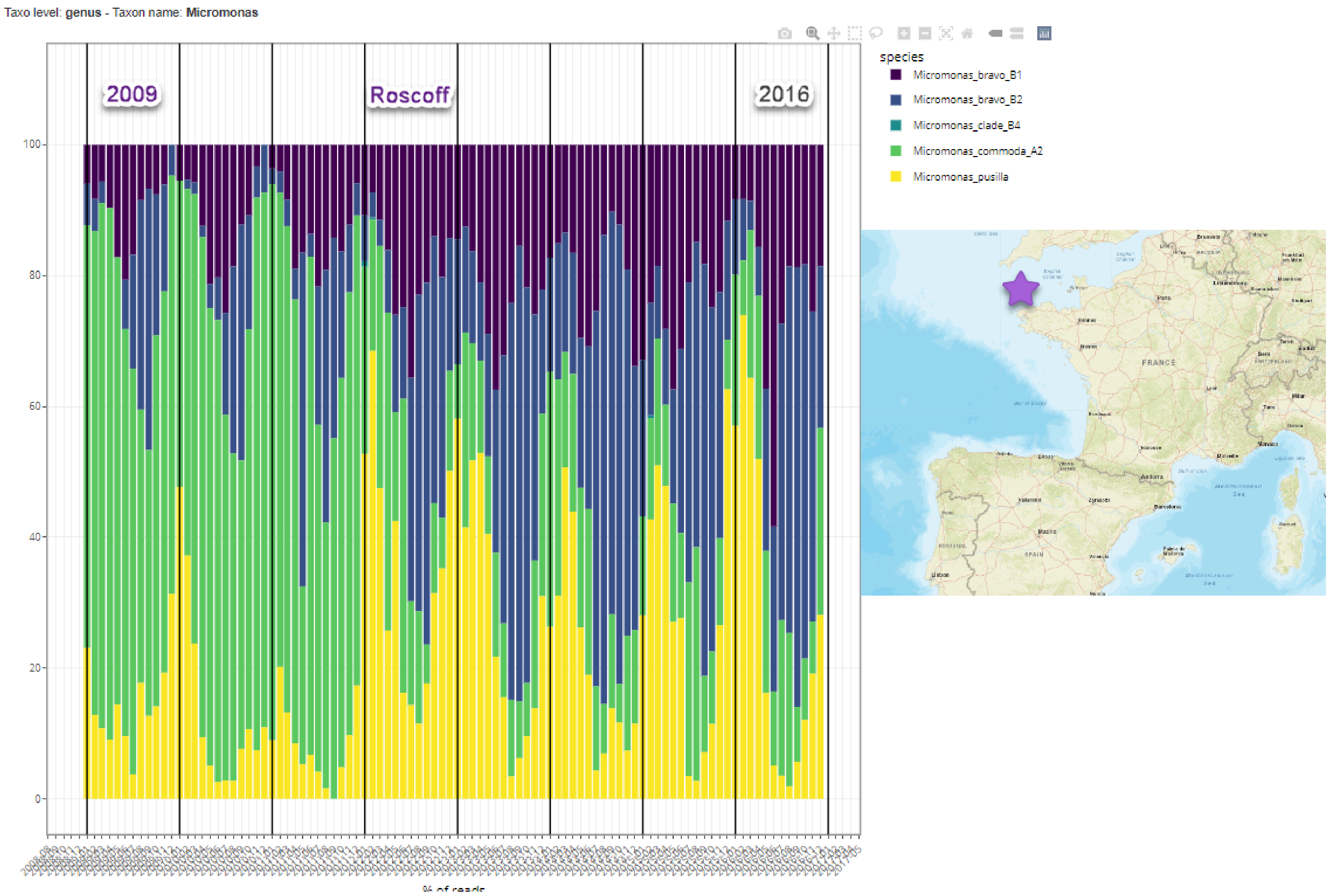
- Domain: Eukaryota
- Supergroup: Archaeplastida
- Division: Chlorophyta
- Subdivision: Chlorophyta_X
- Class: Mamiellophyceae
- Order: Mamielliales
- Family: Bathycoccaceae
- Genus: *Bathycococcus*
- Species: *B. prasinos*

Barplots - Latitude

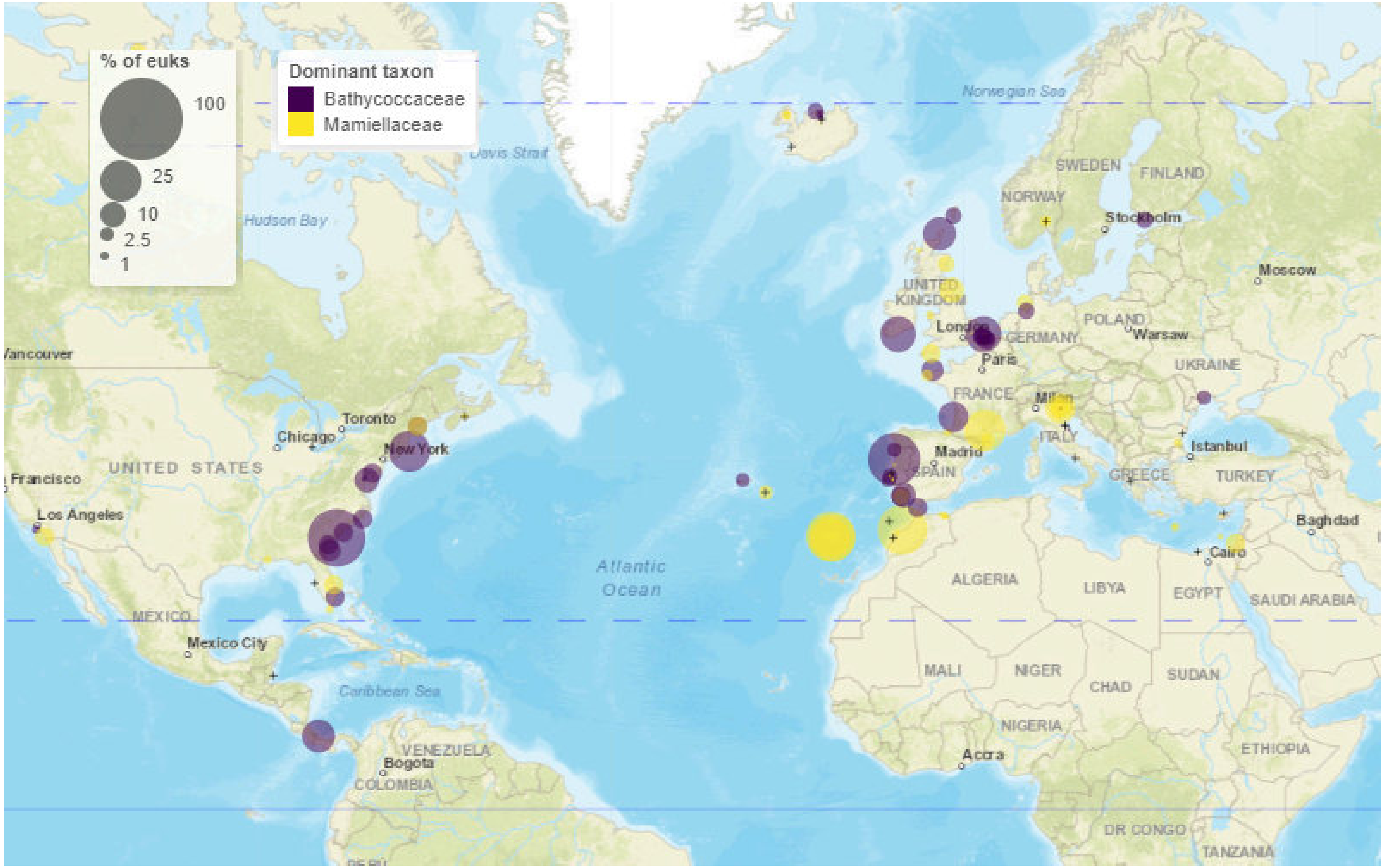


Barplots - Time series

Taxo level: **genus** - Taxon name: **Micromonas**



Maps - Dominant



Maps - Pie charts

Diversity

Number of samples: **960**

Diversity Measure

☑ Chao1 ☑ Shannon ☑ Simpson

X axis

● latitude ○ ecosystem ○ substrate ○ depth_level ○ depth ○ fraction_name ○ DNA_RNA ○ temperature ○ salinity

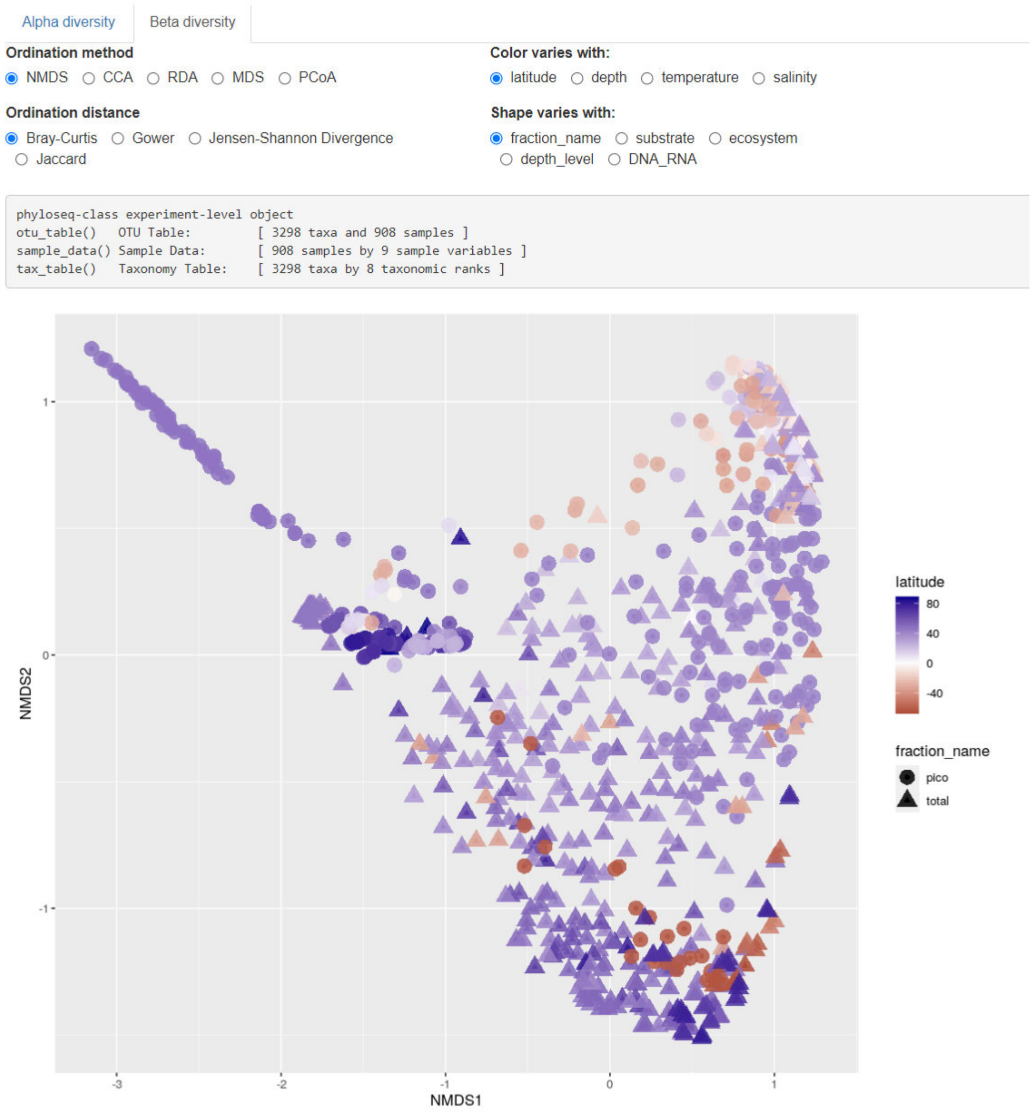
Color

○ latitude ● depth ○ temperature ○ salinity

Shape

● fraction_name ○ substrate ○ ecosystem ○ depth_level ○ DNA_RNA

```
phyloseq-class experiment-level object
otu_table()   OTU Table:         [ 3298 taxa and 908 samples ]
sample_data() Sample Data:       [ 908 samples by 9 sample variables ]
tax_table()   Taxonomy Table:    [ 3298 taxa by 8 taxonomic ranks ]
```



Ordination method

● NMDS ○ CCA ○ RDA ○ MDS ○ PCoA

Ordination distance

● Bray-Curtis ○ Gower ○ Jensen-Shannon Divergence ○ Jaccard

Color varies with:

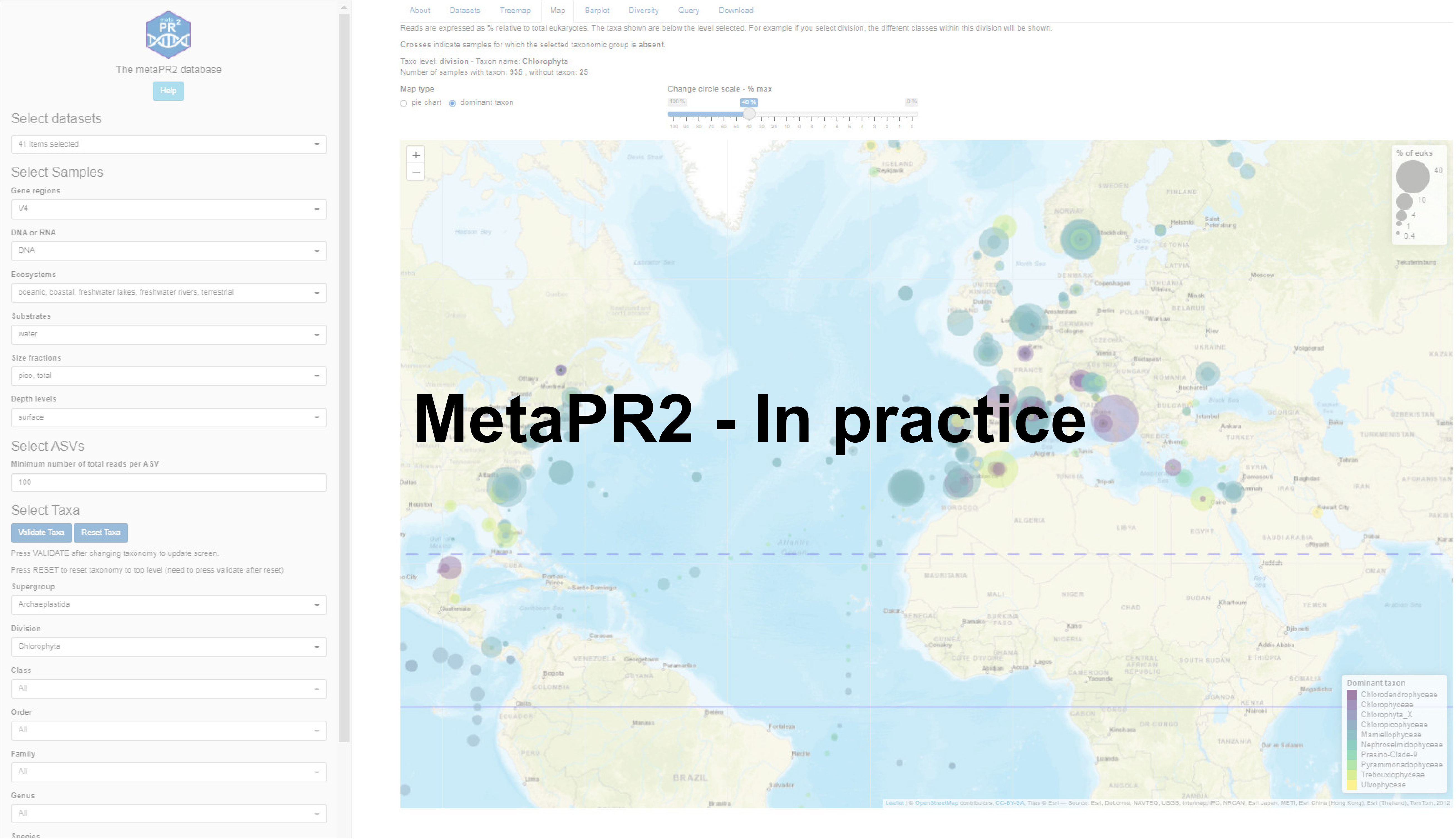
● latitude ○ depth ○ temperature ○ salinity

Shape varies with:

● fraction_name ○ substrate ○ ecosystem ○ depth_level ○ DNA_RNA

```
phyloseq-class experiment-level object
otu_table()   OTU Table:         [ 3298 taxa and 908 samples ]
sample_data() Sample Data:       [ 908 samples by 9 sample variables ]
tax_table()   Taxonomy Table:    [ 3298 taxa by 8 taxonomic ranks ]
```

MetaPR2 - In practice



The metaPR2 database

Select datasets

41 items selected

Select Samples

Gene regions: V4

DNA or RNA: DNA

Ecosystems: oceanic, coastal, freshwater lakes, freshwater rivers, terrestrial

Substrates: water

Size fractions: pico, total

Depth levels: surface

Select ASVs

Minimum number of total reads per ASV: 100

Select Taxa

Press VALIDATE after changing taxonomy to update screen.

Press RESET to reset taxonomy to top level (need to press validate after reset)

Supergroup: Archaeplastida

Division: Chlorophyta

Class: All

Order: All

Family: All

Genus: All

Reads are expressed as % relative to total eukaryotes. The taxa shown are below the level selected. For example if you select division, the different classes within this division will be shown.

Crosses indicate samples for which the selected taxonomic group is absent.

Taxo level: **division** - Taxon name: **Chlorophyta**
Number of samples with taxon: **935** , without taxon: **25**

Map type: ○ pie chart ● dominant taxon

Change circle scale - % max: 40 %

Help and Samples

Help

- Read in detail

Sample table

- dataset_name
- paper (can be useful to read)
- number of samples
- number of ASVs
- number of reads per sample (coverage)

Sample selection

- Major datasets: OSD, Tara, Malaspina
- By habitat: oceanic, coastal etc…
    - Start by “marine global V4”
    - Extend to other habitats/datasets
- V4 vs V9
- DNA vs. RNA
- Ecosystems
- Sustrate: water, ice, soil…
- Size fractions: total, pico…
- Depth level: surface, euphotic…
- Minimum ASV: will filter out rare ASVs (e.g. 1000)
- Selection can be saved (yaml file)

Taxonomy

- Can select several taxa within one level
- Press validate every time you need to refresh
- Can exclude taxa to remove fungi, metazoa…
- Can save taxonomy and reload taxonomy (yaml file)

Treemaps, Maps and Barplots

Treemaps

- Left panel: abundance (number of reads)
  - Reads are “normalized” to 100
- Right panel: diversity (number of ASVs)

Maps

- Read information at top
  - Taxo level
  - Number of samples with/without taxa
- Crosses where taxa absent
- Map types
  - Dominant
  - Pie chart
- Circle scale
  - Moving right increases size

Barplots

- taxonomy vs. function
- variables to use (but this depends on samples selected !)
  - fraction name
  - ecosystem
  - substrate
  - depth level
  - DNA_RNA
  - latitude
  - temperature
  - salinity
  - year, month, day for time series

Diversity

- Hit “Compute…” after refreshing taxonomy
- Time proportional to N samples and taxa
- Information about
  - Number of samples
  - Number of taxa (ASVs)

Alpha diversity

- X: Chao1, Shannon, Simpson (compare)
- Discretize continuous Y
- Change Y (see barplots)
- Change shape
- Change color

Beta diversity

- Ordination method (difference ?)
- Ordination distance (Bray, Jaccard…)
- Change color and shape

Download

- Download
    - datasets (csv)
    - samples (csv)
    - asv list with taxonomy (csv)
    - asv sequences (FASTA)

You can process these data with R (e.g. dplyr and ggplot2)