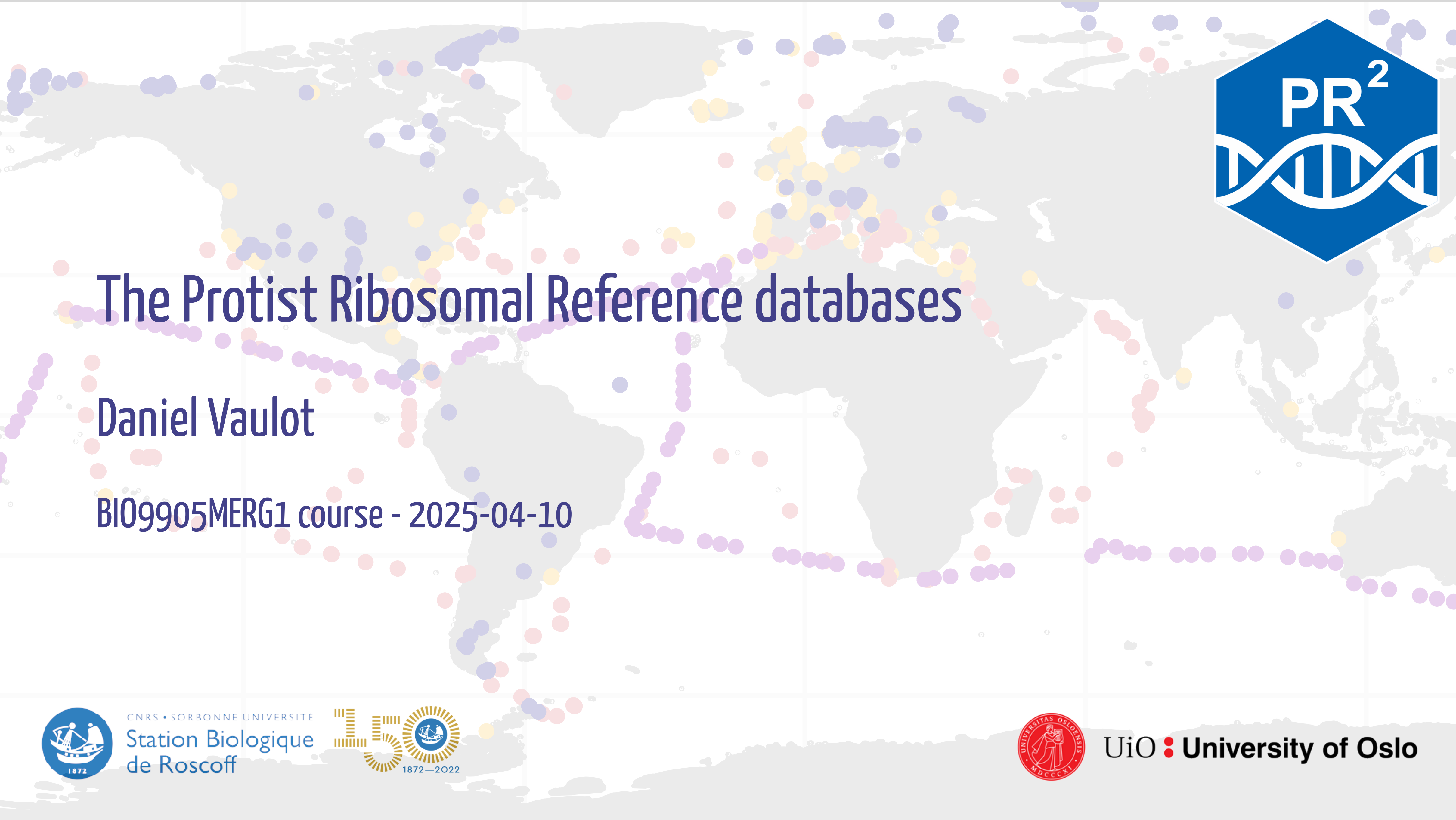


The Protist Ribosomal Reference databases

Daniel Vaulot

BIO9905MERG1 course - 2025-04-10

CNRS • SORBONNE UNIVERSITÉ
Station Biologique
de Roscoff

1872—2022

UiO : University of Oslo

Roscoff


Irlande
Limerick
Cork
ANGLETERRE
PAYS DE GALLES
Cambridge
Oxford
Londres
Bristol
Cardiff
Brighton
Southampton
Plymouth
Manche (mer)
Mer Celtique
Guernesey
Jersey
Brest
Rennes
Nantes
Tours
La Rochelle
Paris
France
Amsterdam
La Haye
Rotterdam
Anvers
Bruxelles
Lille
Belgique
Genève

La Station Biologique de Roscoff

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

1872—2022

Ecologie of Marine Plankton team

- Scientists: 14
- Staff: ~ 40
- Research themes:
  - ☐ Viruses
  - ☐ Bacteria
  - ☐ Cyanobacteria
  - ☐ Symbioses
  - ☐ Parasitism

CNRS • SORBONNE UNIVERSITÉ
Station Biologique de Roscoff

ECOLOGY OF MARINE PLANKTON (ECOMAP)

PERMANENT STAFF

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

Aurélie CHAMBOUVET
CRCN CNRS
Diversity and Ecology of metazoan parasitic protists

Laurence GARCZAREK
DR1 CNRS
Genetic and functional diversity of marine cyanobacteria

Benjamin BAILLEUL
CRCN CNRS
Photophysiology of phytoplankton
49% ECOMAP
51% IBPC

Juliana BERNARDES
MC SU
Machine learning in marine ecology

Colomban DE VARGAS
DR1 CNRS
& FR GOSEE
Plankton system biology

Laure GUILLOU
DR1 CNRS
DU UMR
symbiosis and parasitism

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

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

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

Nathalie SIMON
MC SU
Phytoplankton biodiversity and temporal dynamics

Christophe SIX
MC SU
Adaptation of phytoplankton to temperature

Daniel VAULOT
DRCE CNRS
Emeritus
Diversity and ecology of eukaryotic picoplankton

Norico YAMADA
CRCN CNRS
Evolutionary developmental processes of plastids in "dinotom" dinoflagellates

Estelle BIGEARD
IE CNRS
Marine eukaryotic parasites & viruses
Lab Management (LabCollector)

Florence LE GALL
IEHC CNRS
Marine plankton Lab management
Permanent Training

Camille POIRIER
IE CNRS
Flow cytometry & Lab Management
90% ECOMAP
10% RECYF

Charlotte BERTHELIER
IE CNRS
Bioanalyses of marine plankton

Nicolas HENRY
IR FR2424
Plankton eco-

Morgane RATIN
IR CNRS
Mol. biology & genetics of cyanobacteria
Security & safety

Sarah ROMAC
IE CNRS
Molecular ecology of marine protists
Lab management

Fabienne RIGA
IR

POST-DOCS, PhD STUDENTS & TEMPORARY STAFF

Emile FAURE
Post-Doc
Metagenomics of marine picocyanobacteria

Valeria JIMENEZ
Post-Doc
Dinoflagellate-diatom symbiosis

Natalia LLOPIS MONFERRER
Post-Doc
Silification in Rhizaria

Francesco MATTEI
Post-Doc
Interactions between diatoms and bacteria

Morgan SMITS
Post-Doc
Chemical interactions between host and parasite

Elisa CHAILLER
PhD student
Role of pathobiome on parasitic infections

Emmanuelle JAOUEN
PhD student
Marine viruses

Perrine KERGOAT
PhD student
Niche adaptation of marine cyanobacteria

Heliaz LE BAYON
PhD student
Molecular ecology of shellfish parasites

Morgane GUILLAM
IE SU
Bougainville project
Biological engineer

Erwan LEGEAY
IE SU
Environmental Genomics

Julie LEPETIT
AI
Lipidomics of dinoflagellates

Pauline NOGARET
AI
Marine viruses

ROSCOFF CULTURE COLLECTION

Ian PROBERT
IRHC SU
FR2424
RCC Manager

Priscillia GOURVIL
IR CNRS
FR2424
Curator of the RCC

Charles BACHY
IR CNRS
FR2424
RCC R&D&I manager

Martin GACHENOT
AI SU
FR2424
RCC Strain maintenance & Flow cyto. platform (RECYF)

Sarah GARRIC
IE SU
FR2424
Characterization of cryptophytes

Michele GREGO
AI SU
FR2424
RCC
Cryo-preservation

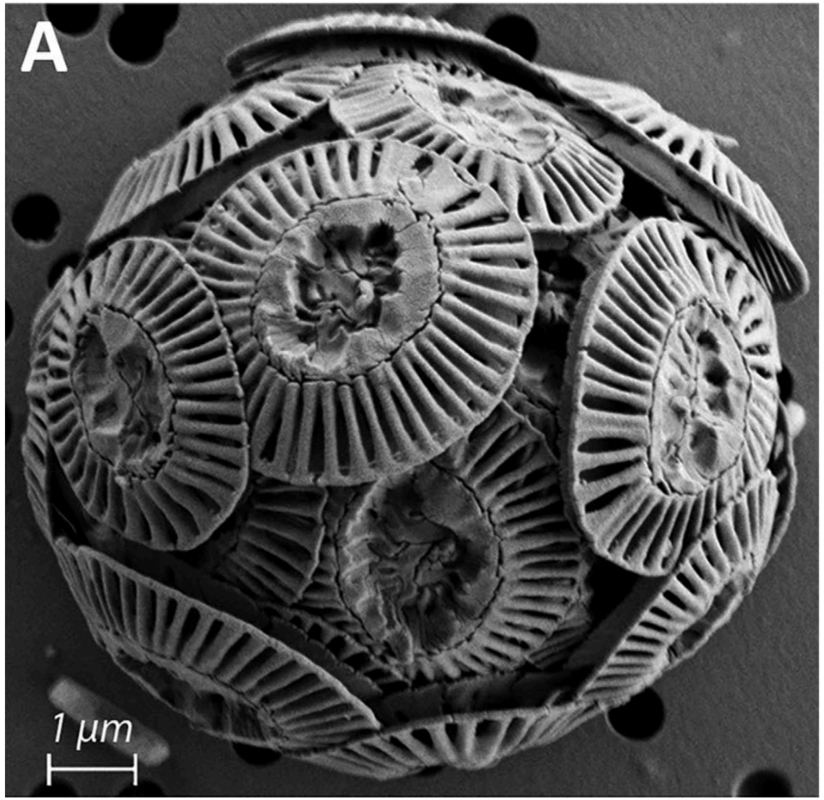
Katell HERVEOU
AI SU
FR2424
Strain maintenance & ATLASEA

Julie JOANNIC
AI SU
FR2424
Molecular biology of phytoplankton

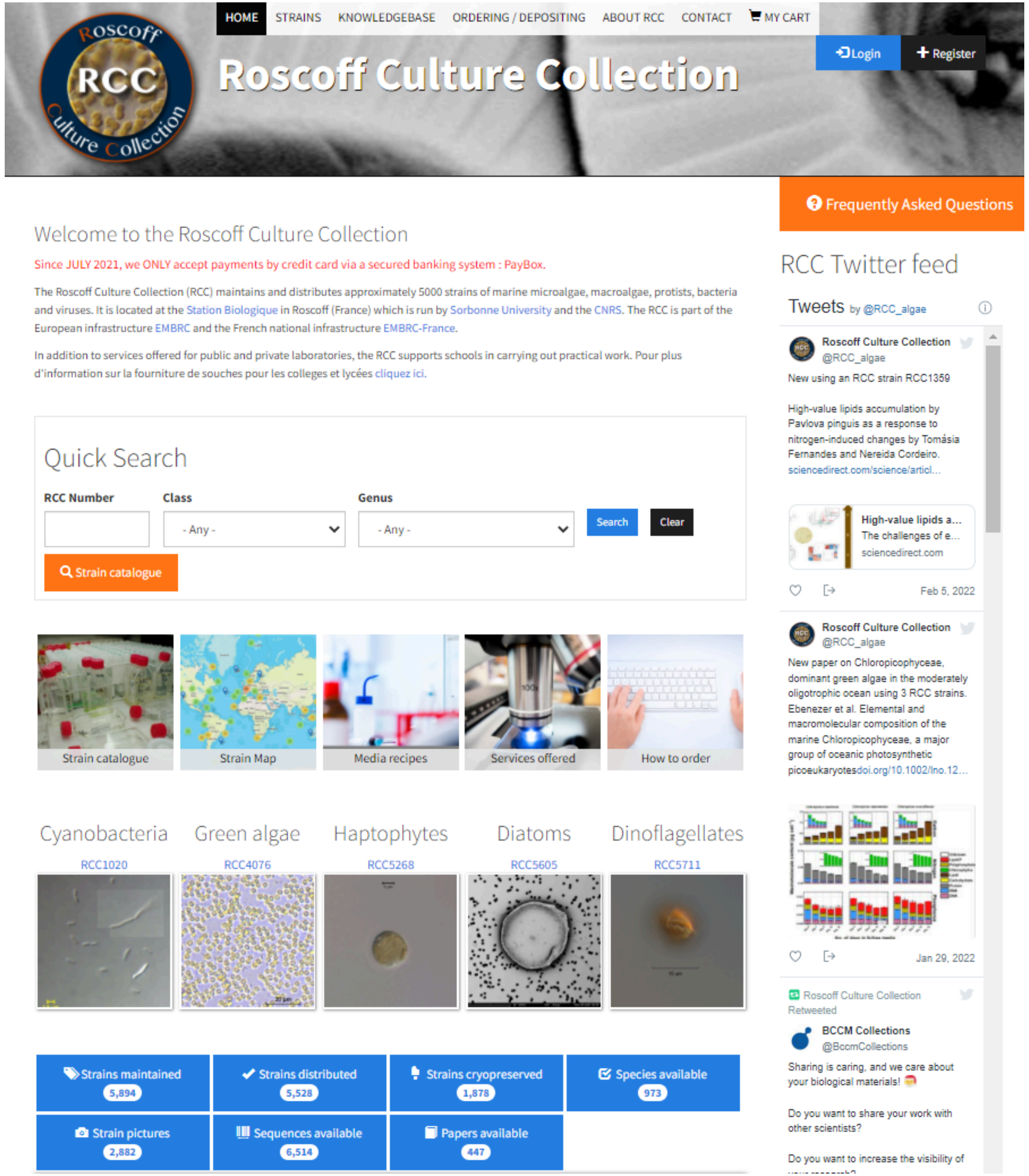
Marie WALDE
IR SU
FR2424
Advanced imaging of marine plankton

Roscoff Culture Collection

- Largest marine algal collection
- 9,500 strains distributed



https://roscoff-culture-collection.org/



Welcome to the Roscoff Culture Collection

Since JULY 2021, we ONLY accept payments by credit card via a secured banking system : PayBox.

The Roscoff Culture Collection (RCC) maintains and distributes approximately 5000 strains of marine microalgae, macroalgae, protists, bacteria and viruses. It is located at the Station Biologique in Roscoff (France) which is run by Sorbonne University and the CNRS. The RCC is part of the European infrastructure EMBRC and the French national infrastructure EMBRC-France.

In addition to services offered for public and private laboratories, the RCC supports schools in carrying out practical work. Pour plus d'information sur la fourniture de souches pour les colleges et lycées cliquez ici.

Quick Search

RCC Number:
Class: - Any -
Genus: - Any -

Strains maintained	Strains distributed	Strains cryopreserved	Species available
5,894	5,528	1,878	973

Strain pictures	Sequences available	Papers available
2,882	6,514	447

RCC Twitter feed

Tweets by @RCC_algae

Roscoff Culture Collection @RCC_algae

New using an RCC strain RCC1359

High-value lipids accumulation by Pavlova pinguis as a response to nitrogen-induced changes by Tomásia Fernandes and Nereida Cordeiro. sciencedirect.com/science/articl...

High-value lipids a...
The challenges of e...
sciencedirect.com

Feb 5, 2022

Roscoff Culture Collection @RCC_algae

New paper on Chloropicophyceae, dominant green algae in the moderately oligotrophic ocean using 3 RCC strains. Ebenezer et al. Elemental and macromolecular composition of the marine Chloropicophyceae, a major group of oceanic photosynthetic picoeukaryotesdoi.org/10.1002/lno.12...

Jan 29, 2022

Roscoff Culture Collection Retweeted

BCCM Collections @BccmCollections

Sharing is caring, and we care about your biological materials! 🎅

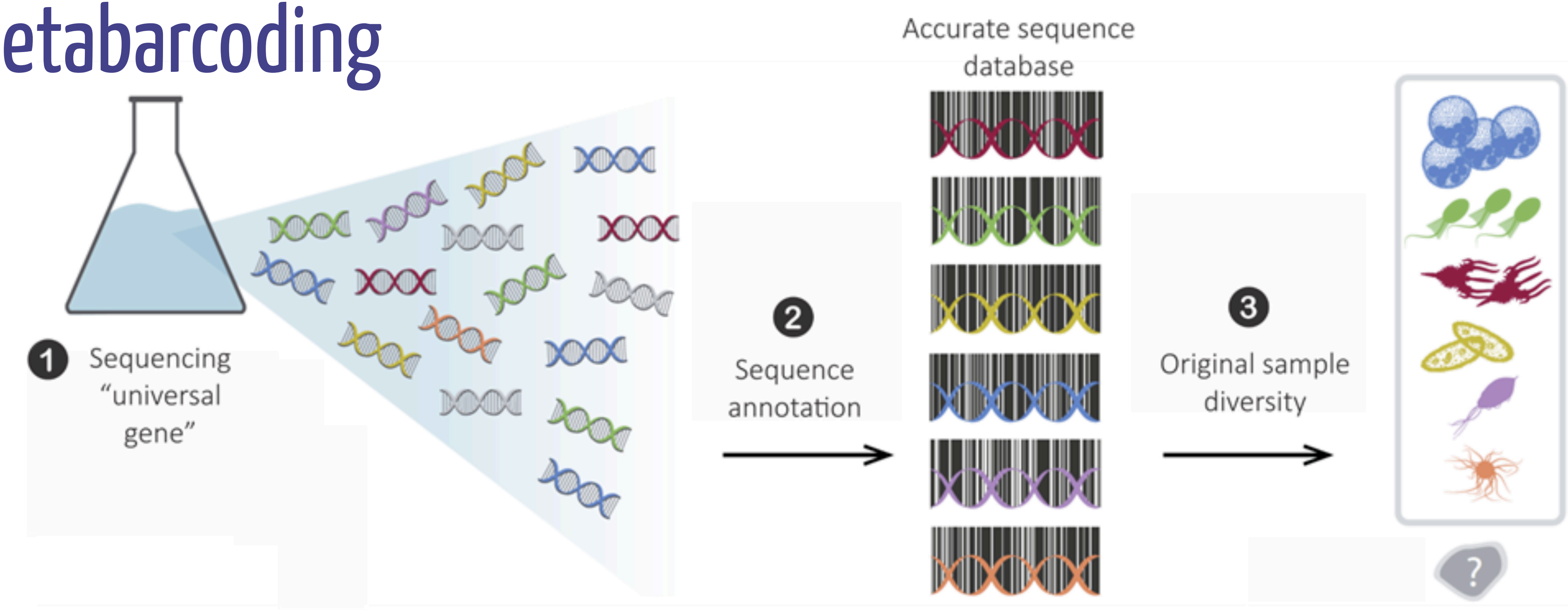
Do you want to share your work with other scientists?

Do you want to increase the visibility of

Outline

- The PR^2 ecosystem
- Protist Ribosomal Reference (PR^2) database
- PR^2 primers
- metaPR^2

Metabarcoding



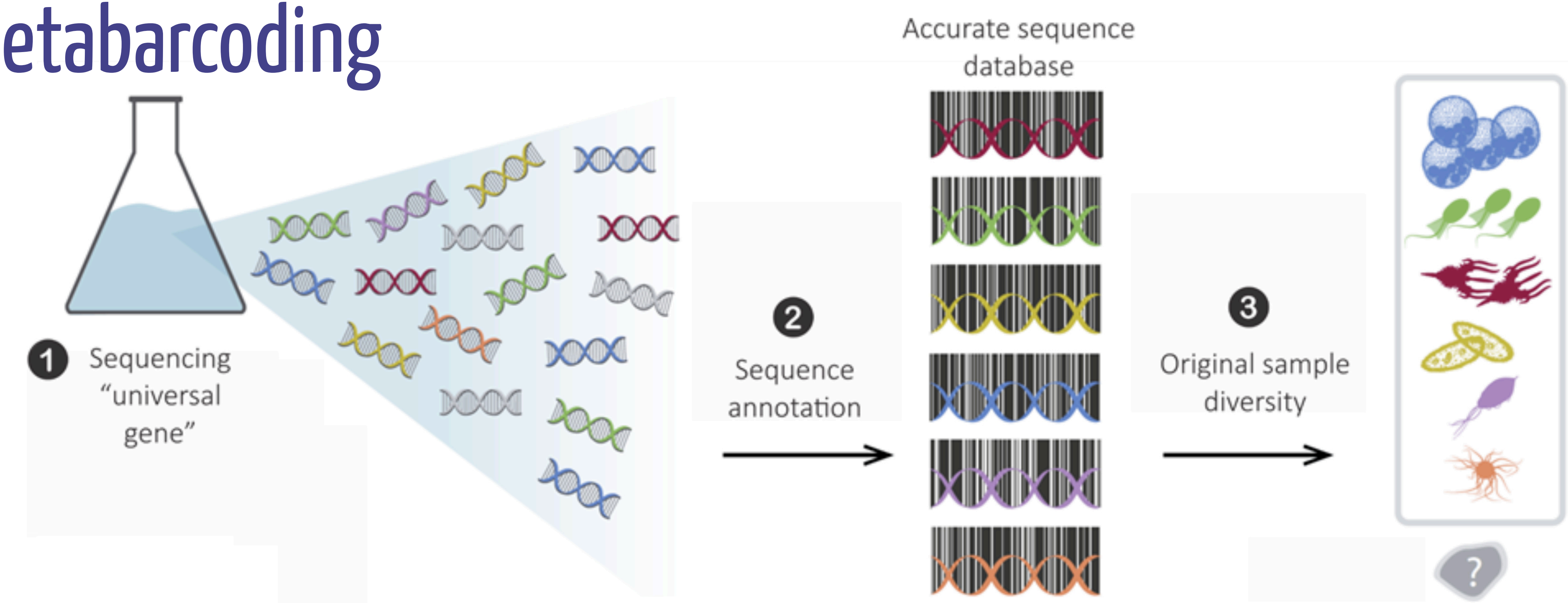
Accurate sequence database

① Sequencing "universal gene"

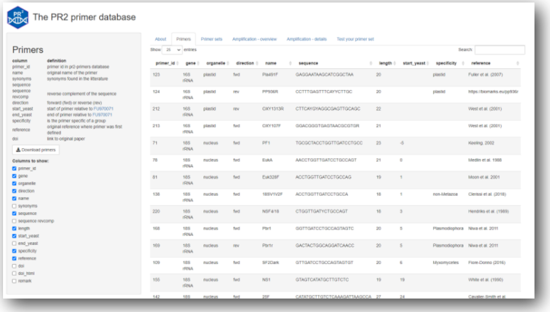
② Sequence annotation

③ Original sample diversity

Metabarcoding



pr2-primers


The PR2 primer database

Primers

pr2-database


PR² database

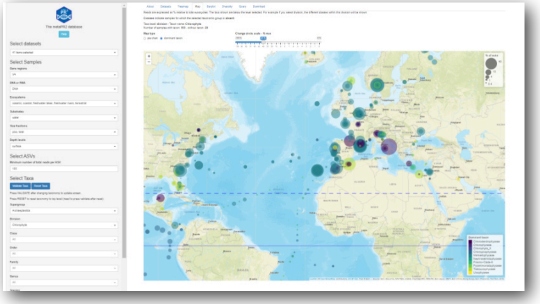
A reference 18S rRNA sequence database

- expert curated taxonomy
- metadata such as geo-localisation
- use to annotate metabarcodes

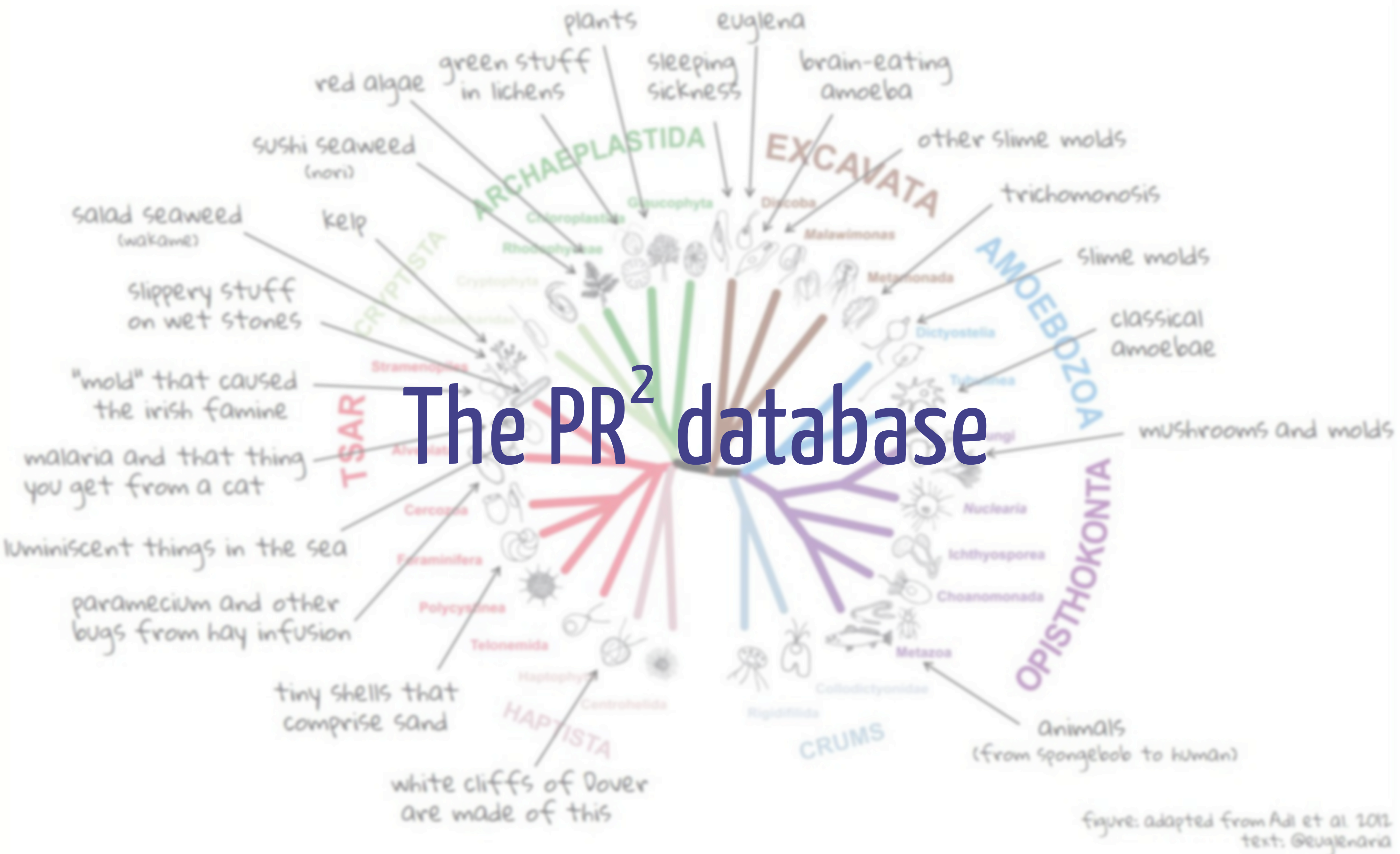
Download latest version of PR²

About PR2

metapr2



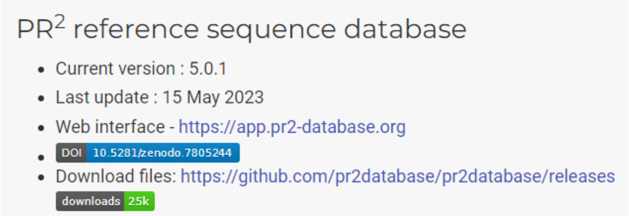
The PR^2 database



PR² reference database

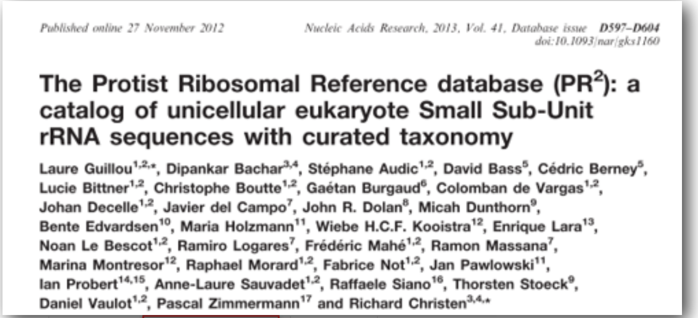
- Created in 2010 (Biomarks project)
- 240 199 annotated sequences
  - ☐ **nuclear 18S rRNA**
  - ☐ plastid 16S rRNA (PhytoRef)
  - ☐ bacteria and archaea 16S rRNA
  - ☐ **rRNA operons** (web interface)
- **Unified taxonomy** (9 ranks from domain to species)
- **Quality control** (e.g. > 500 bp., N < 20, no "NN")
- **Metadata** (e.g. coordinates, environment)
- Link to **other databases** (e.g. GBIF)
- Species functional annotation (e.g. ecological function)
- Web interface / Flat file

- Main use: **taxonomic annotation of metabarcodes**



PR² reference sequence database

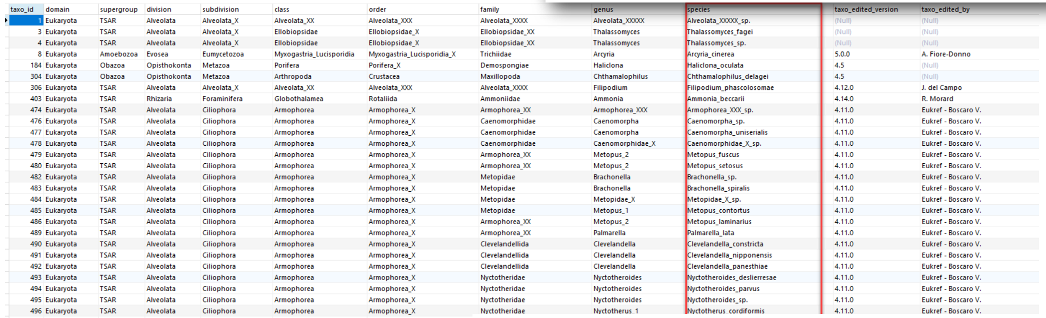
- Current version : 5.0.1
- Last update : 15 May 2023
- Web interface - https://app.pr2-database.org
- DOI 10.5281/zenodo.7805244
- Download files: https://github.com/pr2database/pr2database/releases



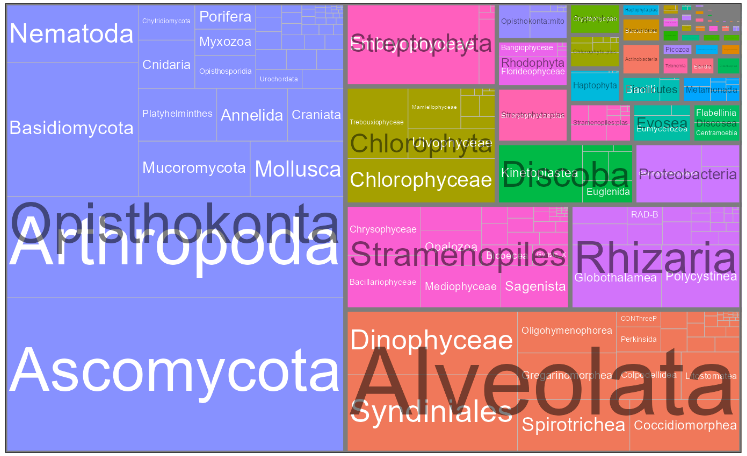
Published online 27 November 2012 Nucleic Acids Research, 2013, Vol. 41, Database issue **D597–D604**
doi:10.1093/nar/gks1160

The Protist Ribosomal Reference database (PR²): a catalog of unicellular eukaryote Small Sub-Unit rRNA sequences with curated taxonomy

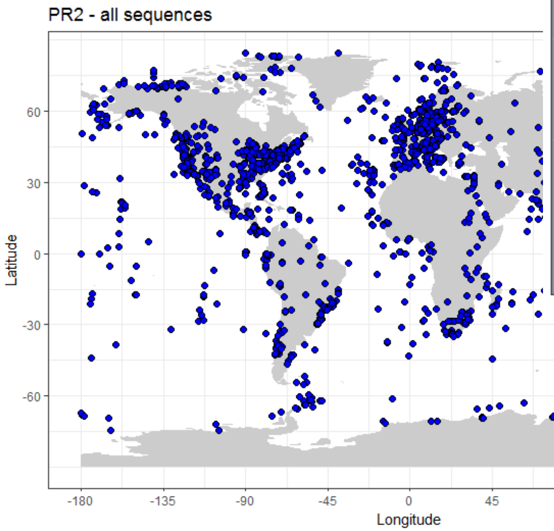
Laure Guillou[1,2,*], Dipankar Bachar[3,4], Stéphane Audic[1,2], David Bass[5], Cédric Berney[5], Lucie Bittner[1,2], Christophe Boutte[1,2], Gaétan Burgaud[6], Colomban de Vargas[1,2], Johan Decelle[1,2], Javier del Campo[7], John R. Dolan[8], Micah Dunthorn[9], Bente Edvardsen[10], Maria Holzmann[11], Wiebe H.C.F. Kooistra[12], Enrique Lara[13], Noan Le Bescot[1,2], Ramiro Logares[7], Frédéric Mahé[1,2], Ramon Massana[7], Marina Montresor[12], Raphael Morard[1,2], Fabrice Not[1,2], Jan Pawlowski[11], Ian Probert[14,15], Anne-Laure Sauvadet[1,2], Raffaele Siano[16], Thorsten Stoeck[9], Daniel Vaulot[1,2], Pascal Zimmermann[17] and Richard Christen[3,4,*]



taxo_id	domain	supergroup	division	subdivision	class	order	family	genus	species	taxo_edited_version	taxo_edited_by
1	Eukaryota	TSAR	Alveolata	Alveolata_X	Alveolata_XX	Alveolata_XXX	Alveolata_XXXX	Alveolata_XXXXX	Alveolata_XXXXX_sp.	(Null)	(Null)
3	Eukaryota	TSAR	Alveolata	Alveolata_X	Ellobiopsidae	Ellobiopsidae_X	Ellobiopsidae_XX	Thalassomyces	Thalassomyces_fagei	(Null)	(Null)
4	Eukaryota	TSAR	Alveolata	Alveolata_X	Ellobiopsidae	Ellobiopsidae_X	Ellobiopsidae_XX	Thalassomyces	Thalassomyces_sp.	(Null)	(Null)
8	Eukaryota	Amoebozoa	Evosea	Eumycetozoa	Myxogastria_Lucisporidia	Myxogastria_Lucisporidia_X	Trichiidae	Arcyria	Arcyria_cinerea	5.0.0	A. Fiore-Donno
184	Eukaryota	Obazoa	Opisthokonta	Metazoa	Porifera	Porifera_X	Demospongiae	Haliclona	Haliclona_oculata	4.5	(Null)
304	Eukaryota	Obazoa	Opisthokonta	Metazoa	Arthropoda	Crustacea	Maxillopoda	Chthamalophilus	Chthamalophilus_delagei	4.5	(Null)
306	Eukaryota	TSAR	Alveolata	Alveolata_X	Alveolata_XX	Alveolata_XXX	Alveolata_XXXX	Filipodium	Filipodium_phascolosomae	4.12.0	J. del Campo
403	Eukaryota	TSAR	Rhizaria	Foraminifera	Globothalamea	Rotaliida	Ammoniidae	Ammonia	Ammonia_beccarii	4.14.0	R. Morard
474	Eukaryota	TSAR	Alveolata	Ciliophora	Armophorea	Armophorea_X	Armophorea_XX	Armophorea_XXX	Armophorea_XXX_sp.	4.11.0	Eukref - Boscaro V.
476	Eukaryota	TSAR	Alveolata	Ciliophora	Armophorea	Armophorea_X	Caenomorphidae	Caenomorpha	Caenomorpha_sp.	4.11.0	Eukref - Boscaro V.
477	Eukaryota	TSAR	Alveolata	Ciliophora	Armophorea	Armophorea_X	Caenomorphidae	Caenomorpha	Caenomorpha_uniserialis	4.11.0	Eukref - Boscaro V.
478	Eukaryota	TSAR	Alveolata	Ciliophora	Armophorea	Armophorea_X	Caenomorphidae	Caenomorphidae_X	Caenomorphidae_X_sp.	4.11.0	Eukref - Boscaro V.
479	Eukaryota	TSAR	Alveolata	Ciliophora	Armophorea	Armophorea_X	Armophorea_XX	Metopus_2	Metopus_fuscus	4.11.0	Eukref - Boscaro V.
480	Eukaryota	TSAR	Alveolata	Ciliophora	Armophorea	Armophorea_X	Armophorea_XX	Metopus_2	Metopus_setosus	4.11.0	Eukref - Boscaro V.
482	Eukaryota	TSAR	Alveolata	Ciliophora	Armophorea	Armophorea_X	Metopidae	Brachonella	Brachonella_sp.	4.11.0	Eukref - Boscaro V.
483	Eukaryota	TSAR	Alveolata	Ciliophora	Armophorea	Armophorea_X	Metopidae	Brachonella	Brachonella_spiralis	4.11.0	Eukref - Boscaro V.
484	Eukaryota	TSAR	Alveolata	Ciliophora	Armophorea	Armophorea_X	Metopidae	Metopidae_X	Metopidae_X_sp.	4.11.0	Eukref - Boscaro V.
485	Eukaryota	TSAR	Alveolata	Ciliophora	Armophorea	Armophorea_X	Metopidae	Metopus_1	Metopus_contortus	4.11.0	Eukref - Boscaro V.
486	Eukaryota	TSAR	Alveolata	Ciliophora	Armophorea	Armophorea_X	Armophorea_XX	Metopus_2	Metopus_laminarius	4.11.0	Eukref - Boscaro V.
489	Eukaryota	TSAR	Alveolata	Ciliophora	Armophorea	Armophorea_X	Armophorea_XX	Palmarella	Palmarella_lata	4.11.0	Eukref - Boscaro V.
490	Eukaryota	TSAR	Alveolata	Ciliophora	Armophorea	Armophorea_X	Clevelandellida	Clevelandella	Clevelandella_constricta	4.11.0	Eukref - Boscaro V.
491	Eukaryota	TSAR	Alveolata	Ciliophora	Armophorea	Armophorea_X	Clevelandellida	Clevelandella	Clevelandella_nipponensis	4.11.0	Eukref - Boscaro V.
492	Eukaryota	TSAR	Alveolata	Ciliophora	Armophorea	Armophorea_X	Clevelandellida	Clevelandella	Clevelandella_panesthiae	4.11.0	Eukref - Boscaro V.
493	Eukaryota	TSAR	Alveolata	Ciliophora	Armophorea	Armophorea_X	Nyctotheridae	Nyctotheroides	Nyctotheroides_deslierresae	4.11.0	Eukref - Boscaro V.
494	Eukaryota	TSAR	Alveolata	Ciliophora	Armophorea	Armophorea_X	Nyctotheridae	Nyctotheroides	Nyctotheroides_parvus	4.11.0	Eukref - Boscaro V.
495	Eukaryota	TSAR	Alveolata	Ciliophora	Armophorea	Armophorea_X	Nyctotheridae	Nyctotheroides	Nyctotheroides_sp.	4.11.0	Eukref - Boscaro V.
496	Eukaryota	TSAR	Alveolata	Ciliophora	Armophorea	Armophorea_X	Nyctotheridae	Nyctotherus_1	Nyctotherus_cordiformis	4.11.0	Eukref - Boscaro V.
497	Eukaryota	TSAR	Alveolata	Ciliophora	Armophorea	Armophorea_X					
499	Eukaryota	TSAR	Alveolata	Ciliophora	Armophorea	Armophorea_X					
503	Eukaryota	TSAR	Alveolata	Ciliophora	CONThreeP	CONThreeP_X					
514	Eukaryota	TSAR	Alveolata	Ciliophora	Colpodea	Colpodea_X					
515	Eukaryota	TSAR	Alveolata	Ciliophora	Colpodea	Colpodea_X					
517	Eukaryota	TSAR	Alveolata	Ciliophora	Colpodea	Colpodea_X					
519	Eukaryota	TSAR	Alveolata	Ciliophora	Colpodea	Colpodea_X					

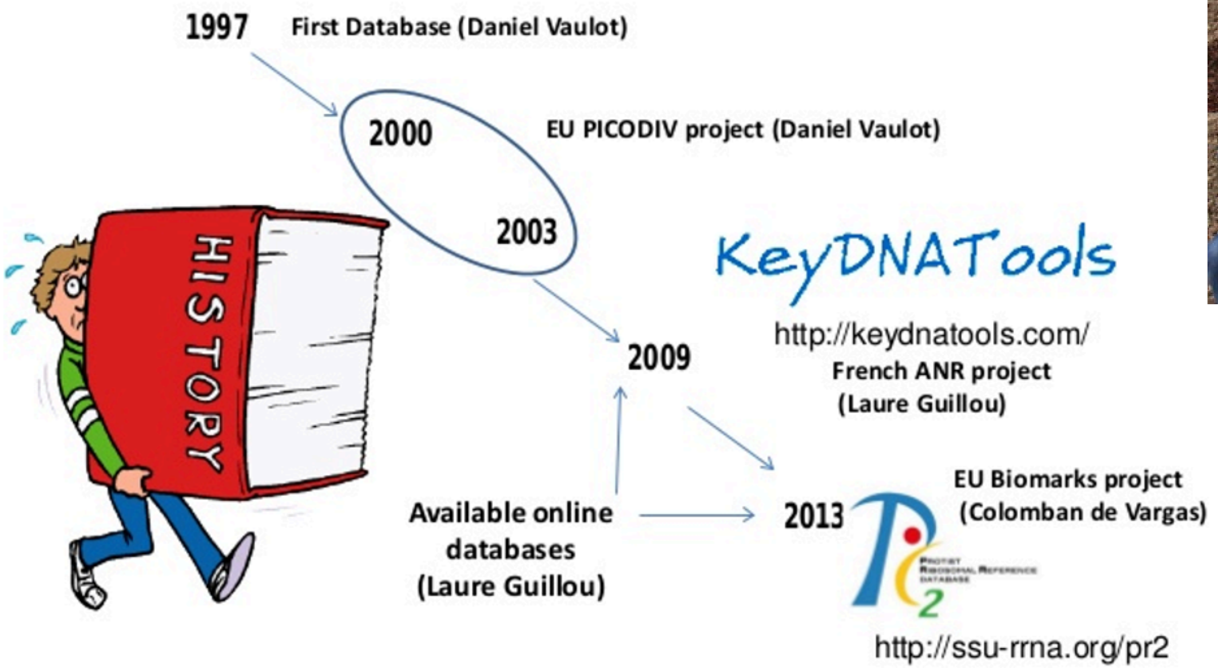


PR2 - all sequences



History

- Item
-



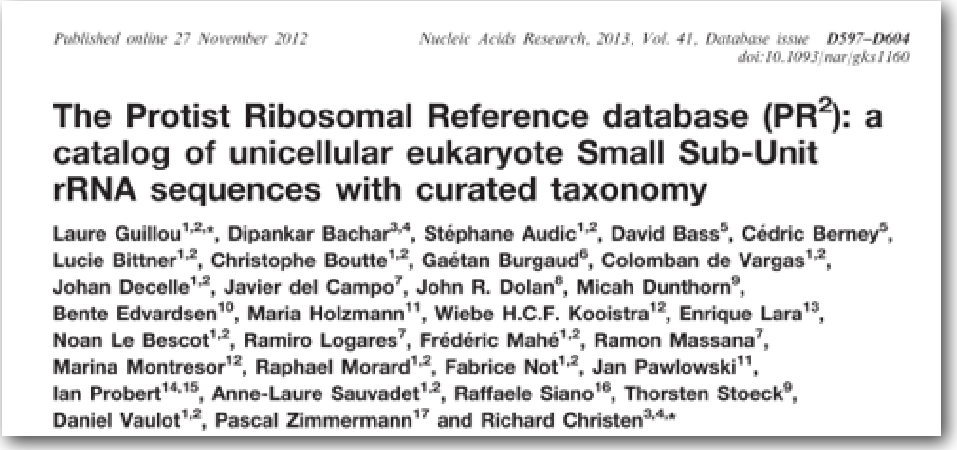
1997 First Database (Daniel Vaulot)

2000 – **2003** EU PICODIV project (Daniel Vaulot)

2009 KeyDNATools
http://keydnatools.com/
French ANR project
(Laure Guillou)

Available online databases (Laure Guillou)

2013 EU Biomarks project (Colomban de Vargas)
http://ssu-rrna.org/pr2



Published online 27 November 2012 *Nucleic Acids Research, 2013, Vol. 41, Database issue* **D597–D604**
doi:10.1093/nar/gks1160

The Protist Ribosomal Reference database (PR^2): a catalog of unicellular eukaryote Small Sub-Unit rRNA sequences with curated taxonomy

Laure Guillou[1,2,*], Dipankar Bachar[3,4], Stéphane Audic[1,2], David Bass[5], Cédric Berney[5], Lucie Bittner[1,2], Christophe Boutte[1,2], Gaétan Burgaud[6], Colomban de Vargas[1,2], Johan Decelle[1,2], Javier del Campo[7], John R. Dolan[8], Micah Dunthorn[9], Bente Edvardsen[10], Maria Holzmann[11], Wiebe H.C.F. Kooistra[12], Enrique Lara[13], Noan Le Bescot[1,2], Ramiro Logares[7], Frédéric Mahé[1,2], Ramon Massana[7], Marina Montresor[12], Raphael Morard[1,2], Fabrice Not[1,2], Jan Pawlowski[11], Ian Probert[14,15], Anne-Laure Sauvadet[1,2], Raffaele Siano[16], Thorsten Stoeck[9], Daniel Vaulot[1,2], Pascal Zimmermann[17] and Richard Christen[3,4,*]

Citation

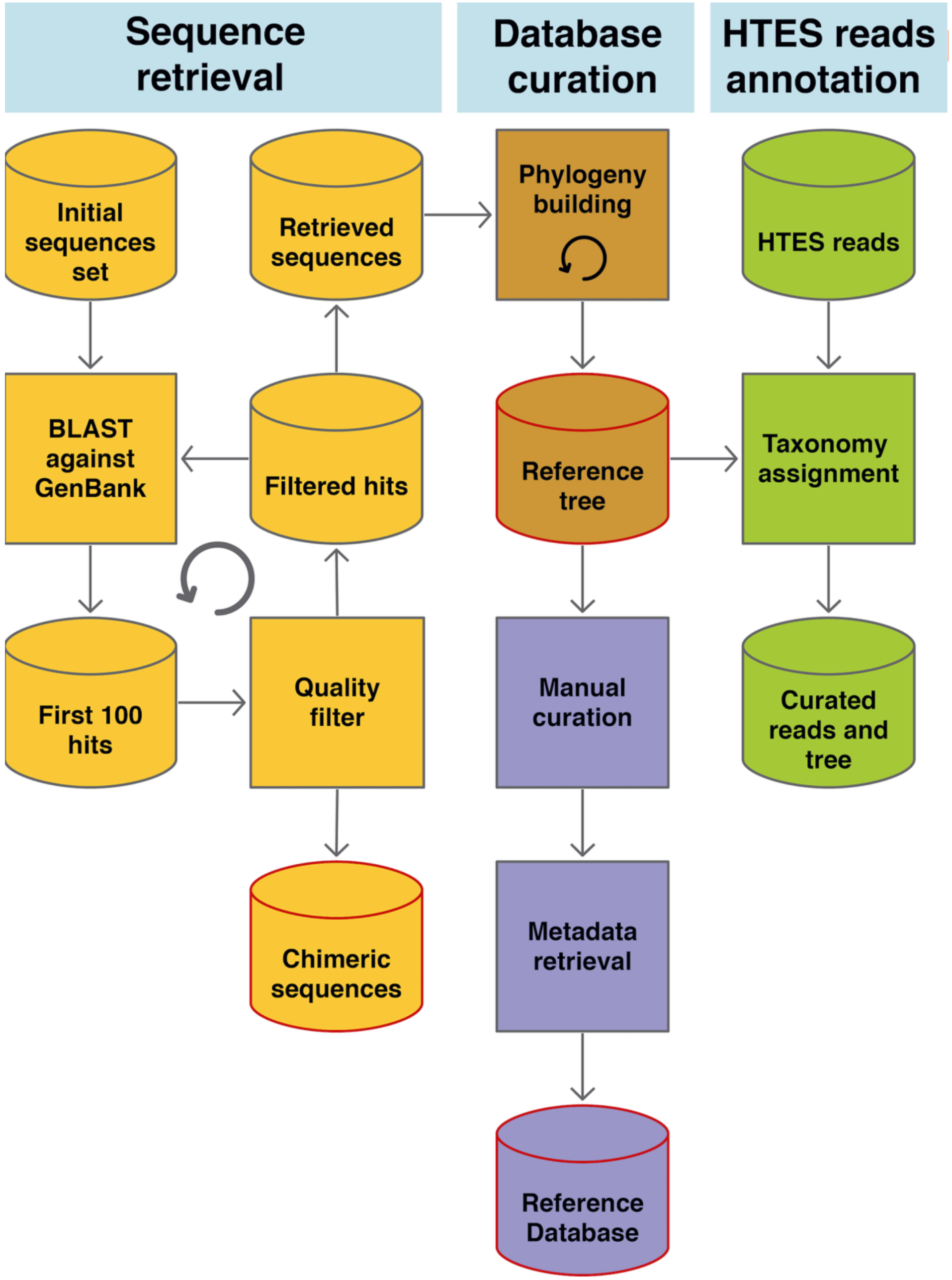
Annotation - Eukref

Workshop Roscoff 2018

Javier del Campo
Last updated on 2019-11-29 · eukref

Our third workshop took place at the **Station Biologique Roscoff** in **Roscoff**, France from **November 5 to 9 2018**. This effort brought together taxonomists with expertise in individual lineages that span the eukaryotic tree of life to curate reference 18S rDNA sequences of these lineages by incorporating knowledge of phylogenetic, morphological, and/or environmental contextual data. **This third edition was particularly (but not exclusively) focused on Diatoms, Dinoflagellates, and Green Algae**. During the workshop,

https://pr2-database.org/eukref/about/



Sequence retrieval

Initial sequences set
BLAST against GenBank
First 100 hits
Quality filter
Chimeric sequences
Filtered hits
Retrieved sequences

Database curation

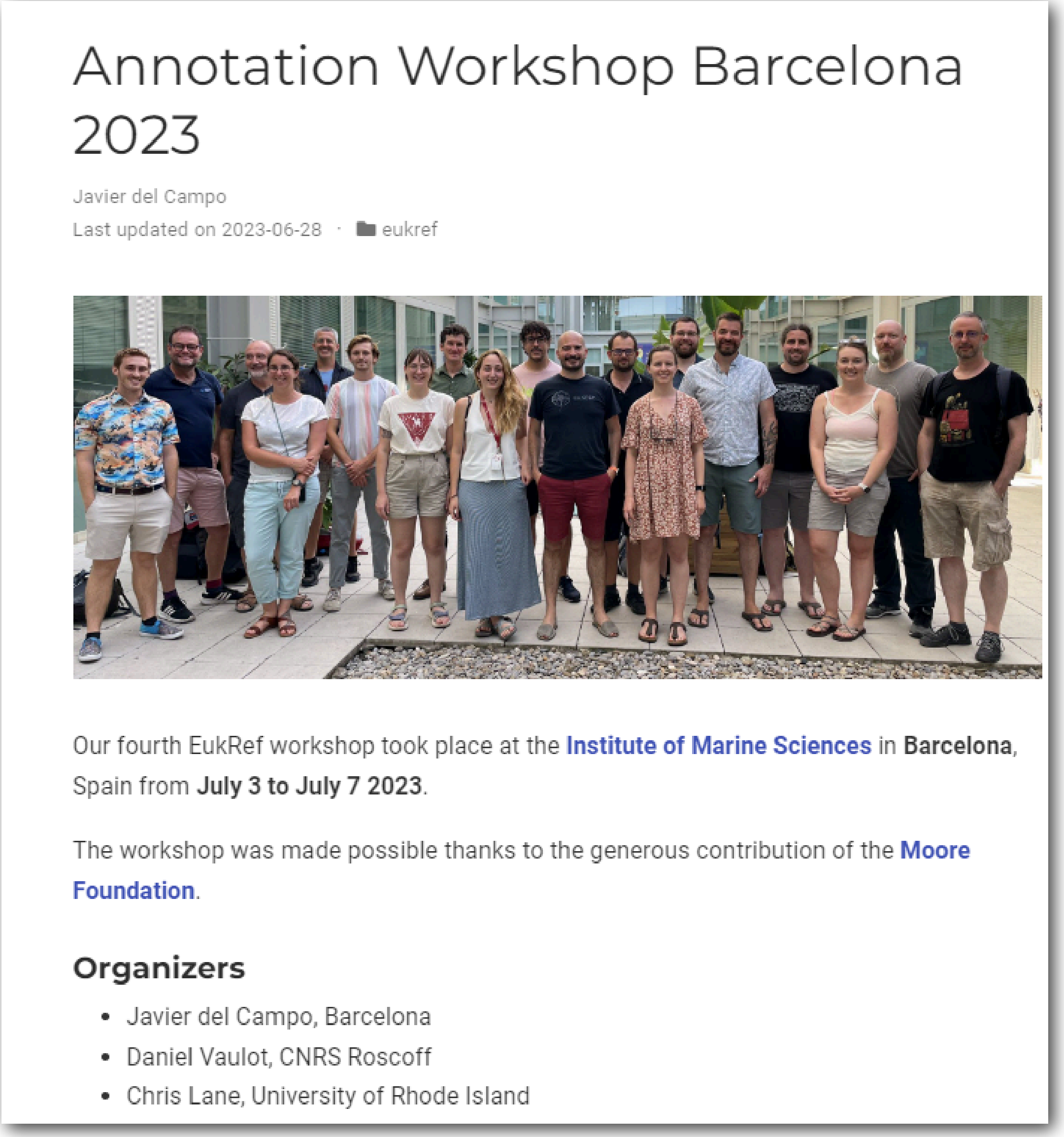
Phylogeny building
Reference tree
Manual curation
Metadata retrieval
Reference Database

HTES reads annotation

HTES reads
Taxonomy assignment
Curated reads and tree

Annotation - Eukref

- Simplified pipeline
- Adapt more closely to PR2 structure
- Groups annotated
  - Picozoa
  - Chloropicophyceae
  - Red algae
  - Metazoa
  - Dinoflagellates
  - Basal fungi
  - …



Annotation Workshop Barcelona 2023

Javier del Campo
Last updated on 2023-06-28 · eukref

Our fourth EukRef workshop took place at the **Institute of Marine Sciences** in **Barcelona**, Spain from **July 3 to July 7 2023**.

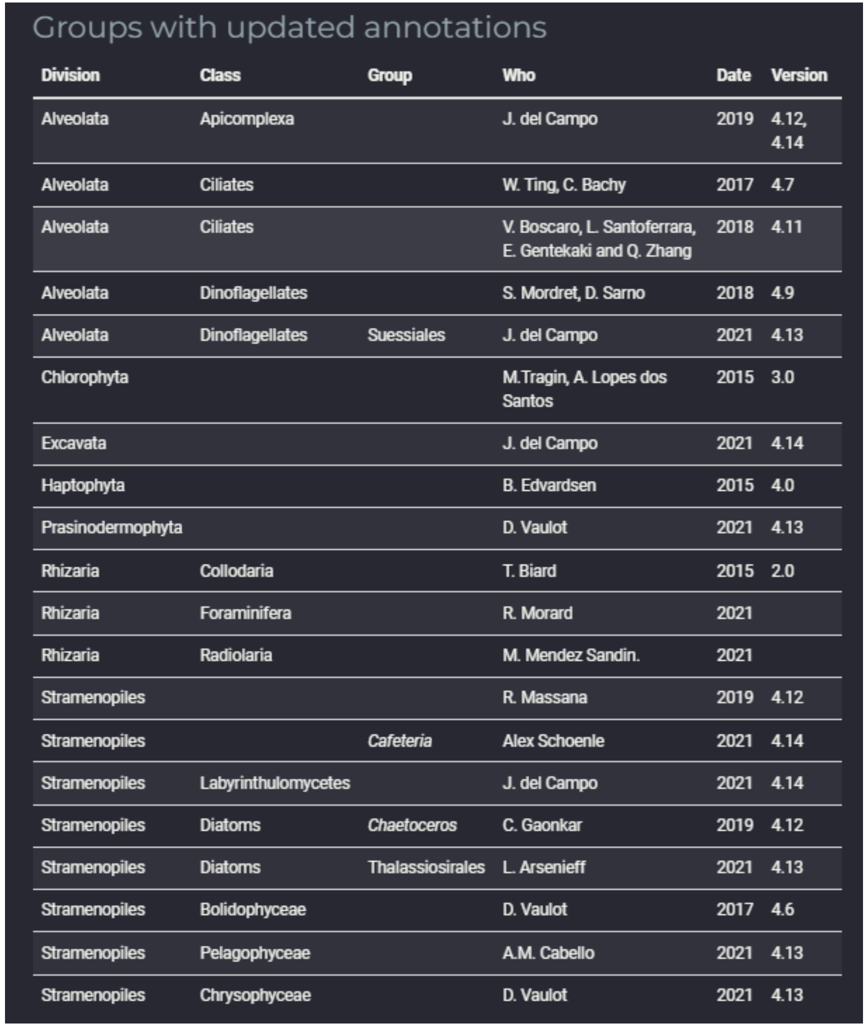
The workshop was made possible thanks to the generous contribution of the **Moore Foundation**.

Organizers

- Javier del Campo, Barcelona
- Daniel Vaulot, CNRS Roscoff
- Chris Lane, University of Rhode Island

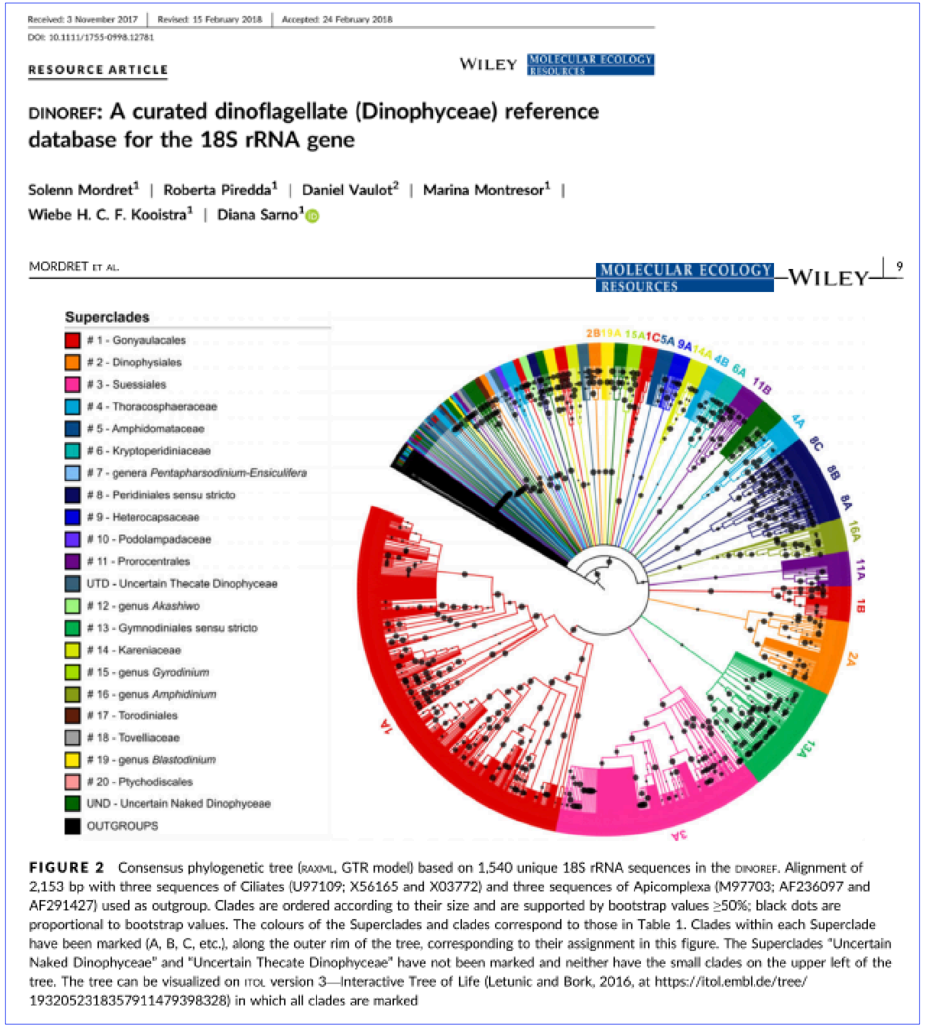
https://pr2-database.org/eukref/about/

Annotation - Contributions



Groups with updated annotations

Division	Class	Group	Who	Date	Version
Alveolata	Apicomplexa		J. del Campo	2019	4.12, 4.14
Alveolata	Ciliates		W. Ting, C. Bachy	2017	4.7
Alveolata	Ciliates		V. Boscaro, L. Santoferrara, E. Gentekaki and Q. Zhang	2018	4.11
Alveolata	Dinoflagellates		S. Mordret, D. Sarno	2018	4.9
Alveolata	Dinoflagellates	Suessiales	J. del Campo	2021	4.13
Chlorophyta			M.Tragin, A. Lopes dos Santos	2015	3.0
Excavata			J. del Campo	2021	4.14
Haptophyta			B. Edvardsen	2015	4.0
Prasinodermophyta			D. Vaulot	2021	4.13
Rhizaria	Collodaria		T. Biard	2015	2.0
Rhizaria	Foraminifera		R. Morard	2021	
Rhizaria	Radiolaria		M. Mendez Sandin.	2021	
Stramenopiles			R. Massana	2019	4.12
Stramenopiles		*Cafeteria*	Alex Schoenle	2021	4.14
Stramenopiles	Labyrinthulomycetes		J. del Campo	2021	4.14
Stramenopiles	Diatoms	*Chaetoceros*	C. Gaonkar	2019	4.12
Stramenopiles	Diatoms	Thalassiosirales	L. Arsenieff	2021	4.13
Stramenopiles	Bolidophyceae		D. Vaulot	2017	4.6
Stramenopiles	Pelagophyceae		A.M. Cabello	2021	4.13
Stramenopiles	Chrysophyceae		D. Vaulot	2021	4.13



Received: 3 November 2017 | Revised: 15 February 2018 | Accepted: 24 February 2018

DOI: 10.1111/1755-0998.12781

RESOURCE ARTICLE

WILEY MOLECULAR ECOLOGY RESOURCES

DINOREF: A curated dinoflagellate (Dinophyceae) reference database for the 18S rRNA gene

Solenn Mordret[1] | Roberta Piredda[1] | Daniel Vaulot[2] | Marina Montresor[1] | Wiebe H. C. F. Kooistra[1] | Diana Sarno[1]

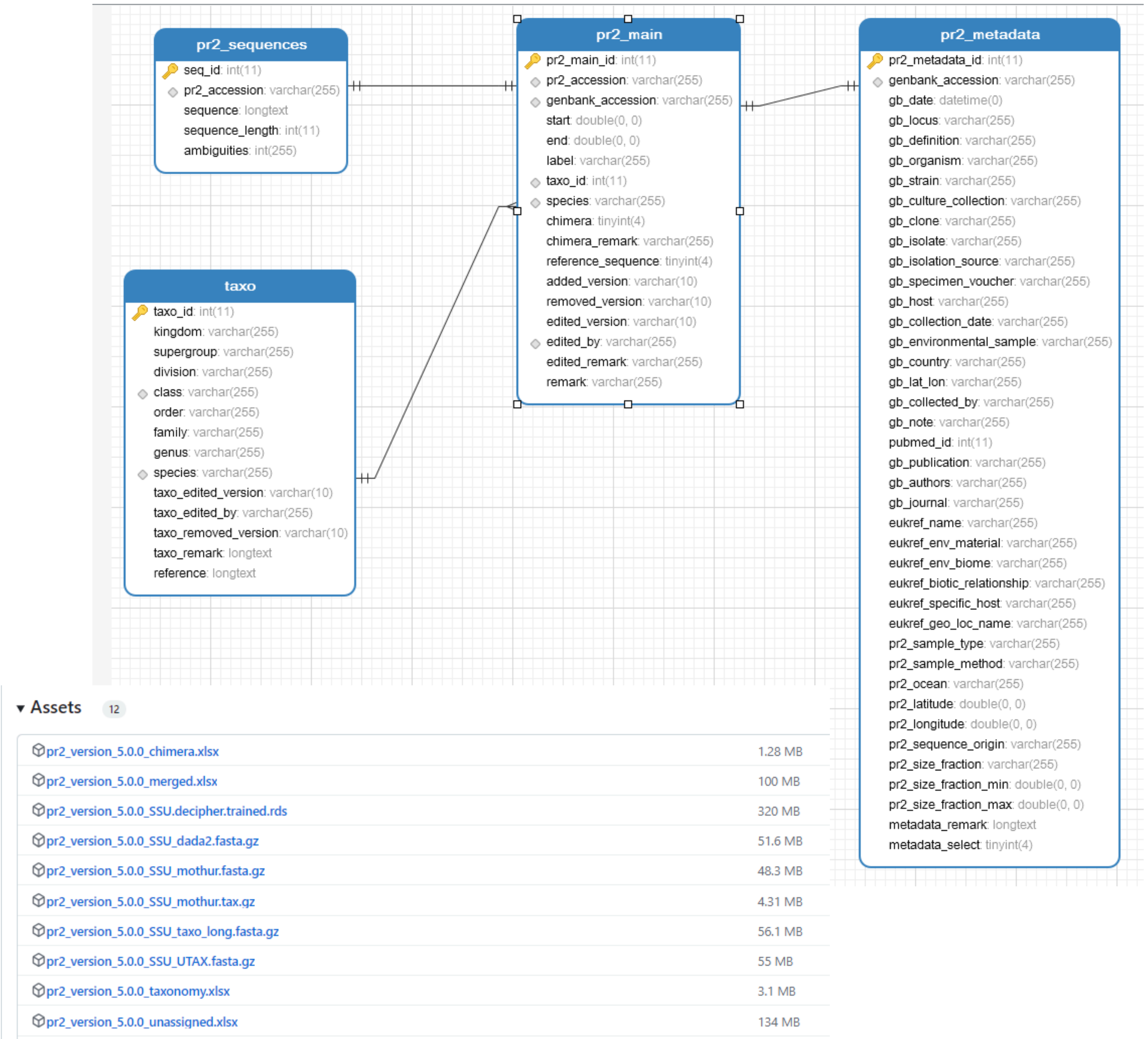
MORDRET ET AL. MOLECULAR ECOLOGY RESOURCES WILEY 9

FIGURE 2 Consensus phylogenetic tree (RAXML, GTR model) based on 1,540 unique 18S rRNA sequences in the DINOREF. Alignment of 2,153 bp with three sequences of Ciliates (U97109; X56165 and X03772) and three sequences of Apicomplexa (M97703; AF236097 and AF291427) used as outgroup. Clades are ordered according to their size and are supported by bootstrap values ≥50%; black dots are proportional to bootstrap values. The colours of the Superclades and clades correspond to those in Table 1. Clades within each Superclade have been marked (A, B, C, etc.), along the outer rim of the tree, corresponding to their assignment in this figure. The Superclades "Uncertain Naked Dinophyceae" and "Uncertain Thecate Dinophyceae" have not been marked and neither have the small clades on the upper left of the tree. The tree can be visualized on ITOL version 3—Interactive Tree of Life (Letunic and Bork, 2016, at https://itol.embl.de/tree/1932052318357911479398328) in which all clades are marked

https://pr2-database.org/documentation/pr2-taxonomic-groups/

Management

- **MySQL database**
- **R scripts** for:
  - importing
  - exporting
  - validating
- **Data provided as**
  - text files (for dada2, mothur)
  - fasta (phylogeny)
  - R package



pr2_sequences

seq_id: int(11)
pr2_accession: varchar(255)
sequence: longtext
sequence_length: int(11)
ambiguities: int(255)

pr2_main

pr2_main_id: int(11)
pr2_accession: varchar(255)
genbank_accession: varchar(255)
start: double(0, 0)
end: double(0, 0)
label: varchar(255)
taxo_id: int(11)
species: varchar(255)
chimera: tinyint(4)
chimera_remark: varchar(255)
reference_sequence: tinyint(4)
added_version: varchar(10)
removed_version: varchar(10)
edited_version: varchar(10)
edited_by: varchar(255)
edited_remark: varchar(255)
remark: varchar(255)

pr2_metadata

pr2_metadata_id: int(11)
genbank_accession: varchar(255)
gb_date: datetime(0)
gb_locus: varchar(255)
gb_definition: varchar(255)
gb_organism: varchar(255)
gb_strain: varchar(255)
gb_culture_collection: varchar(255)
gb_clone: varchar(255)
gb_isolate: varchar(255)
gb_isolation_source: varchar(255)
gb_specimen_voucher: varchar(255)
gb_host: varchar(255)
gb_collection_date: varchar(255)
gb_environmental_sample: varchar(255)
gb_country: varchar(255)
gb_lat_lon: varchar(255)
gb_collected_by: varchar(255)
gb_note: varchar(255)
pubmed_id: int(11)
gb_publication: varchar(255)
gb_authors: varchar(255)
gb_journal: varchar(255)
eukref_name: varchar(255)
eukref_env_material: varchar(255)
eukref_env_biome: varchar(255)
eukref_biotic_relationship: varchar(255)
eukref_specific_host: varchar(255)
eukref_geo_loc_name: varchar(255)
pr2_sample_type: varchar(255)
pr2_sample_method: varchar(255)
pr2_ocean: varchar(255)
pr2_latitude: double(0, 0)
pr2_longitude: double(0, 0)
pr2_sequence_origin: varchar(255)
pr2_size_fraction: varchar(255)
pr2_size_fraction_min: double(0, 0)
pr2_size_fraction_max: double(0, 0)
metadata_remark: longtext
metadata_select: tinyint(4)

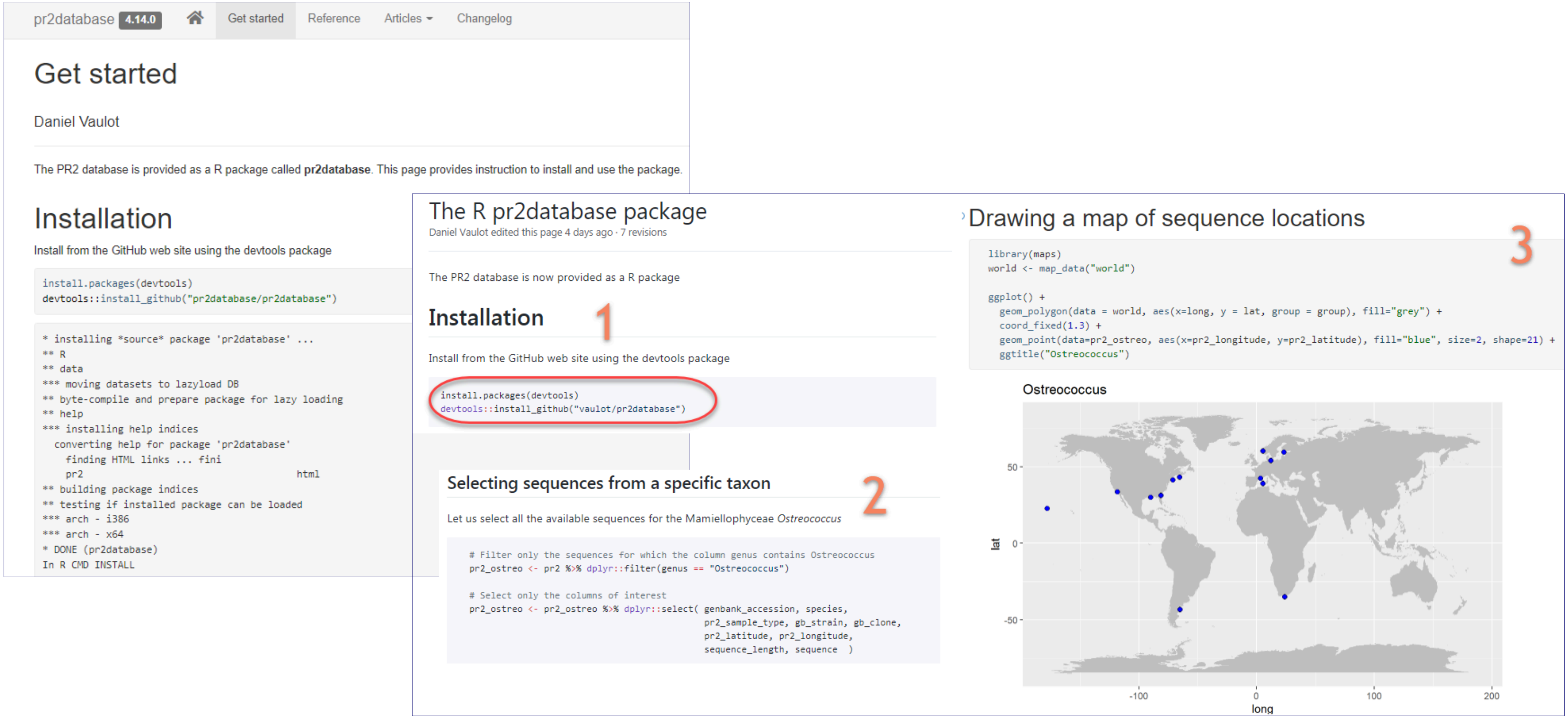
taxo

taxo_id: int(11)
kingdom: varchar(255)
supergroup: varchar(255)
division: varchar(255)
class: varchar(255)
order: varchar(255)
family: varchar(255)
genus: varchar(255)
species: varchar(255)
taxo_edited_version: varchar(10)
taxo_edited_by: varchar(255)
taxo_removed_version: varchar(10)
taxo_remark: longtext
reference: longtext

▾ Assets 12

pr2_version_5.0.0_chimera.xlsx	1.28 MB
pr2_version_5.0.0_merged.xlsx	100 MB
pr2_version_5.0.0_SSU.decipher.trained.rds	320 MB
pr2_version_5.0.0_SSU_dada2.fasta.gz	51.6 MB
pr2_version_5.0.0_SSU_mothur.fasta.gz	48.3 MB
pr2_version_5.0.0_SSU_mothur.tax.gz	4.31 MB
pr2_version_5.0.0_SSU_taxo_long.fasta.gz	56.1 MB
pr2_version_5.0.0_SSU_UTAX.fasta.gz	55 MB
pr2_version_5.0.0_taxonomy.xlsx	3.1 MB
pr2_version_5.0.0_unassigned.xlsx	134 MB

R package



Get started

Daniel Vaulot

The PR2 database is provided as a R package called **pr2database**. This page provides instruction to install and use the package.

Installation

Install from the GitHub web site using the devtools package

```
install.packages(devtools)
devtools::install_github("pr2database/pr2database")
```

```
* installing *source* package 'pr2database' ...
** R
** data
*** moving datasets to lazyload DB
** byte-compile and prepare package for lazy loading
** help
*** installing help indices
  converting help for package 'pr2database'
    finding HTML links ... fini
    pr2                                     html
** building package indices
** testing if installed package can be loaded
*** arch - i386
*** arch - x64
* DONE (pr2database)
In R CMD INSTALL
```

The R pr2database package

Daniel Vaulot edited this page 4 days ago · 7 revisions

The PR2 database is now provided as a R package

Installation

1

Install from the GitHub web site using the devtools package

```
install.packages(devtools)
devtools::install_github("vaulot/pr2database")
```

Selecting sequences from a specific taxon

2

Let us select all the available sequences for the Mamiellophyceae *Ostreococcus*

```
# Filter only the sequences for which the column genus contains Ostreococcus
pr2_ostreo <- pr2 %>% dplyr::filter(genus == "Ostreococcus")

# Select only the columns of interest
pr2_ostreo <- pr2_ostreo %>% dplyr::select( genbank_accession, species,
                                            pr2_sample_type, gb_strain, gb_clone,
                                            pr2_latitude, pr2_longitude,
                                            sequence_length, sequence  )
```

Drawing a map of sequence locations

3

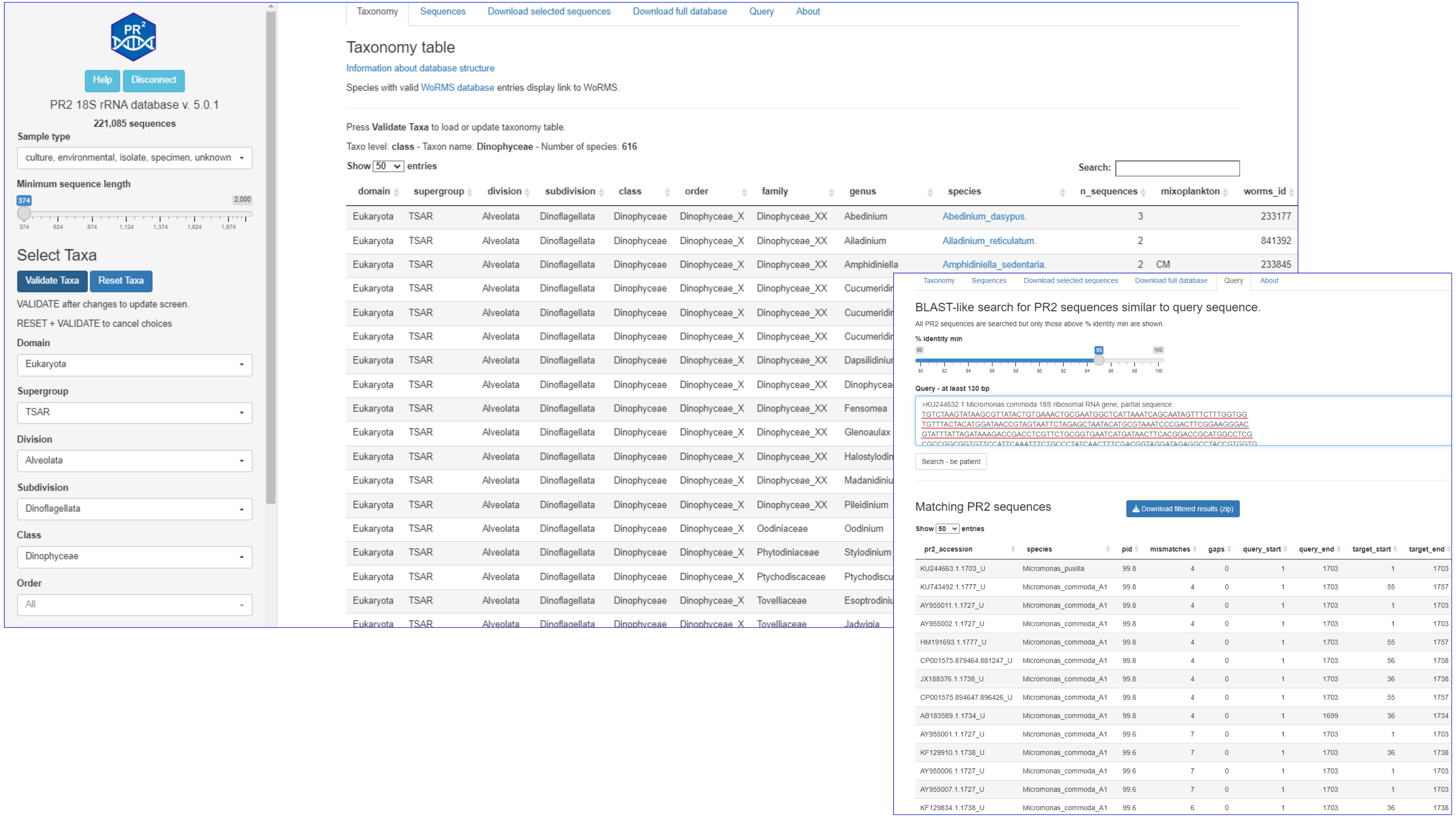
```
library(maps)
world <- map_data("world")

ggplot() +
  geom_polygon(data = world, aes(x=long, y = lat, group = group), fill="grey") +
  coord_fixed(1.3) +
  geom_point(data=pr2_ostreo, aes(x=pr2_longitude, y=pr2_latitude), fill="blue", size=2, shape=21) +
  ggtitle("Ostreococcus")
```

Web interface

https://app.pr2-database.org

- Built with R shiny
- Panels
  - ☐ Taxonomy
  - ☐ Sequences
  - ☐ Download
    - ☐ selected
    - ☐ full
  - ☐ Query



PR2 18S rRNA database v. 5.0.1

221,085 sequences

Sample type
culture, environmental, isolate, specimen, unknown

Minimum sequence length
374 – 2,000

Select Taxa

VALIDATE after changes to update screen.

RESET + VALIDATE to cancel choices

Domain
Eukaryota

Supergroup
TSAR

Division
Alveolata

Subdivision
Dinoflagellata

Class
Dinophyceae

Order
All

Taxonomy table

Information about database structure

Species with valid WoRMS database entries display link to WoRMS.

Press **Validate Taxa** to load or update taxonomy table.

Taxo level: **class** - Taxon name: **Dinophyceae** - Number of species: **616**

Show 50 **entries**

Search:

domain	supergroup	division	subdivision	class	order	family	genus	species	n_sequences	mixoplankton	worms_id
Eukaryota	TSAR	Alveolata	Dinoflagellata	Dinophyceae	Dinophyceae_X	Dinophyceae_XX	Abedinium	Abedinium_dasypus.	3		233177
Eukaryota	TSAR	Alveolata	Dinoflagellata	Dinophyceae	Dinophyceae_X	Dinophyceae_XX	Ailadinium	Ailadinium_reticulatum.	2		841392
Eukaryota	TSAR	Alveolata	Dinoflagellata	Dinophyceae	Dinophyceae_X	Dinophyceae_XX	Amphidiniella	Amphidiniella_sedentaria.	2	CM	233845
Eukaryota	TSAR	Alveolata	Dinoflagellata	Dinophyceae	Dinophyceae_X	Dinophyceae_XX	Cucumeridi				
Eukaryota	TSAR	Alveolata	Dinoflagellata	Dinophyceae	Dinophyceae_X	Dinophyceae_XX	Cucumeridi				
Eukaryota	TSAR	Alveolata	Dinoflagellata	Dinophyceae	Dinophyceae_X	Dinophyceae_XX	Cucumeridi				
Eukaryota	TSAR	Alveolata	Dinoflagellata	Dinophyceae	Dinophyceae_X	Dinophyceae_XX	Dapsilidiniu				
Eukaryota	TSAR	Alveolata	Dinoflagellata	Dinophyceae	Dinophyceae_X	Dinophyceae_XX	Dinophycea				
Eukaryota	TSAR	Alveolata	Dinoflagellata	Dinophyceae	Dinophyceae_X	Dinophyceae_XX	Fensomea				
Eukaryota	TSAR	Alveolata	Dinoflagellata	Dinophyceae	Dinophyceae_X	Dinophyceae_XX	Glenoaulax				
Eukaryota	TSAR	Alveolata	Dinoflagellata	Dinophyceae	Dinophyceae_X	Dinophyceae_XX	Halostylodin				
Eukaryota	TSAR	Alveolata	Dinoflagellata	Dinophyceae	Dinophyceae_X	Dinophyceae_XX	Madanidiniu				
Eukaryota	TSAR	Alveolata	Dinoflagellata	Dinophyceae	Dinophyceae_X	Dinophyceae_XX	Pileidinium				
Eukaryota	TSAR	Alveolata	Dinoflagellata	Dinophyceae	Dinophyceae_X	Oodiniaceae	Oodinium				
Eukaryota	TSAR	Alveolata	Dinoflagellata	Dinophyceae	Dinophyceae_X	Phytodiniaceae	Stylodinium				
Eukaryota	TSAR	Alveolata	Dinoflagellata	Dinophyceae	Dinophyceae_X	Ptychodiscaceae	Ptychodiscu				
Eukaryota	TSAR	Alveolata	Dinoflagellata	Dinophyceae	Dinophyceae_X	Tovelliaceae	Esoptrodiniu				
Eukaryota	TSAR	Alveolata	Dinoflagellata	Dinophyceae	Dinophyceae_X	Tovelliaceae	Jadwigia				

BLAST-like search for PR2 sequences similar to query sequence.

All PR2 sequences are searched but only those above % identity min are shown

% identity min
80 – 95 – 100

Query - at least 130 bp

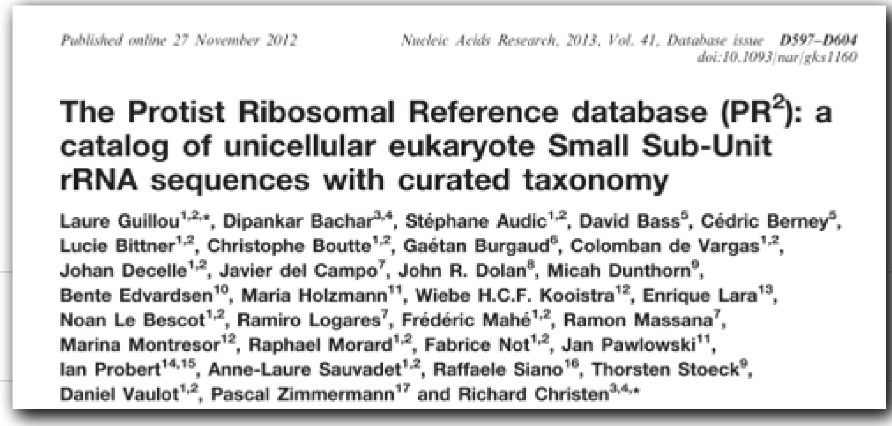
```
>KU244632.1 Micromonas commoda 18S ribosomal RNA gene, partial sequence
TGTCTAAGTATAAGCGTTATACTGTGAAACTGCGAATGGCTCATTAAATCAGCAATAGTTTCTTTGGTGG
TGTTTACTACATGGATAACCGTAGTAATTCTAGAGCTAATACATGCGTAAATCCCGACTTCGGAAGGGAC
GTATTTATTAGATAAAGACCGACCTCGTTCTGCGGTGAATCATGATAACTTCACGGACCGCATGGCCTCG
CGCCGGCGGTGTTCCATTCAAATTTCTGCCCTATCAACTTTCGACGGTAGGATAGAGGCCTACCGTGGTG
```

Matching PR2 sequences

Show 50 **entries**

pr2_accession	species	pid	mismatches	gaps	query_start	query_end	target_start	target_end
KU244663.1.1703_U	Micromonas_pusilla	99.8	4	0	1	1703	1	1703
KU743492.1.1777_U	Micromonas_commoda_A1	99.8	4	0	1	1703	55	1757
AY955011.1.1727_U	Micromonas_commoda_A1	99.8	4	0	1	1703	1	1703
AY955002.1.1727_U	Micromonas_commoda_A1	99.8	4	0	1	1703	1	1703
HM191693.1.1777_U	Micromonas_commoda_A1	99.8	4	0	1	1703	55	1757
CP001575.879464.881247_U	Micromonas_commoda_A1	99.8	4	0	1	1703	56	1758
JX188376.1.1738_U	Micromonas_commoda_A1	99.8	4	0	1	1703	36	1738
CP001575.894647.896426_U	Micromonas_commoda_A1	99.8	4	0	1	1703	55	1757
AB183589.1.1734_U	Micromonas_commoda_A1	99.8	4	0	1	1699	36	1734
AY955001.1.1727_U	Micromonas_commoda_A1	99.6	7	0	1	1703	1	1703
KF129910.1.1738_U	Micromonas_commoda_A1	99.6	7	0	1	1703	36	1738
AY955006.1.1727_U	Micromonas_commoda_A1	99.6	7	0	1	1703	1	1703
AY955007.1.1727_U	Micromonas_commoda_A1	99.6	7	0	1	1703	1	1703
KF129834.1.1738_U	Micromonas_commoda_A1	99.6	6	0	1	1703	36	1738

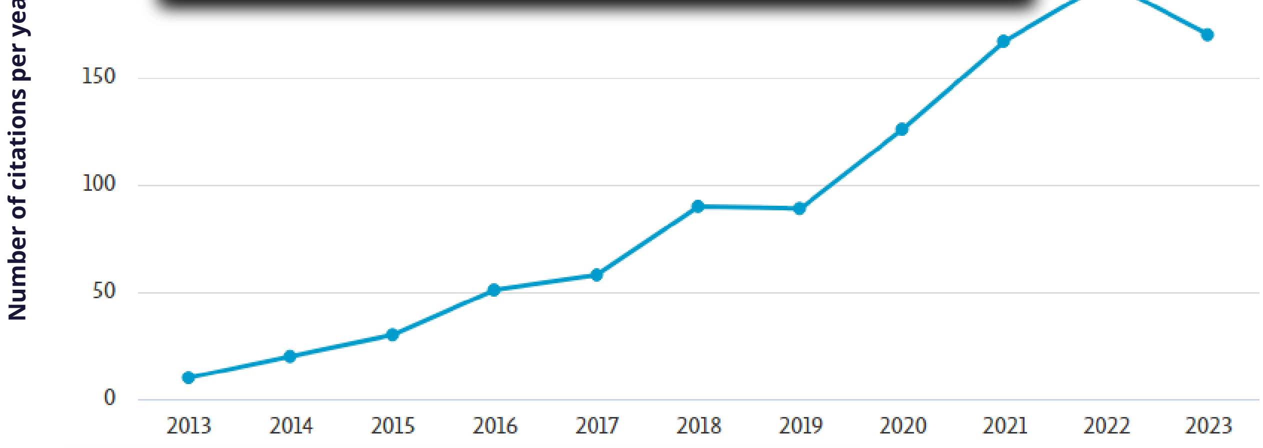
More than 1,700 papers citing PR^2

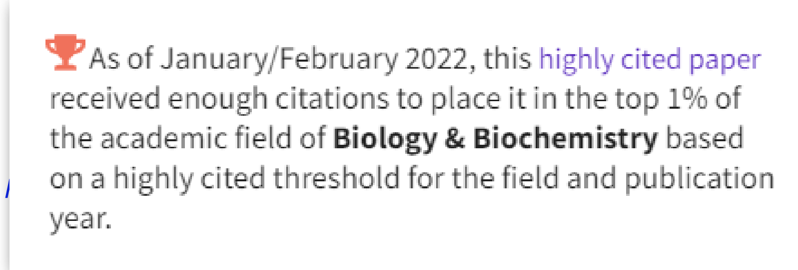


Nucleic Acids Research, 2013, Vol. 41, Database issue D597–D604
doi:10.1093/nar/gks1160

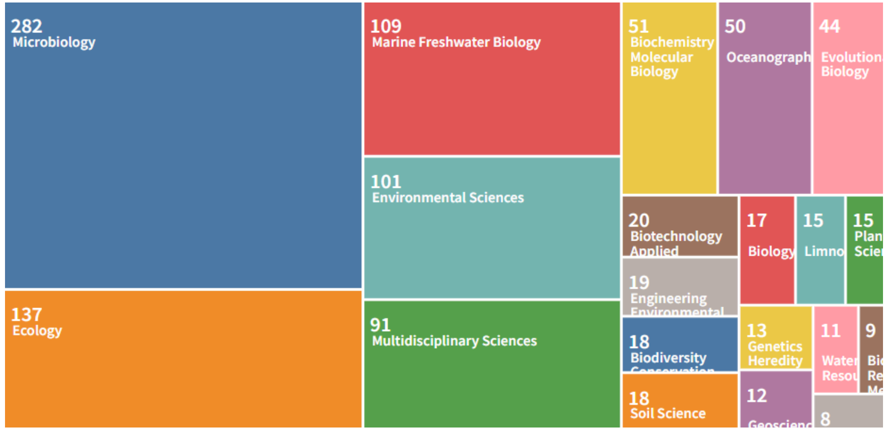
The Protist Ribosomal Reference database (PR^2): a catalog of unicellular eukaryote Small Sub-Unit rRNA sequences with curated taxonomy

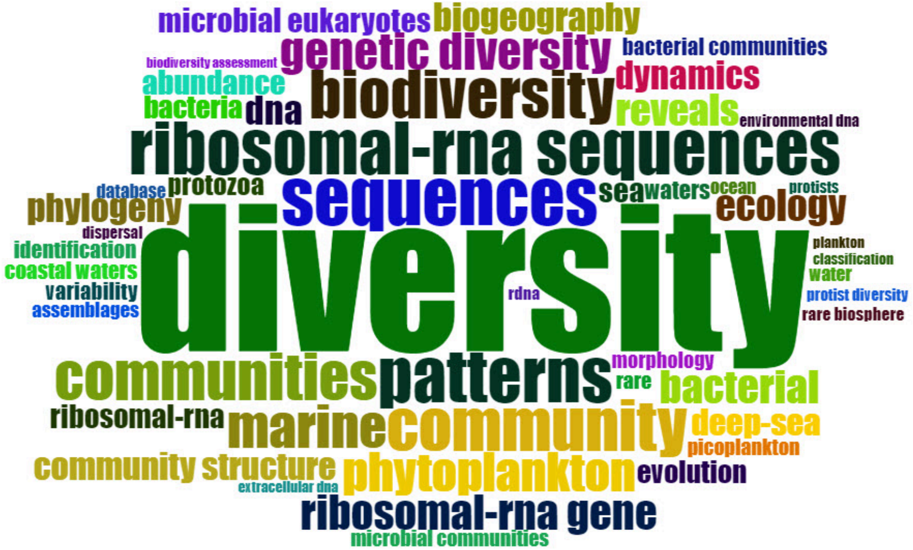
Laure Guillou[1,2,*], Dipankar Bachar[3,4], Stéphane Audic[1,2], David Bass[5], Cédric Berney[5], Lucie Bittner[1,2], Christophe Boutte[1,2], Gaétan Burgaud[6], Colomban de Vargas[1,2], Johan Decelle[1,2], Javier del Campo[7], John R. Dolan[8], Micah Dunthorn[9], Bente Edvardsen[10], Maria Holzmann[11], Wiebe H.C.F. Kooistra[12], Enrique Lara[13], Noan Le Bescot[1,2], Ramiro Logares[7], Frédéric Mahé[1,2], Ramon Massana[7], Marina Montresor[12], Raphael Morard[1,2], Fabrice Not[1,2], Jan Pawlowski[11], Ian Probert[14,15], Anne-Laure Sauvadet[1,2], Raffaele Siano[16], Thorsten Stoeck[9], Daniel Vaulot[1,2], Pascal Zimmermann[17] and Richard Christen[3,4,*]





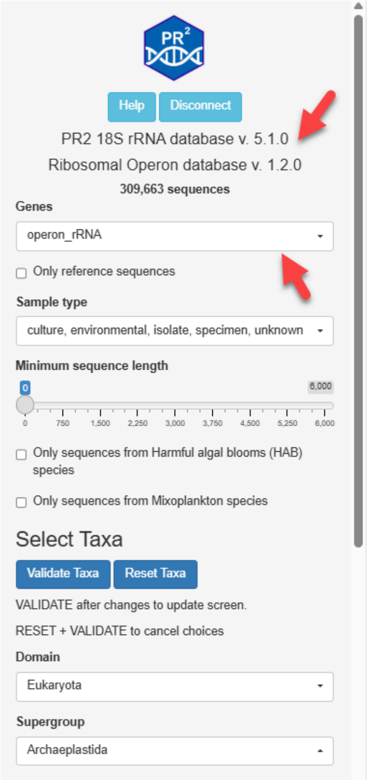
🏆 As of January/February 2022, this highly cited paper received enough citations to place it in the top 1% of the academic field of **Biology & Biochemistry** based on a highly cited threshold for the field and publication year.





Latest version - 5.1 - April 2025

- Update of some groups
  - Picozoa
  - Choropicophyceae
  - Rhodophyta
  - Dinophyceae
  - Foraminifera
- New sequences from GenBank
- Fixing taxonomy
- Link to GBIF
- Functional traits
  - HAB (Harmful Algal Bloom) species
  - ecological function - from Sommeria-Klein et al. 2021
- Web interface to Ribosomal Operon Database (ROD)



PR2 18S rRNA database v. 5.1.0

Ribosomal Operon database v. 1.2.0

309,663 sequences

Genes

operon_rRNA

☐ Only reference sequences

Sample type

culture, environmental, isolate, specimen, unknown

Minimum sequence length

☐ Only sequences from Harmful algal blooms (HAB) species

☐ Only sequences from Mixoplankton species

Select Taxa

VALIDATE after changes to update screen.

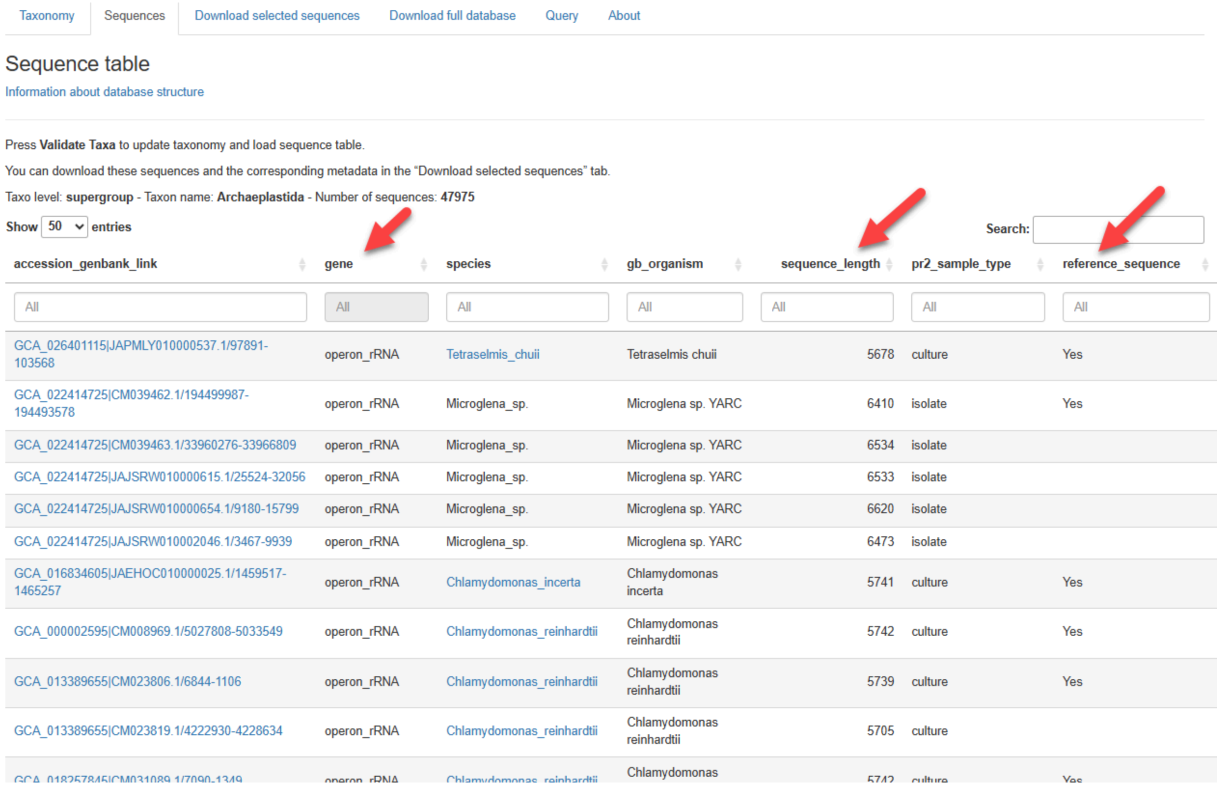
RESET + VALIDATE to cancel choices

Domain

Eukaryota

Supergroup

Archaeplastida



Sequence table

Press **Validate Taxa** to update taxonomy and load sequence table.

You can download these sequences and the corresponding metadata in the "Download selected sequences" tab.

Taxo level: **supergroup** - Taxon name: **Archaeplastida** - Number of sequences: **47975**

accession_genbank_link	gene	species	gb_organism	sequence_length	pr2_sample_type	reference_sequence
All	All	All	All	All	All	All
GCA_026401115|JAPMLY010000537.1/97891-103568	operon_rRNA	Tetraselmis_chuii	Tetraselmis chuii	5678	culture	Yes
GCA_022414725|CM039462.1/194499987-194493578	operon_rRNA	Microglena_sp.	Microglena sp. YARC	6410	isolate	Yes
GCA_022414725|CM039463.1/33960276-33966809	operon_rRNA	Microglena_sp.	Microglena sp. YARC	6534	isolate	
GCA_022414725|JAJSRW010000615.1/25524-32056	operon_rRNA	Microglena_sp.	Microglena sp. YARC	6533	isolate	
GCA_022414725|JAJSRW010000654.1/9180-15799	operon_rRNA	Microglena_sp.	Microglena sp. YARC	6620	isolate	
GCA_022414725|JAJSRW010002046.1/3467-9939	operon_rRNA	Microglena_sp.	Microglena sp. YARC	6473	isolate	
GCA_016834605|JAEHOC010000025.1/1459517-1465257	operon_rRNA	Chlamydomonas_incerta	Chlamydomonas incerta	5741	culture	Yes
GCA_000002595|CM008969.1/5027808-5033549	operon_rRNA	Chlamydomonas_reinhardtii	Chlamydomonas reinhardtii	5742	culture	Yes
GCA_013389655|CM023806.1/6844-1106	operon_rRNA	Chlamydomonas_reinhardtii	Chlamydomonas reinhardtii	5739	culture	Yes
GCA_013389655|CM023819.1/4222930-4228634	operon_rRNA	Chlamydomonas_reinhardtii	Chlamydomonas reinhardtii	5705	culture	
GCA_018257845|CM031089.1/7090-1349	operon_rRNA	Chlamydomonas_reinhardtii	Chlamydomonas	5742	culture	Yes

What's next for pr2-database

- Integration of taxonomy into GBIF (in progress)
- Need more participation from taxonomy experts
- Full operon
  - Include metagenomic data ?
- Functional annotation
  - Trophic mode (photo, hetero, mixo…)
  - Size group (pico, nano, micro…)
- Other genes: *coi*

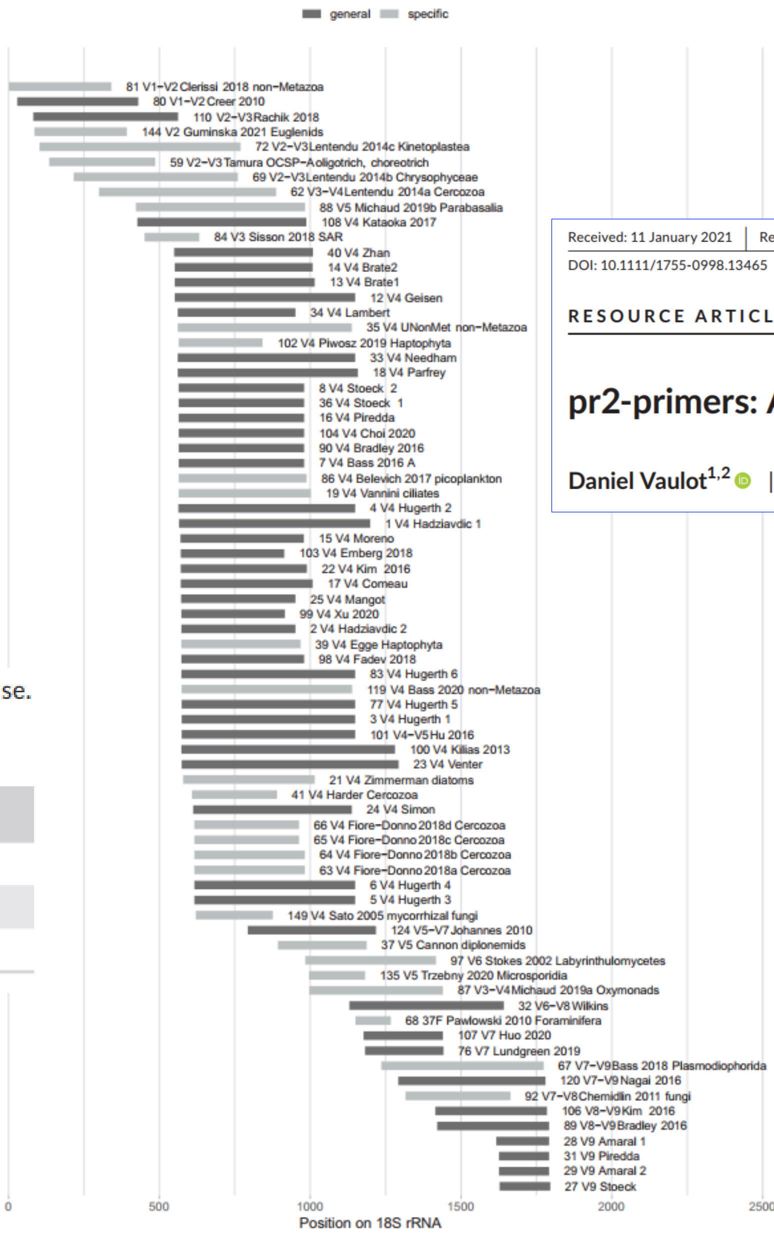
The PR^2 primer database

18S rRNA primers

- Wide diversity of primers and sets
- No database for protists
- Taxonomic specificity of primers?

TABLE 1 Summary of primers listed in the pr2-primers database. General primers target all eukaryotes and specific primers only certain taxonomic groups

Direction	General primers	Specific primers
fwd	55	89
rev	53	88
Total	108	177



Received: 11 January 2021 | Revised: 31 May 2021 | Accepted: 6 July 2021

DOI: 10.1111/1755-0998.13465

MOLECULAR ECOLOGY RESOURCES WILEY

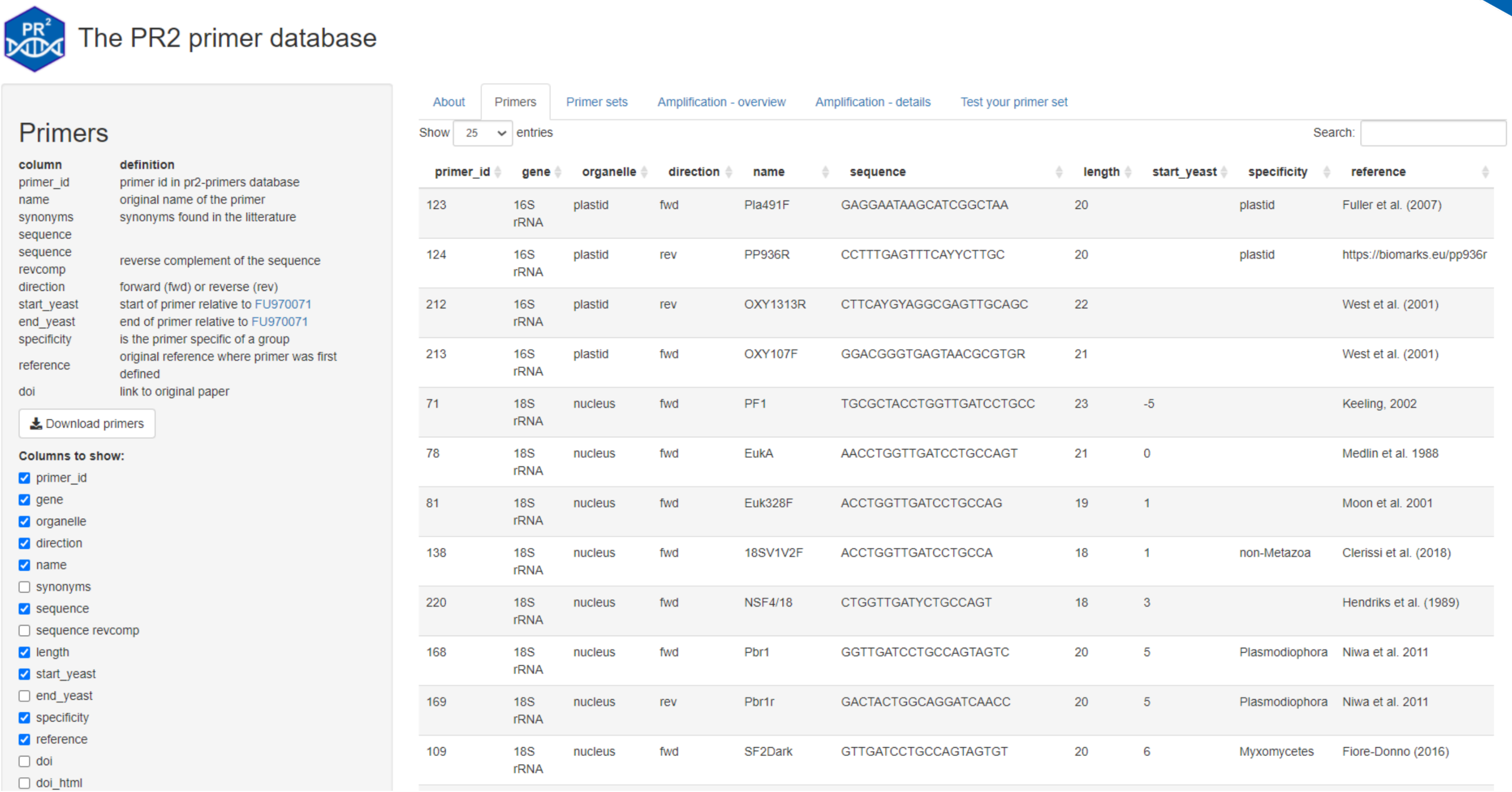
RESOURCE ARTICLE

pr2-primers: An 18S rRNA primer database for protists

Daniel Vaulot[1,2] | Stefan Geisen[3,4,5] | Frédéric Mahé[6,7] | David Bass[8,9]

Vaulot, D., Mahé, F., Bass, D., Geisen, S., 2022. pr2-primer: an 18S rRNA primer database for protists. Molecular Ecology Resources 22, 168–179. https://doi.org/10.1111/1755-0998.13465

pr2-primers database



The PR2 primer database

Primers

column	definition
primer_id	primer id in pr2-primers database
name	original name of the primer
synonyms	synonyms found in the litterature
sequence	
sequence revcomp	reverse complement of the sequence
direction	forward (fwd) or reverse (rev)
start_yeast	start of primer relative to FU970071
end_yeast	end of primer relative to FU970071
specificity	is the primer specific of a group
reference	original reference where primer was first defined
doi	link to original paper

Columns to show:

- ☑ primer_id
- ☑ gene
- ☑ organelle
- ☑ direction
- ☑ name
- ☐ synonyms
- ☑ sequence
- ☐ sequence revcomp
- ☑ length
- ☑ start_yeast
- ☐ end_yeast
- ☑ specificity
- ☑ reference
- ☐ doi
- ☐ doi_html

Show 25 entries Search:

primer_id	gene	organelle	direction	name	sequence	length	start_yeast	specificity	reference
123	16S rRNA	plastid	fwd	Pla491F	GAGGAATAAGCATCGGCTAA	20		plastid	Fuller et al. (2007)
124	16S rRNA	plastid	rev	PP936R	CCTTTGAGTTTCAYYCTTGC	20		plastid	https://biomarks.eu/pp936r
212	16S rRNA	plastid	rev	OXY1313R	CTTCAYGYAGGCGAGTTGCAGC	22			West et al. (2001)
213	16S rRNA	plastid	fwd	OXY107F	GGACGGGTGAGTAACGCGTGR	21			West et al. (2001)
71	18S rRNA	nucleus	fwd	PF1	TGCGCTACCTGGTTGATCCTGCC	23	-5		Keeling, 2002
78	18S rRNA	nucleus	fwd	EukA	AACCTGGTTGATCCTGCCAGT	21	0		Medlin et al. 1988
81	18S rRNA	nucleus	fwd	Euk328F	ACCTGGTTGATCCTGCCAG	19	1		Moon et al. 2001
138	18S rRNA	nucleus	fwd	18SV1V2F	ACCTGGTTGATCCTGCCA	18	1	non-Metazoa	Clerissi et al. (2018)
220	18S rRNA	nucleus	fwd	NSF4/18	CTGGTTGATYCTGCCAGT	18	3		Hendriks et al. (1989)
168	18S rRNA	nucleus	fwd	Pbr1	GGTTGATCCTGCCAGTAGTC	20	5	Plasmodiophora	Niwa et al. 2011
169	18S rRNA	nucleus	rev	Pbr1r	GACTACTGGCAGGATCAACC	20	5	Plasmodiophora	Niwa et al. 2011
109	18S rRNA	nucleus	fwd	SF2Dark	GTTGATCCTGCCAGTAGTGT	20	6	Myxomycetes	Fiore-Donno (2016)

https://app.pr2-primers.org

In silico analysis against PR^2

- Mismatches #
- Mismatches position
- Amplicon size

The PR2 primer database

About | Primers | Primer sets | Amplification - overview | Amplification - details | Test your primer set

Precomputed results for primer sets

Against PR2 sequence database

Primer set
004 - V4 Hugerth_2 - general

Kingdom
Eukaryota

Supergroup
All

Division
All

Class
All

Update plot

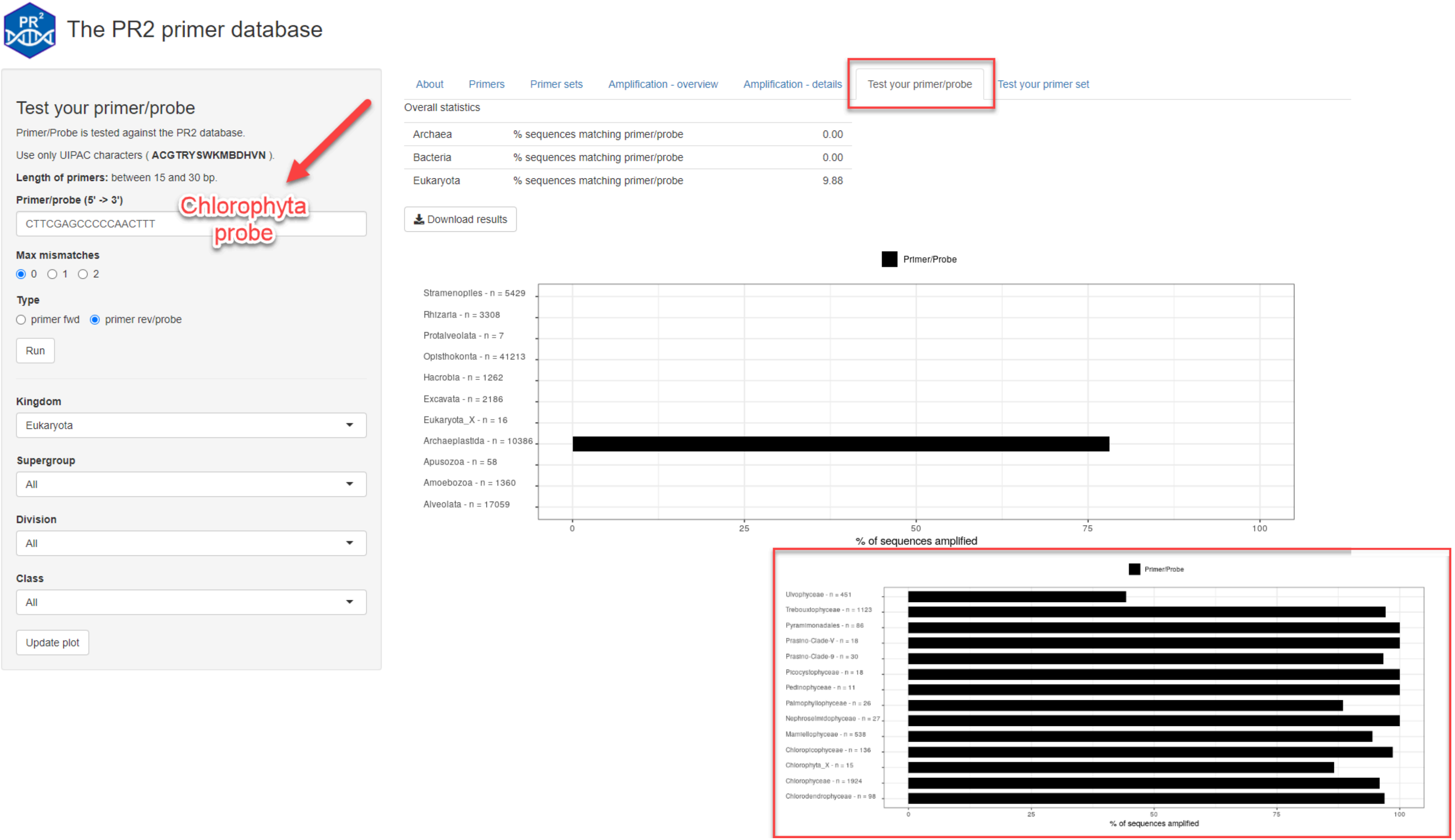
Overall statistics

Archaea	% sequences matching forward primer	82.29
Archaea	% sequences matching reverse primer	100.00
Archaea	% sequences amplified	82.29
Archaea	mean amplicon size	414.77
Bacteria	% sequences matching forward primer	98.99
Bacteria	% sequences matching reverse primer	98.70
Bacteria	% sequences amplified	97.74
Bacteria	mean amplicon size	409.80
Eukaryota	% sequences matching forward primer	95.76
Eukaryota	% sequences matching reverse primer	98.29
Eukaryota	% sequences amplified	94.38
Eukaryota	mean amplicon size	604.08

Precomputed

Top panel: Location of mismatches for forward and reverse primer. **Bottom panel left** number of mismatches. **Bottom panel right:** amplicon size

Test your own



The PR2 primer database

Test your primer/probe

Primer/Probe is tested against the PR2 database.

Use only UIPAC characters (**ACGTRYSWKMBDHVN**).

Length of primers: between 15 and 30 bp.

Primer/probe (5' -> 3'): CTTCGAGCCCCCAACTTT

Chlorophyta probe

Max mismatches: ● 0 ○ 1 ○ 2

Type: ○ primer fwd ● primer rev/probe

Kingdom: Eukaryota

Supergroup: All

Division: All

Class: All

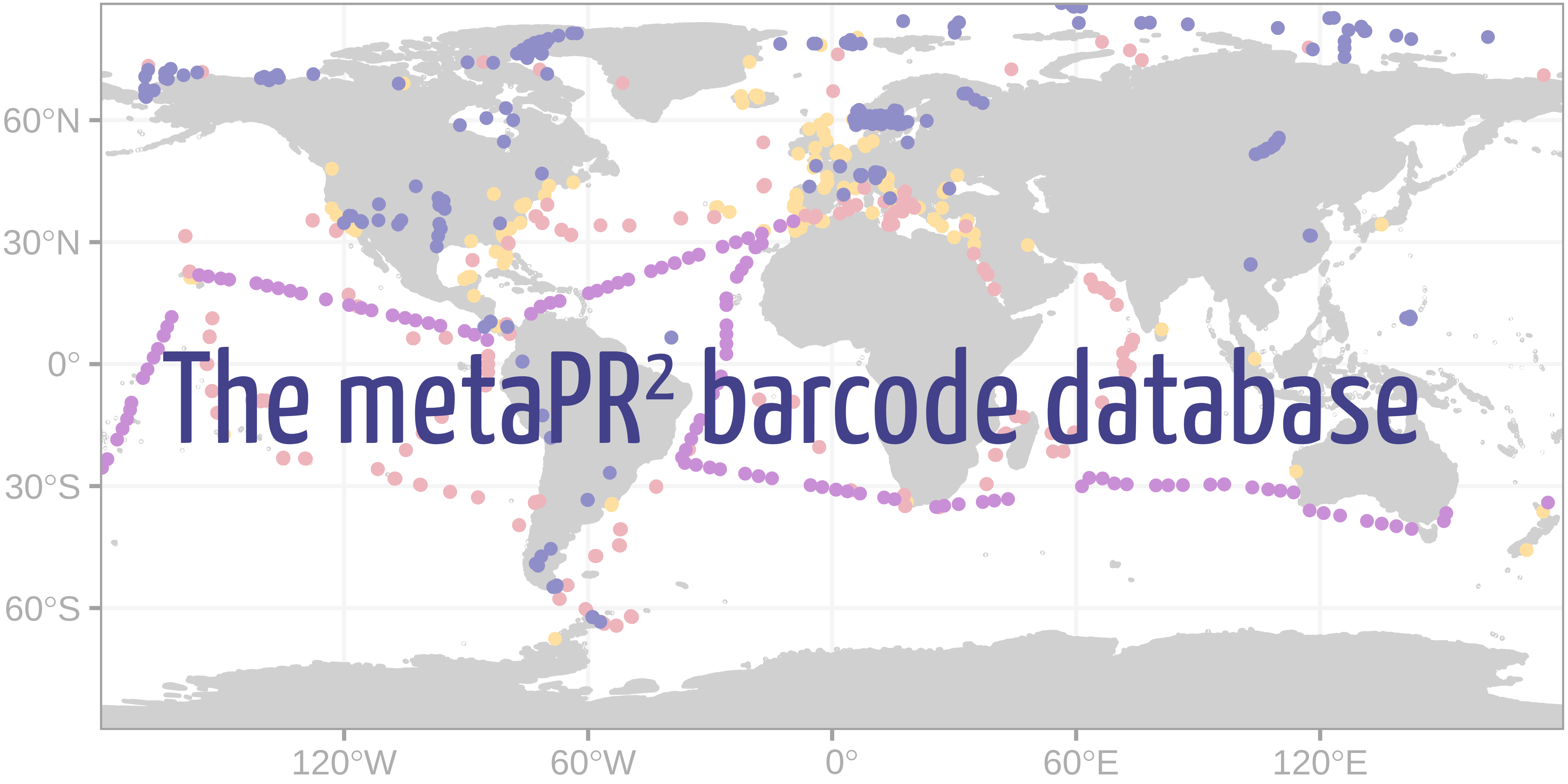
Overall statistics

Archaea	% sequences matching primer/probe	0.00
Bacteria	% sequences matching primer/probe	0.00
Eukaryota	% sequences matching primer/probe	9.88

What's next for pr2-primers

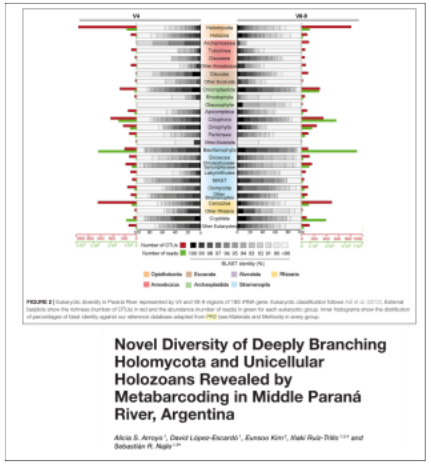
- Update PR^2 to 5.1.0
- Include ROD database
- Add more primers
- ITS/28S primers

The metaPR^2 barcode database

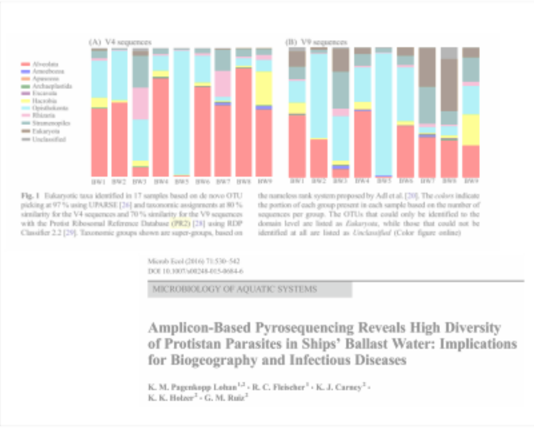


Motivation

- In the last decade, many metabarcoding studies
- Data hard to compare:
    - ◻ Different primers
    - ◻ Different processing
    - ◻ Different similarity levels
- Processed data usually not available
- Metadata not available
- Few global datasets used (*Tara*, Malaspina)
- These datasets only temperate and tropical marine


Novel Diversity of Deeply Branching Holomycota and Unicellular Holozoans Revealed by Metabarcoding in Middle Paraná River, Argentina

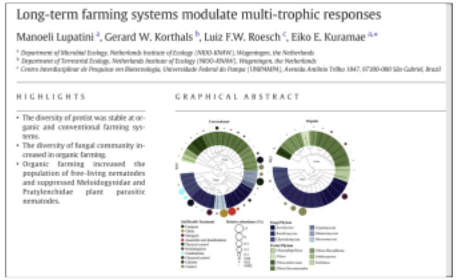
Alicia S. Arroyo, David López-Escardó, Eunsoo Kim, Iñaki Ruiz-Trillo and Sebastián R. Najle


Microb Ecol (2016) 71:530–542
DOI 10.1007/s00248-015-0684-6

MICROBIOLOGY OF AQUATIC SYSTEMS

Amplicon-Based Pyrosequencing Reveals High Diversity of Protistan Parasites in Ships' Ballast Water: Implications for Biogeography and Infectious Diseases

K. M. Pagenkopp Lohan[1,2] · R. C. Fleischer[1] · K. J. Carney[2] · K. K. Holzer[2] · G. M. Ruiz[2]

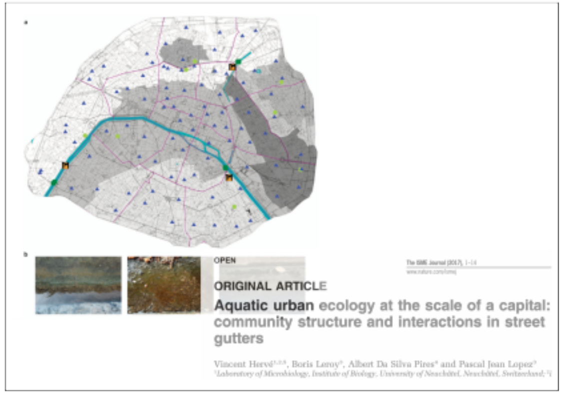

Long-term farming systems modulate multi-trophic responses

Manoeli Lupatini[a], Gerard W. Korthals[b], Luiz F.W. Roesch[c], Eiko E. Kuramae[a,*]

HIGHLIGHTS

GRAPHICAL ABSTRACT

- The diversity of protist was stable at organic and conventional farming systems.
- The diversity of fungal community increased in organic farming.

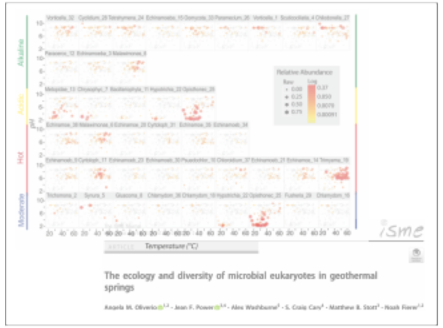

OPEN

The ISME Journal (2017), 1–14
www.nature.com/ismej

ORIGINAL ARTICLE

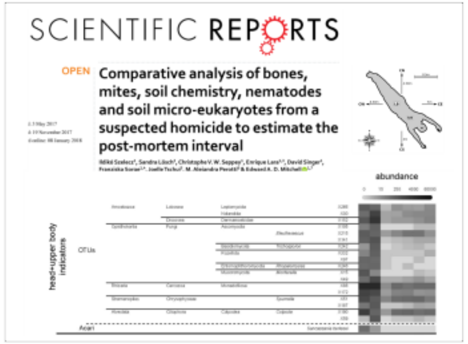
Aquatic urban ecology at the scale of a capital: community structure and interactions in street gutters

Vincent Hervé, Boris Leroy, Albert Da Silva Pires and Pascal Jean Lopez


ARTICLE

The ecology and diversity of microbial eukaryotes in geothermal springs

Angela M. Oliverio · Jean F. Power · Alex Washburne · S. Craig Cary · Matthew B. Stott · Noah Fierer

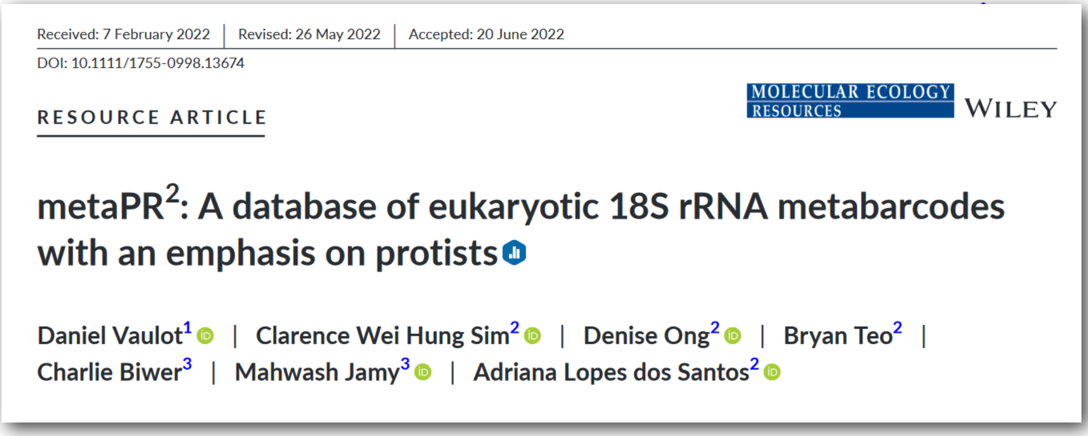

SCIENTIFIC REPORTS

OPEN

Comparative analysis of bones, mites, soil chemistry, nematodes and soil micro-eukaryotes from a suspected homicide to estimate the post-mortem interval

metapr2 - strategy

- Scan papers
- Start from raw data (fastq) available from GenBank SRA
- Use dada2 pipeline producing ASVs
  - Different datasets are comparable
- Annotate taxonomy with PR^2
- Integrate metadata
  - Latitude and longitude
  - Depth
  - Substrate (water, ice, soil)
- Data stored in MySQL database
- Develop web interface using R shiny



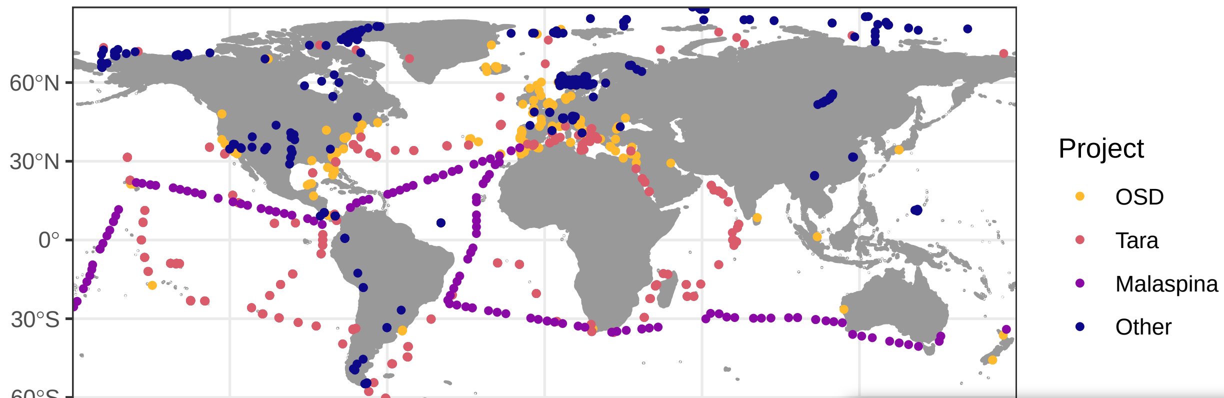
Received: 7 February 2022 | Revised: 26 May 2022 | Accepted: 20 June 2022

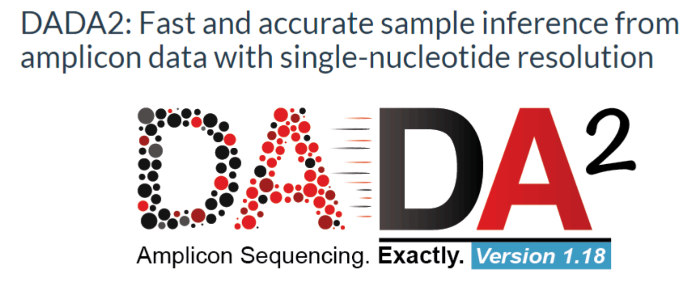
DOI: 10.1111/1755-0998.13674

RESOURCE ARTICLE

metaPR^2: A database of eukaryotic 18S rRNA metabarcodes with an emphasis on protists

Daniel Vaulot[1] | Clarence Wei Hung Sim[2] | Denise Ong[2] | Bryan Teo[2] | Charlie Biwer[3] | Mahwash Jamy[3] | Adriana Lopes dos Santos[2]





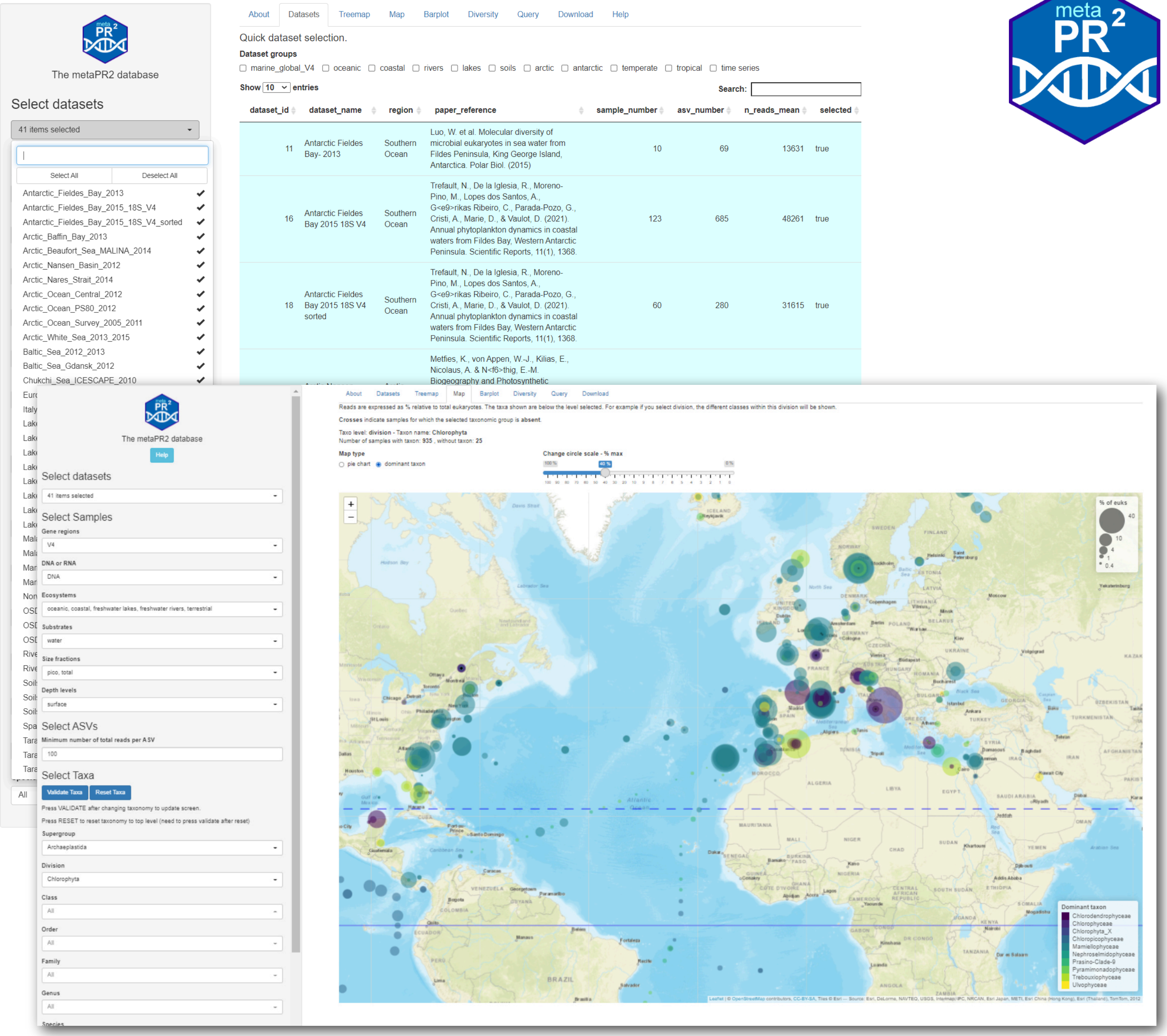
DADA2: Fast and accurate sample inference from amplicon data with single-nucleotide resolution

Vaulot, D., Sim, C.W.H., Ong, D., Teo, B., Biwer, C., Jamy, M., Lopes dos Santos, A., 2022. metaPR2: a database of eukaryotic 18S rRNA metabarcodes with an emphasis on protists. Molecular Ecology Ressources

metapr2 - version 2.1

- Datasets: 59
  - ☐ *Tara* Oceans (reprocessed with dada2)
  - ☐ Malaspina
  - ☐ Ocean Sampling Day - 2014 & 2015
  - ☐ Arctic datasets
  - ☐ Deep Sea
  - ☐ Lakes, Rivers, Soils
- Samples: 6,200
- ASVs: 90,000
- Annotated with PR2 version 5.0.1

https://app.metapr2.org



The metaPR2 database

Select datasets

41 items selected

Antarctic_Fieldes_Bay_2013 ✓
Antarctic_Fieldes_Bay_2015_18S_V4 ✓
Antarctic_Fieldes_Bay_2015_18S_V4_sorted ✓
Arctic_Baffin_Bay_2013 ✓
Arctic_Beaufort_Sea_MALINA_2014 ✓
Arctic_Nansen_Basin_2012 ✓
Arctic_Nares_Strait_2014 ✓
Arctic_Ocean_Central_2012 ✓
Arctic_Ocean_PS80_2012 ✓
Arctic_Ocean_Survey_2005_2011 ✓
Arctic_White_Sea_2013_2015 ✓
Baltic_Sea_2012_2013 ✓
Baltic_Sea_Gdansk_2012 ✓
Chukchi_Sea_ICESCAPE_2010 ✓
Euro
Italy
Lake
Lake
Lake
Lake
Lake
Lake
Lake
Lake
Mala
Mala
Mari
Mari
Norv
OSD
OSD
OSD
Rive
Rive
Soils
Soils
Soils
Spa
Tara
Tara
Tara

All

Quick dataset selection.

Dataset groups

☐ marine_global_V4 ☐ oceanic ☐ coastal ☐ rivers ☐ lakes ☐ soils ☐ arctic ☐ antarctic ☐ temperate ☐ tropical ☐ time series

Show 10 entries

Search:

dataset_id	dataset_name	region	paper_reference	sample_number	asv_number	n_reads_mean	selected
11	Antarctic Fieldes Bay- 2013	Southern Ocean	Luo, W. et al. Molecular diversity of microbial eukaryotes in sea water from Fildes Peninsula, King George Island, Antarctica. Polar Biol. (2015)	10	69	13631	true
16	Antarctic Fieldes Bay 2015 18S V4	Southern Ocean	Trefault, N., De la Iglesia, R., Moreno-Pino, M., Lopes dos Santos, A., G<e9>rikas Ribeiro, C., Parada-Pozo, G., Cristi, A., Marie, D., & Vaulot, D. (2021). Annual phytoplankton dynamics in coastal waters from Fildes Bay, Western Antarctic Peninsula. Scientific Reports, 11(1), 1368.	123	685	48261	true
18	Antarctic Fieldes Bay 2015 18S V4 sorted	Southern Ocean	Trefault, N., De la Iglesia, R., Moreno-Pino, M., Lopes dos Santos, A., G<e9>rikas Ribeiro, C., Parada-Pozo, G., Cristi, A., Marie, D., & Vaulot, D. (2021). Annual phytoplankton dynamics in coastal waters from Fildes Bay, Western Antarctic Peninsula. Scientific Reports, 11(1), 1368.	60	280	31615	true
			Metfies, K., von Appen, W.-J., Kilias, E., Nicolaus, A. & N<f6>thig, E.-M. Biogeography and Photosynthetic				

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

Species

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**.

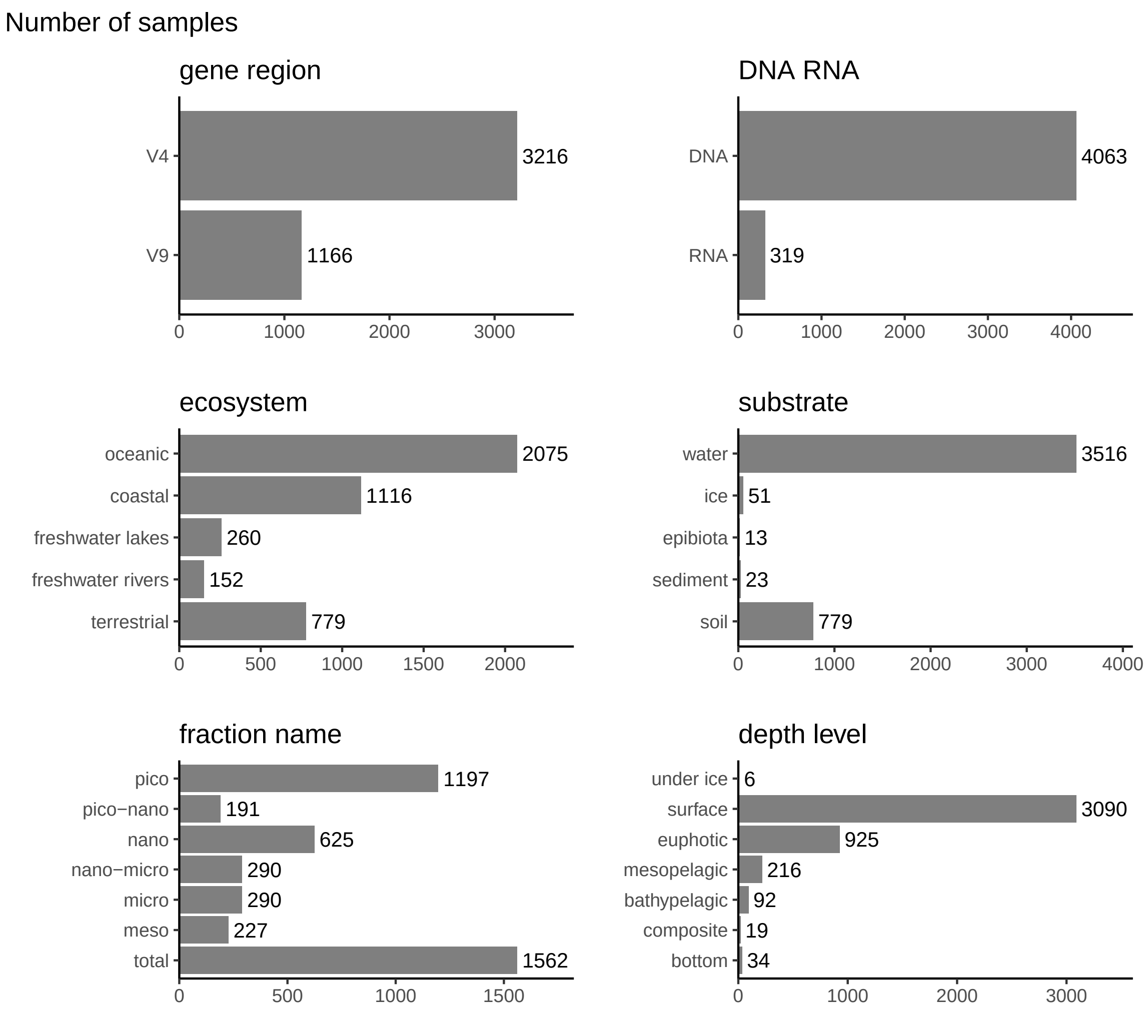
Taxo level: **division** - Taxon name: **Chlorophyta**
Number of samples with taxon: **935** , without taxon: **25**

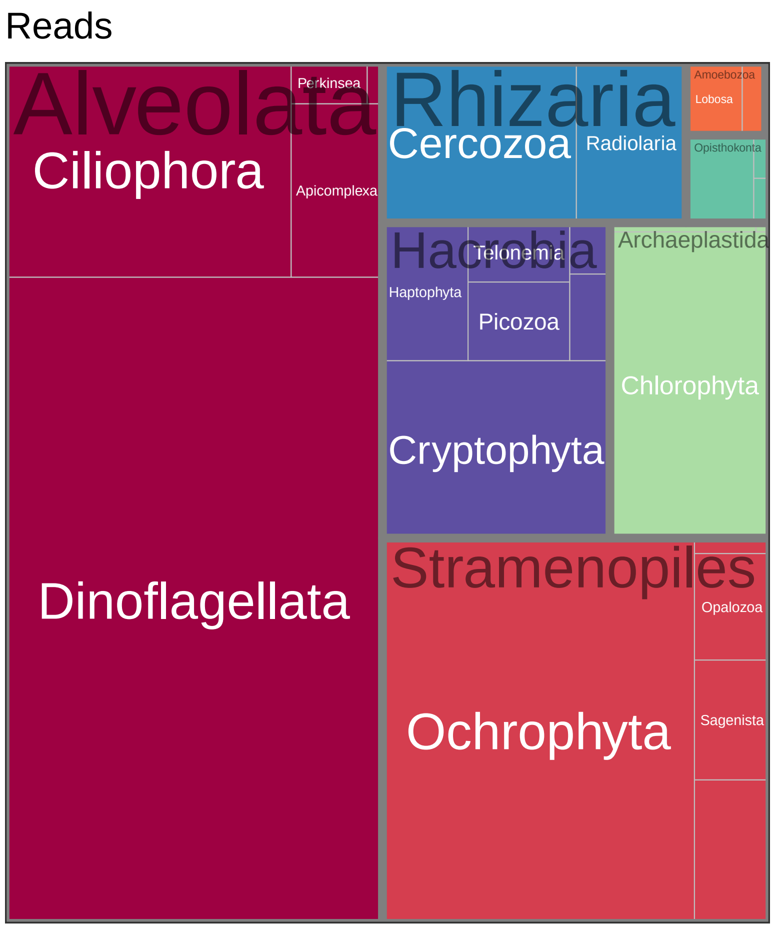
Map type

○ pie chart ● dominant taxon

Change circle scale - % max: 40 %

Samples and ASVs



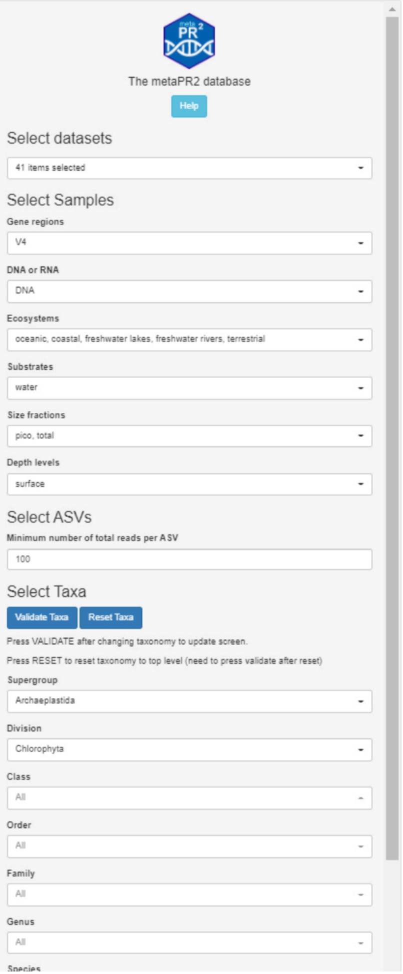


Protists only

Web interface

https://app.metapr2.org

- Built with R shiny
  - ☐ Available also as R package
- Panels
  - ☐ Datasets
  - ☐ Treemaps
  - ☐ Maps
  - ☐ Barplots
  - ☐ Diversity
  - ☐ Query
  - ☐ Download



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

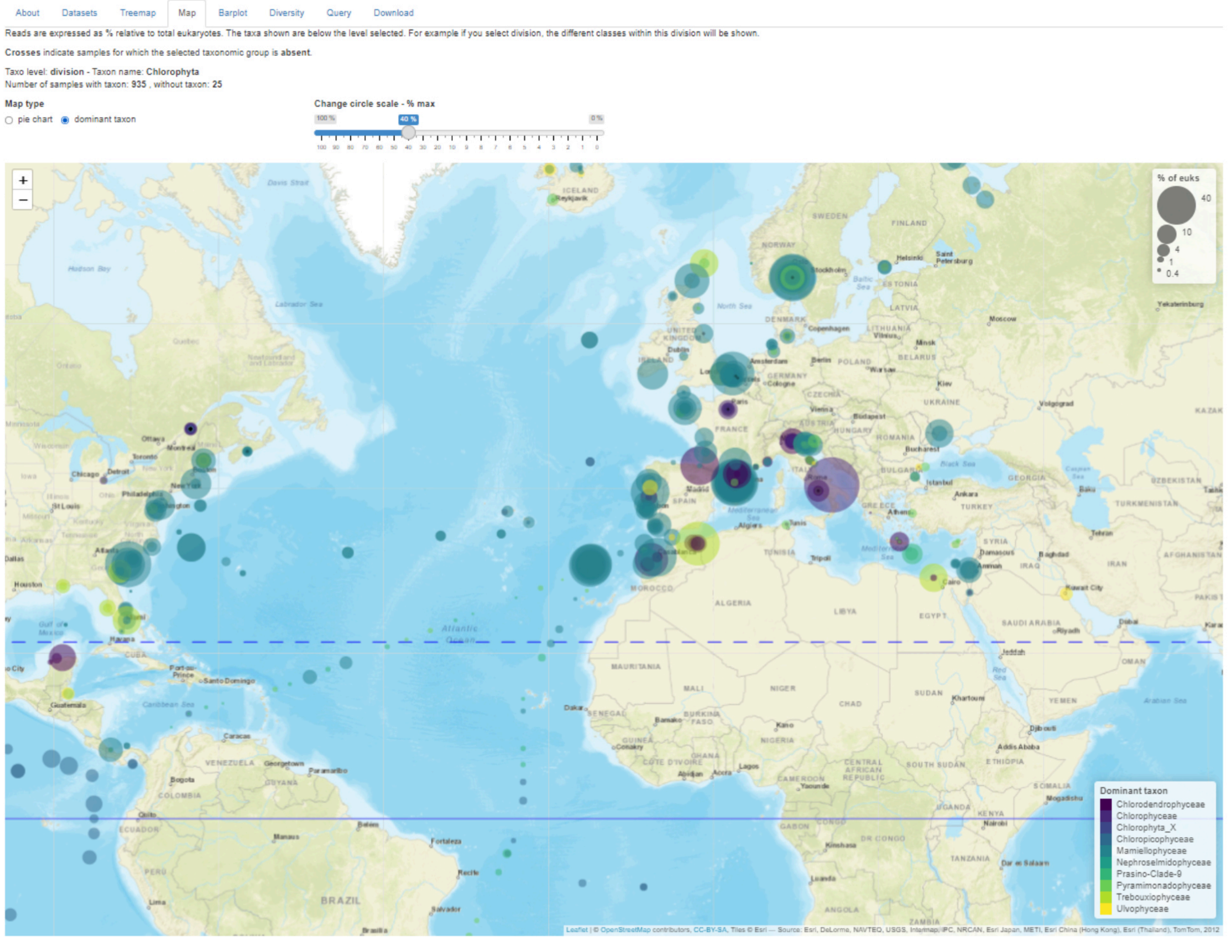
Class: All

Order: All

Family: All

Genus: All

Species



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**.

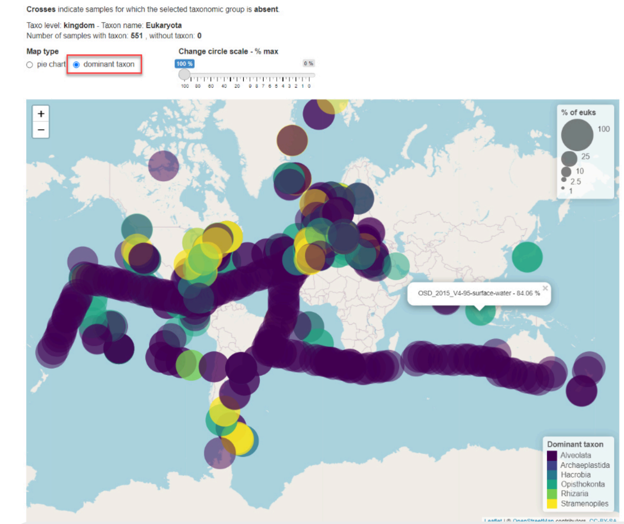
Taxo level: **division** - Taxon name: **Chlorophyta**
Number of samples with taxon: **935** , without taxon: **25**

Map type
○ pie chart ● dominant taxon

Change circle scale - % max

Web interface

Maps

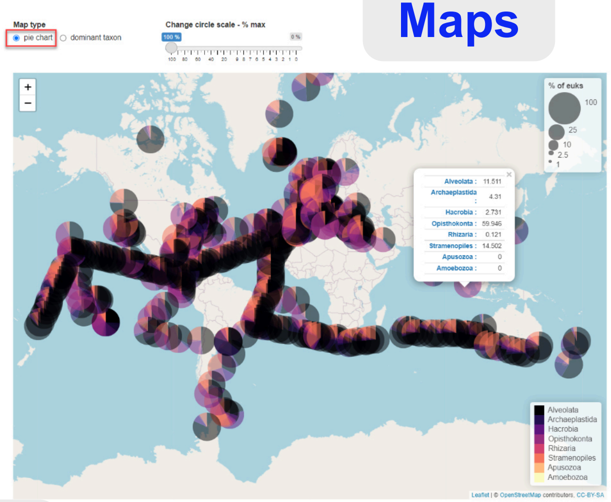


Crosses indicate samples for which the selected taxonomic group is **absent**.

Taxo level: **kingdom** - Taxon name: **Eukaryota**
Number of samples with taxon: **551** , without taxon: **0**

Map type
○ pie chart ● dominant taxon

Change circle scale - % max

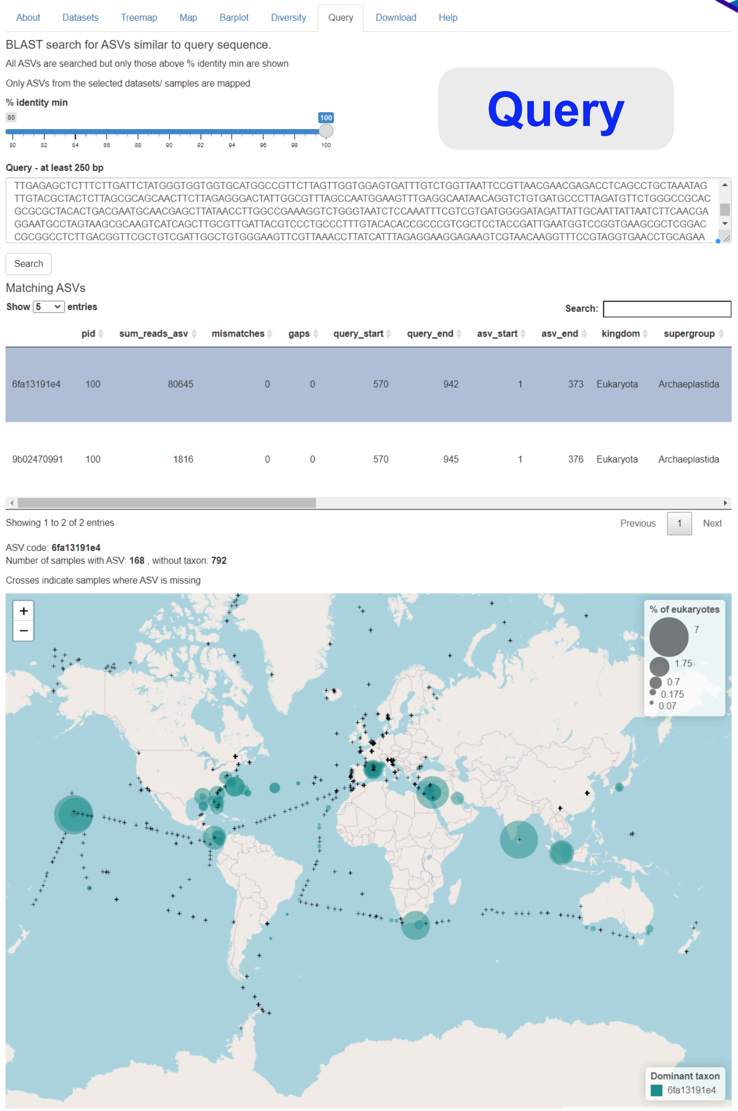


Map type
● pie chart ○ dominant taxon

Change circle scale - % max

Alveolata :	11.511
Archaeplastida :	4.31
Hacrobia :	2.731
Opisthokonta :	59.946
Rhizaria :	0.121
Stramenopiles :	14.502
Apusozoa :	0
Amoebozoa :	0

Query



BLAST search for ASVs similar to query sequence.

All ASVs are searched but only those above % identity min are shown

Only ASVs from the selected datasets/ samples are mapped

% identity min

Query - at least 250 bp

```
TTGAGAGCTCTTTCTTGATTCTATGGGTGGTGGTGCATGGCCGTTCTTAGTTGGTGGAGTGATTTGTCTGGTTAATTCCGTTAACGAACGAGACCTCAGCCTGCTAAATAG
TTGTACGCTACTCTTAGCGCAGCAACTTCTTAGAGGGACTATTGGCGTTTAGCCAATGGAAGTTTGAGGCAATAACAGGTCTGTGATGCCCTTAGATGTTCTGGGCCGCAC
GCGCGCTACACTGACGAATGCAACGAGCTTATAACCTTGGCCGAAAGGTCTGGGTAATCTCCAAATTTCGTCGTGATGGGGATAGATTATTGCAATTATTAATCTTCAACGA
GGAATGCCTAGTAAGCGCAAGTCATCAGCTTGCGTTGATTACGTCCCTGCCCTTTGTACACACCGCCCGTCGCTCCTACCGATTGAATGGTCCGGTGAAGCGCTCGGAC
CGCGGCCTCTTGACGGTTCGCTGTCGATTGGCTGTGGGAAGTTCGTTAAACCTTATCATTTAGAGGAAGGAGAAGTCGTAACAAGGTTTCCGTAGGTGAACCTGCAGAA
```

Matching ASVs

Show 5 **entries** **Search:**

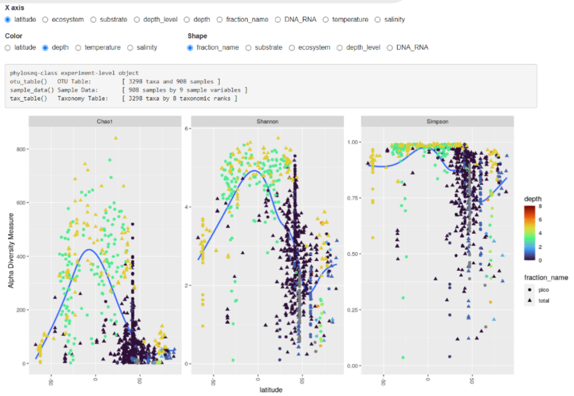
	pid	sum_reads_asv	mismatches	gaps	query_start	query_end	asv_start	asv_end	kingdom	supergroup
6fa13191e4	100	80645	0	0	570	942	1	373	Eukaryota	Archaeplastida
9b02470991	100	1816	0	0	570	945	1	376	Eukaryota	Archaeplastida

Showing 1 to 2 of 2 entries

ASV code: **6fa13191e4**
Number of samples with ASV: **168** , without taxon: **792**

Crosses indicate samples where ASV is missing

Alpha and beta diversity

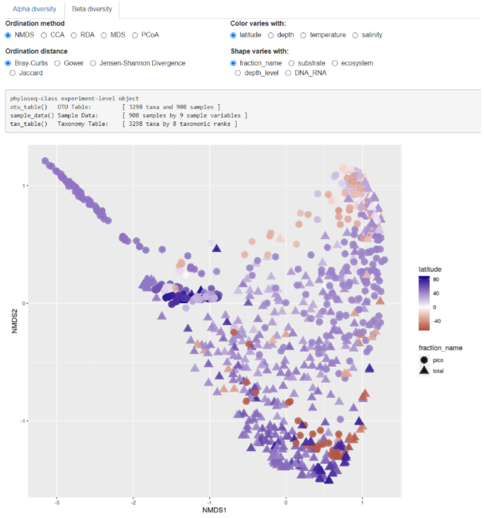


X axis
● latitude ○ ecosystem ○ substrate ○ depth_level ○ depth ○ fraction_name ○ DNA_RNA ○ temperature ○ salinity

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

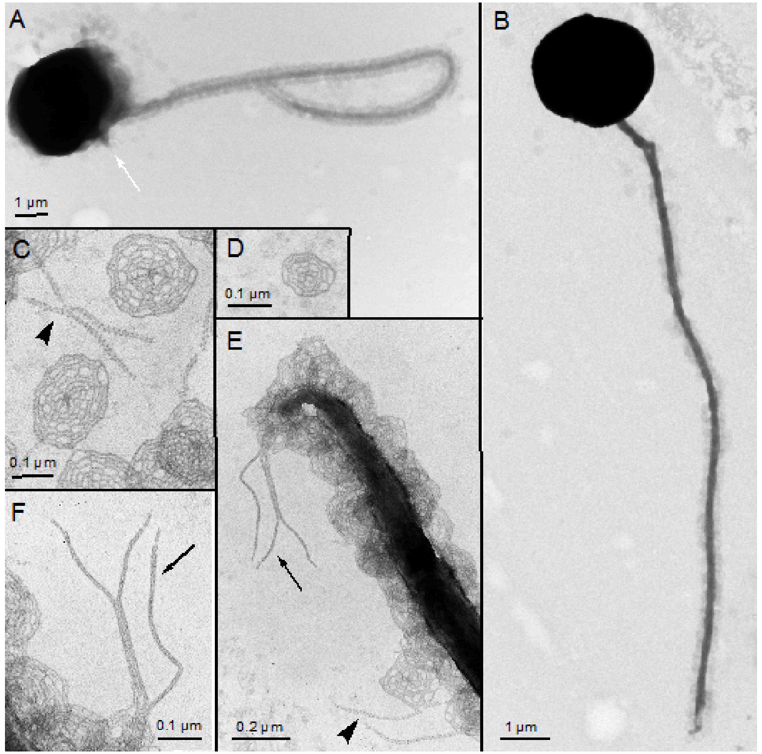
Ordination distance
● Bray-Curtis ○ Gower ○ Jensen-Shannon Divergence
○ Jaccard

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 ]
```

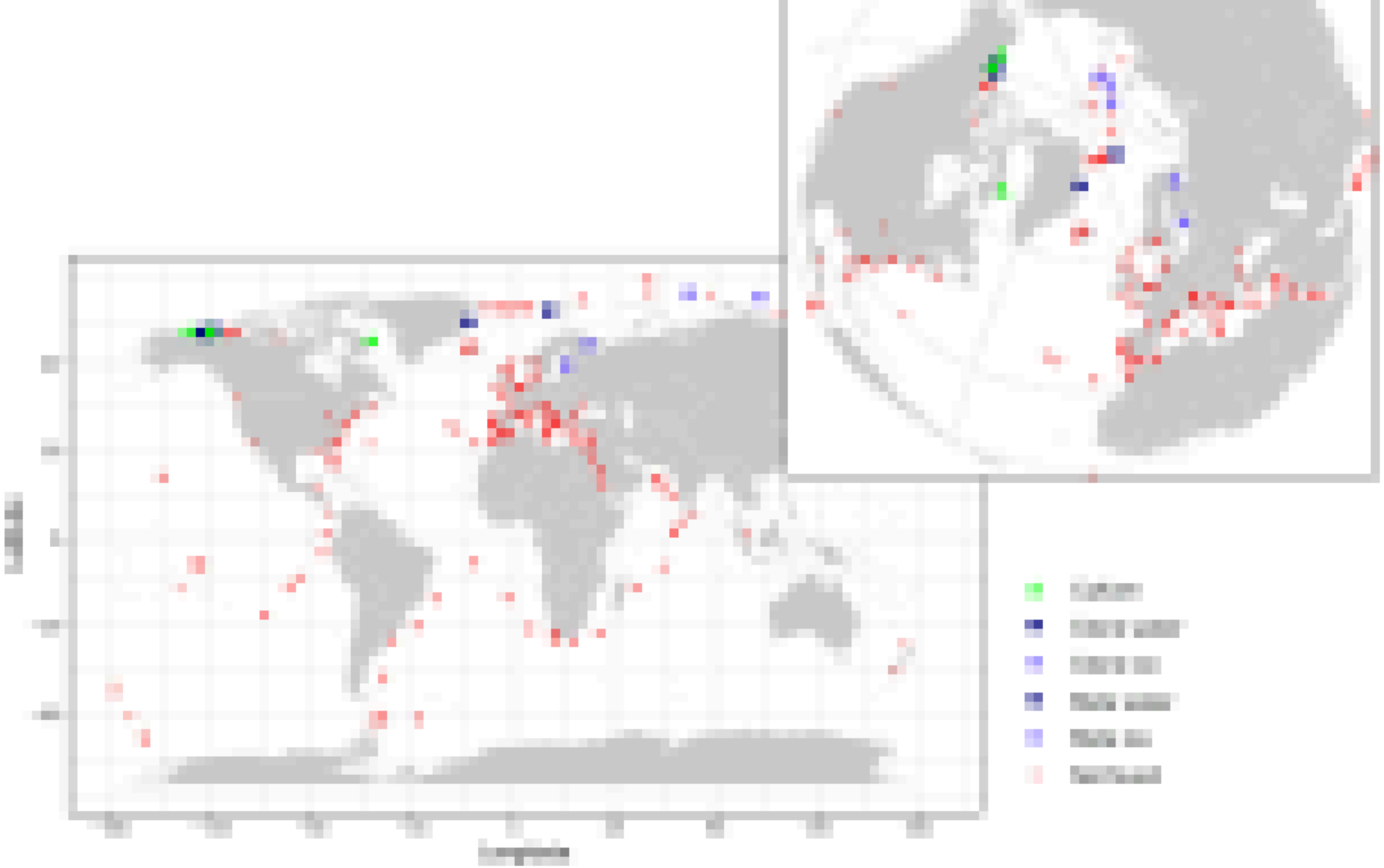
Biogeography




J. Phycol. 56, 37–51 (2020)
© 2019 Phycological Society of America
DOI: 10.1111/jpy.12932

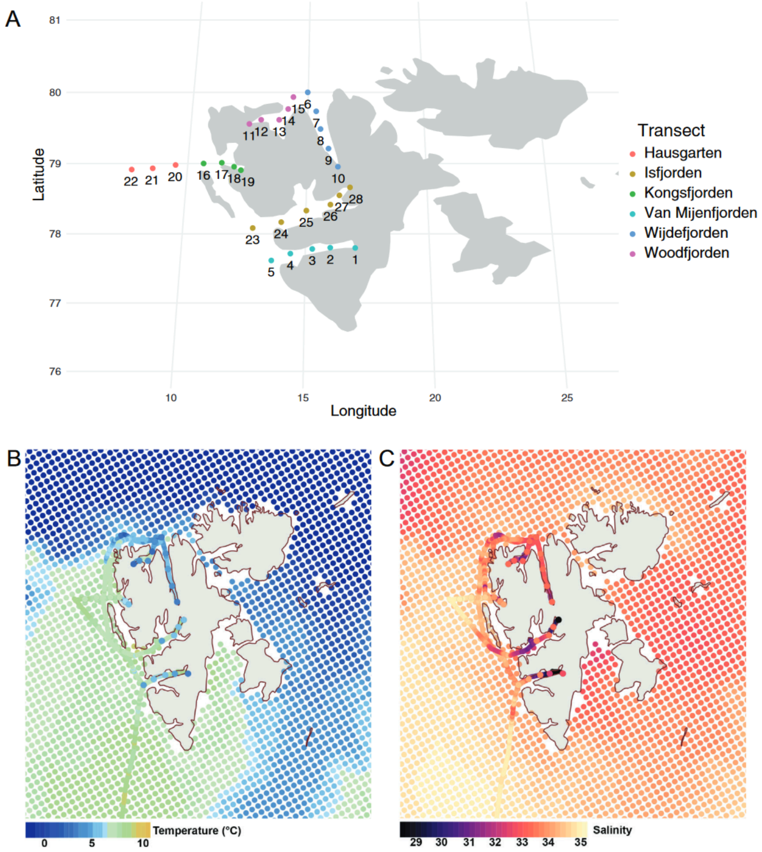
MANTONIELLA BEAUFORTII AND MANTONIELLA BAFFINENSIS SP. NOV. (MAMIELLALES, MAMIELLOPHYCEAE), TWO NEW GREEN ALGAL SPECIES FROM THE HIGH ARCTIC[1]

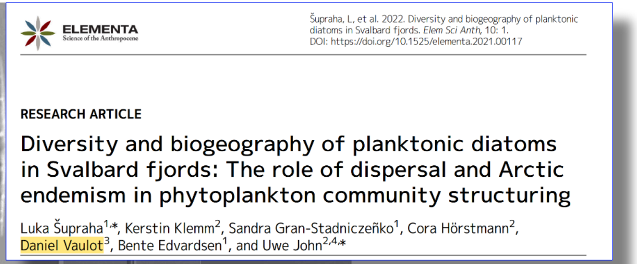
Sheree Yau[2,3]
Integrative Marine Biology Laboratory (BIOM), CNRS, UMR7232, Sorbonne Université, Banyuls sur Mer, France



Yau et al.., 2020. Mantoniella beaufortii and Mantoniella baffinensis sp. nov. (Mamiellales, Mamiellophyceae), two new green algal species from the high arctic. Journal of Phycology 56, 37–51. https://doi.org/10.1111/jpy.12932

Svalbard diatoms





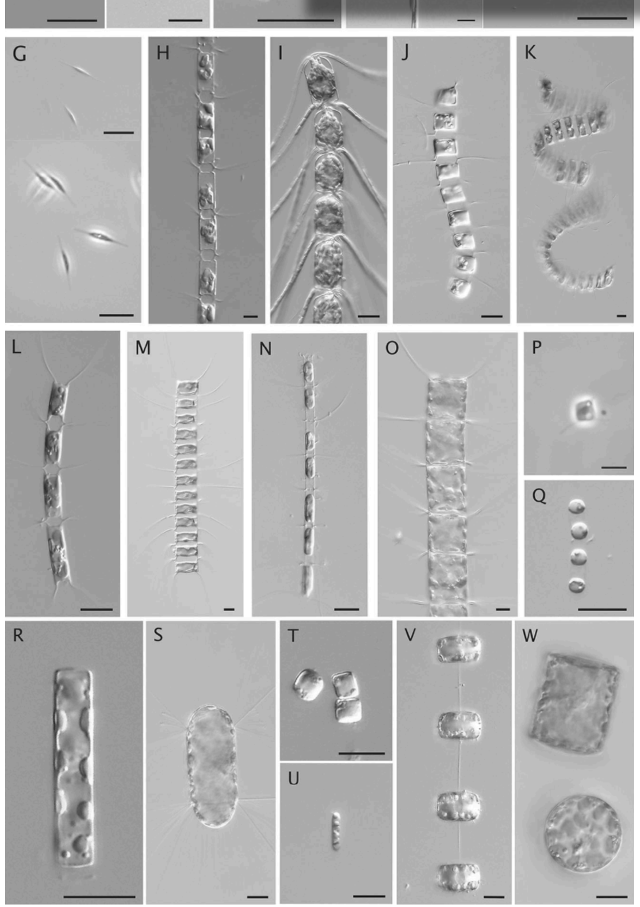
ELEMENTA
Science of the Anthropocene

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

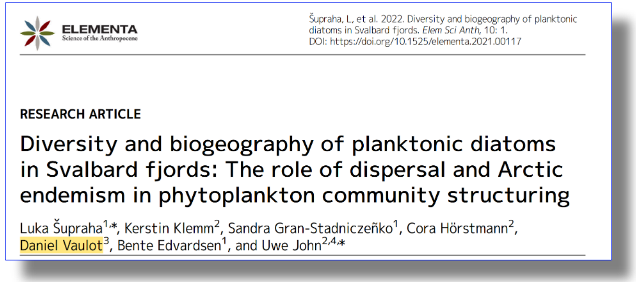
Diversity and biogeography of planktonic diatoms in Svalbard fjords: The role of dispersal and Arctic endemism in phytoplankton community structuring

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 phytoplankton community structuring. Elementa: Science of the Anthropocene 10, 00117

Biogeographic distribution types



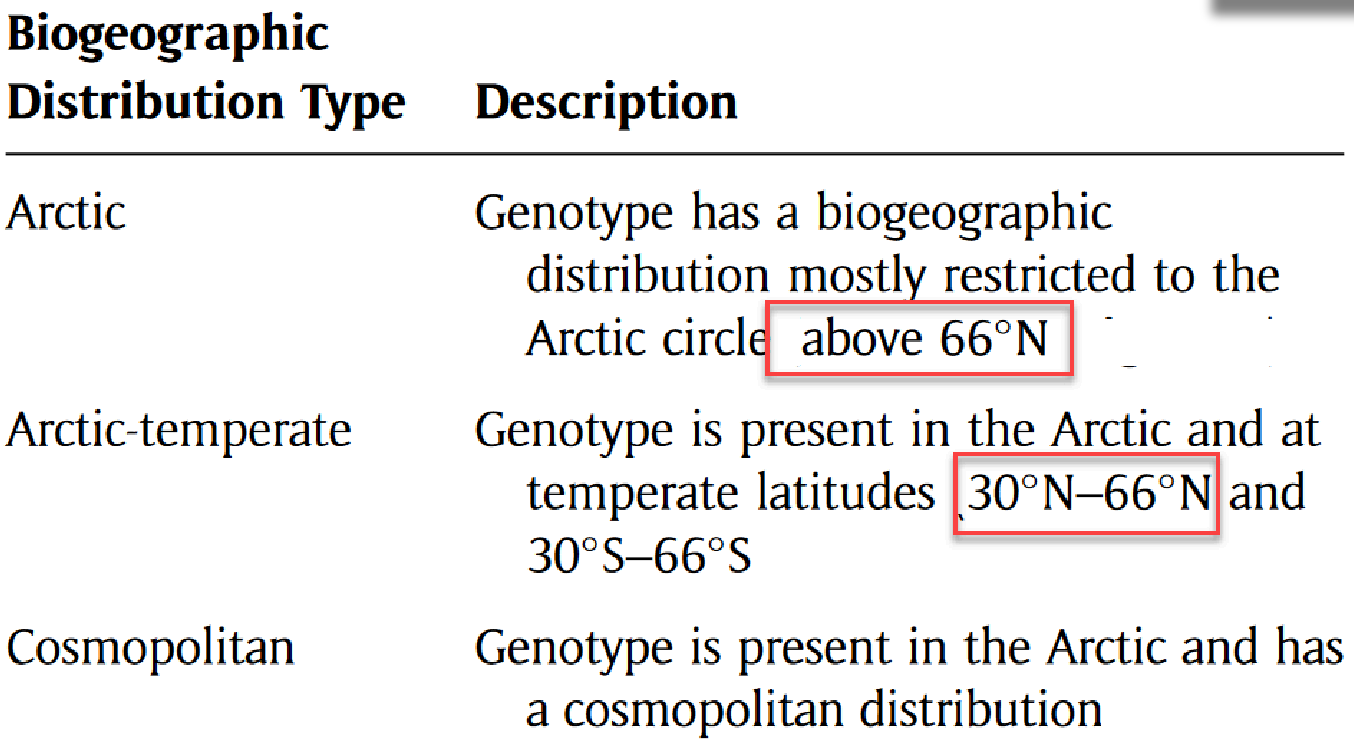
ELEMENTA
Science of the Anthropocene

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

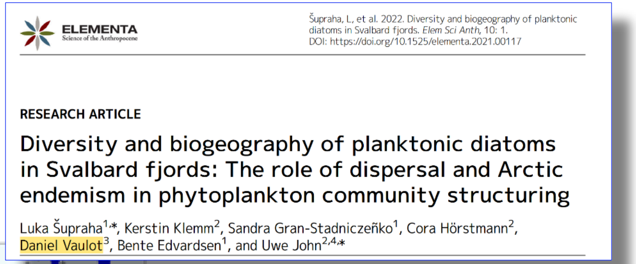
Luka Šupraha[1,*], Kerstin Klemm[2], Sandra Gran-Stadniczeñko[1], Cora Hörstmann[2], Daniel Vaulot[3], Bente Edvardsen[1], and Uwe John[2,4,*]



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

Šupraha, L. et al., 2022. Diversity and biogeography of planktonic diatoms in Svalbard fjords: the role of dispersal and Arctic endemism in phytoplankton community structuring. Elementa: Science of the Anthropocene 10, 00117

Polar genotypes



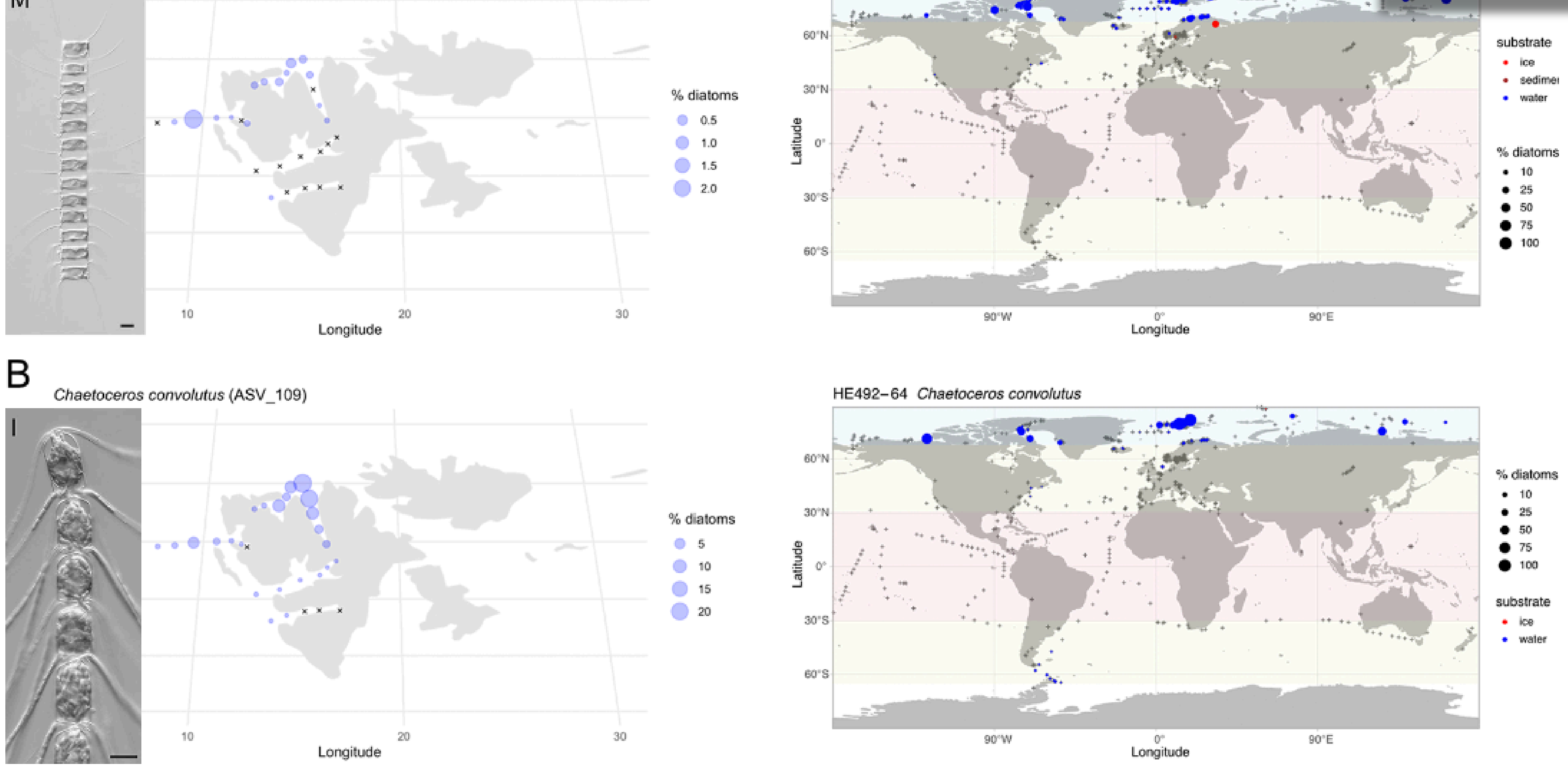
ELEMENTA
Science of the Anthropocene

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

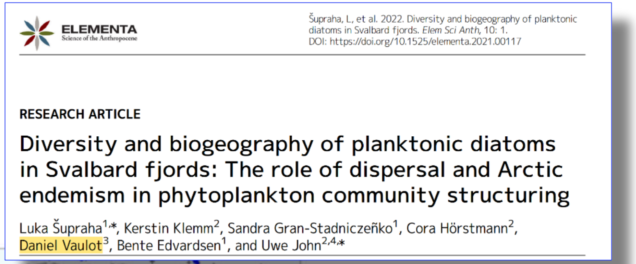
Diversity and biogeography of planktonic diatoms in Svalbard fjords: The role of dispersal and Arctic endemism in phytoplankton community structuring

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 phytoplankton community structuring. Elementa: Science of the Anthropocene 10, 00117

Cosmopolitan genotypes



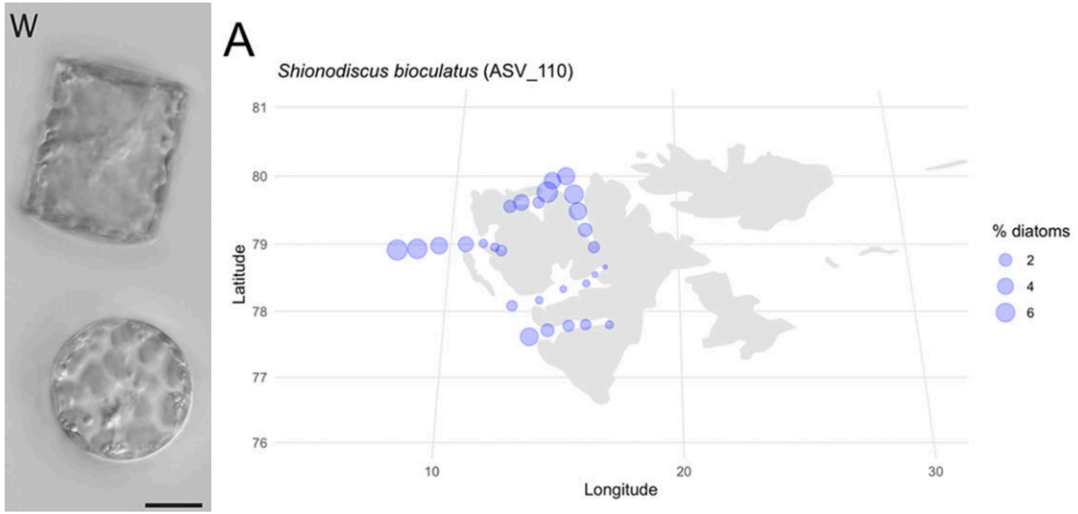
ELEMENTA
Science of the Anthropocene

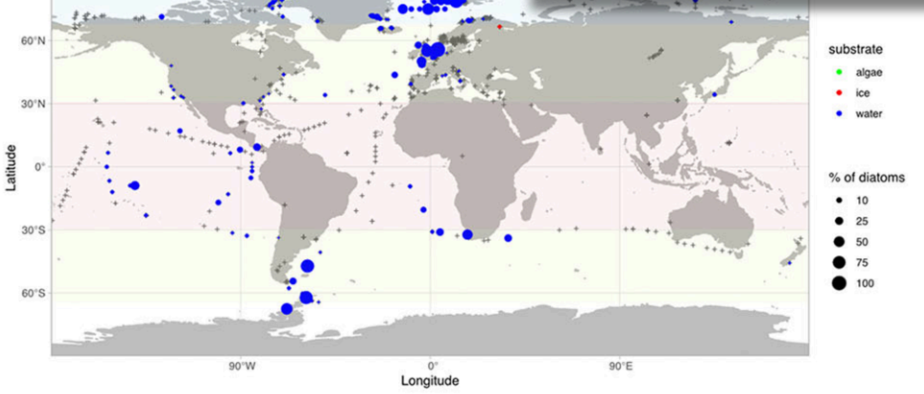
Šupraha, L., et al. 2022. Diversity and biogeography of planktonic diatoms in Svalbard fjords. *Elem Sci Anth*, 10: 1.
DOI: https://doi.org/10.1525/elementa.2021.00117

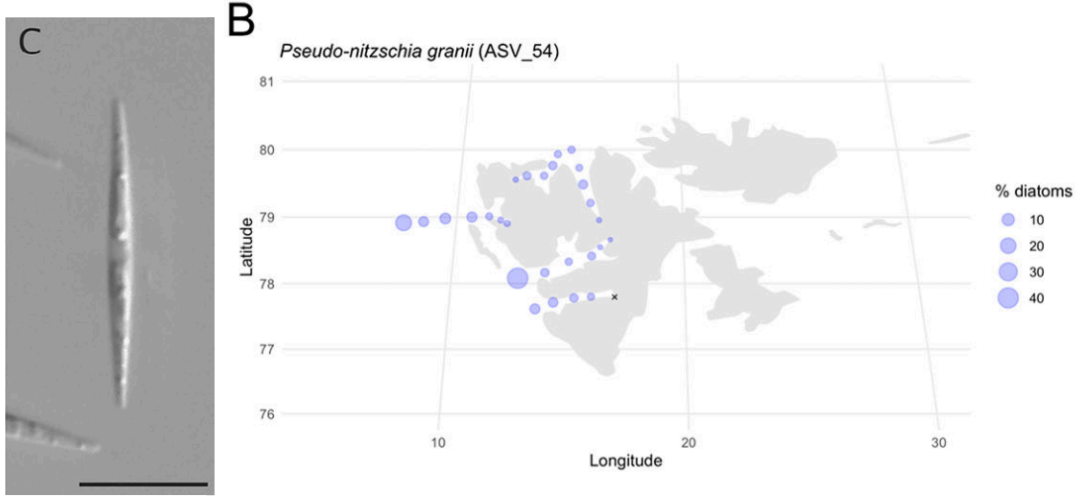
RESEARCH ARTICLE

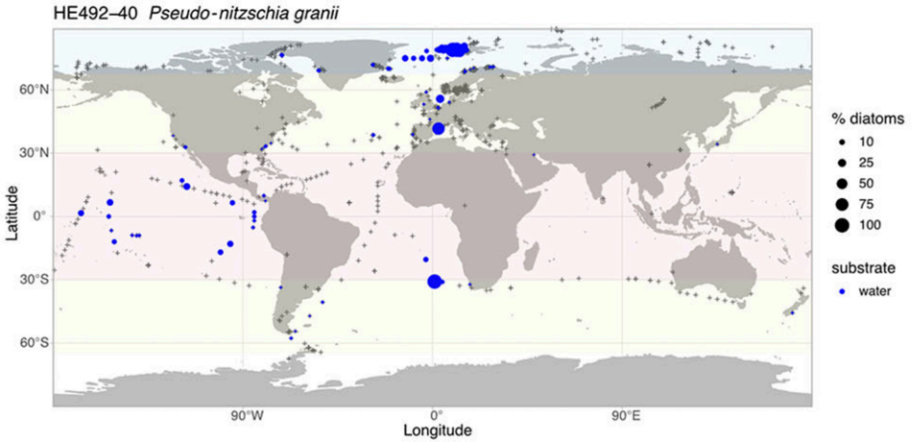
Diversity and biogeography of planktonic diatoms in Svalbard fjords: The role of dispersal and Arctic endemism in phytoplankton community structuring

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



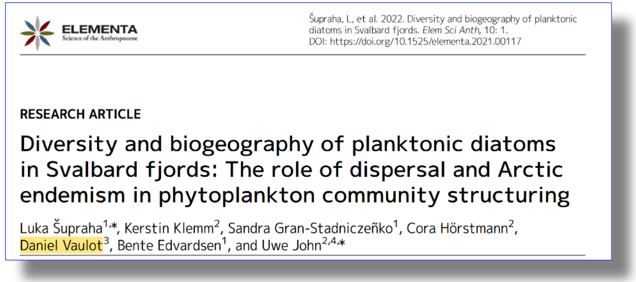






Šupraha, L. et al., 2022. Diversity and biogeography of planktonic diatoms in Svalbard fjords: the role of dispersal and Arctic endemism in phytoplankton community structuring. Elementa: Science of the Anthropocene 10, 00117

Distribution in Svalbard fjords

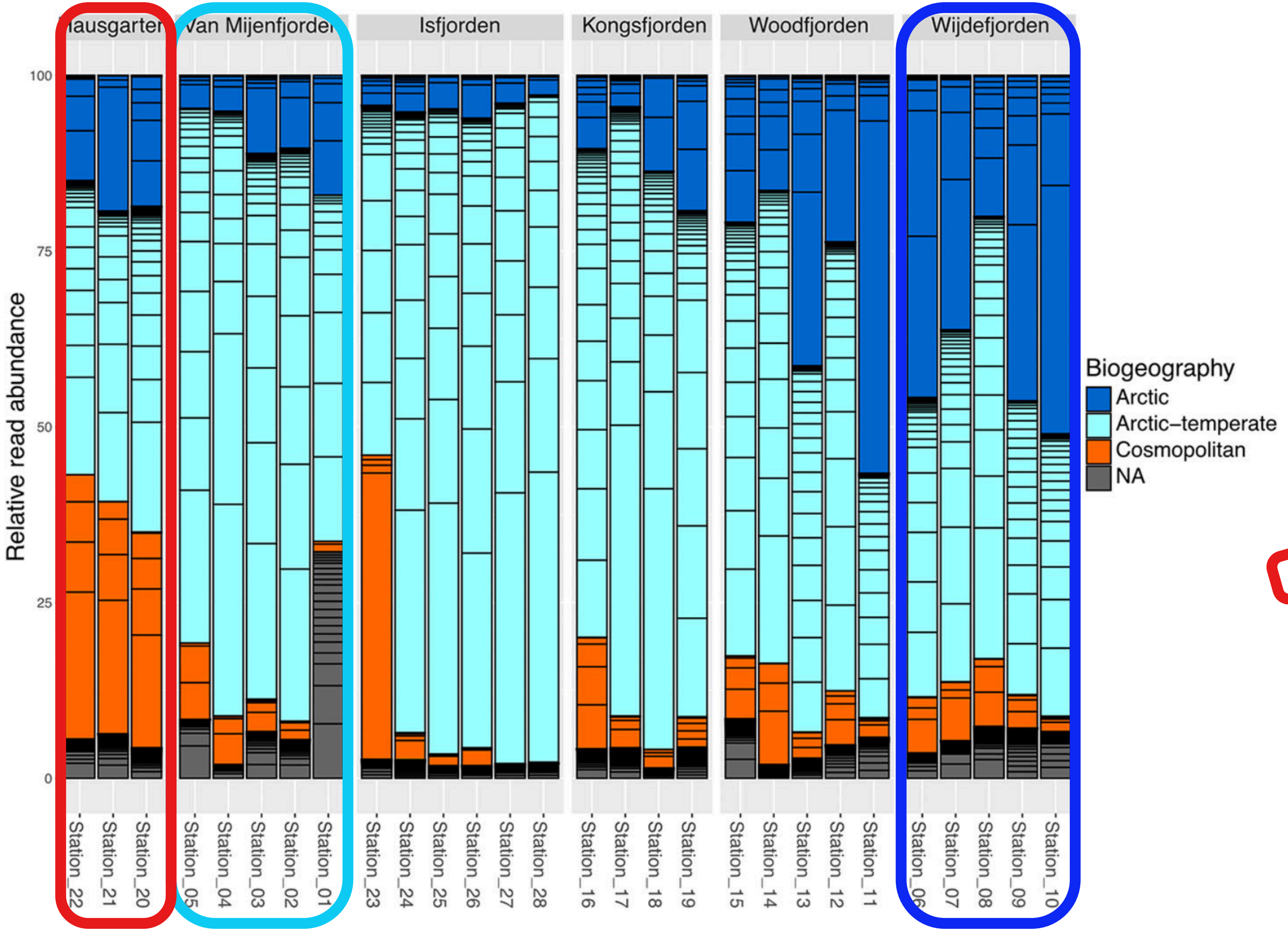

ELEMENTA
Science of the Anthropocene

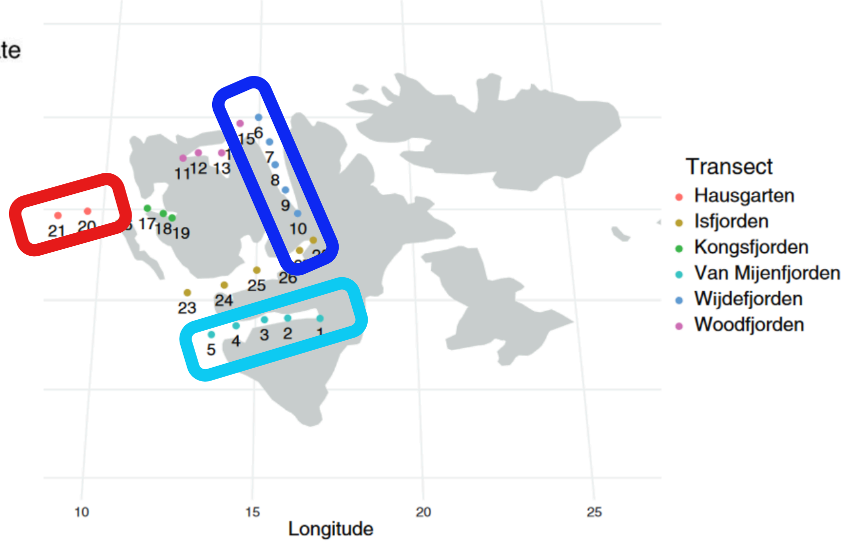
Šupraha, L, et al. 2022. Diversity and biogeography of planktonic diatoms in Svalbard fjords. *Elem Sci Anth*, 10: 1.
DOI: https://doi.org/10.1525/elementa.2021.00117

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 phytoplankton community structuring. Elementa: Science of the Anthropocene 10, 00117

What's next for metapr2

- More datasets
  - antarctic, tropical, equatorial
  - soils, sediments
  - freshwater
  - microbiomes
- Reannotate with PR² version 5.1
- Include long read metabarcodes ?

Thanks to all PR^2 contributors


Oslo 2003

Contributors PR2 reference


Javier del Campo
Institute of Evolutionary Biology UPF-CSIC
PR2 reference, EukRef


Mahwash Jamy
Swedish University of Agricultural Sciences, Uppsala
Eukaryotic phylogeny


Anna Maria Fiore-Donno
Universität zu Köln
Amoebozoa


Raphaël Morard
Marum, University of Bremen
Foraminifera

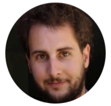

Miguel Mendez Sandin
Uppsala U., Sweden
Radiolaria


Alexandra Schönle
Universität zu Köln
Stramenopiles


Laure Arsenieff
CNRS, France
Diatoms


Ana Maria Cabello
IEO-Malaga
Pelagophyceae


Charles Bachy
CNRS, France
Ciliates


Tristan Biard
U. Littoral, France
Radiolaria


Vittorio Boscaro
UBC, Canada
Ciliates


Fabien Burki
Uppsala U., Sweden
Eukaryotic phylogeny


Andrea de Oliveira da Rocha Franco
FURG, Brazil
Diatoms


Johan Decelle
CEA, France
Plastids


Richard Dorrell
Institut de biologie de l'École normale supérieure, Paris
Plastid 16S


Manon Dünn
Universität zu Köln, Germany
Percolomonads


Bente Edvardsen
U. Oslo, Norway
Haptophyta


Maximilian Ganser
Paris-Lodron University of Salzburg, Austria
Ciliates


Chetan Gaonkar
Texas A&M, USA
Diatoms


Megan Gross
University of Kaiserslautern, Germany
Microsporidia


Haifeng Gu
Third Institute of Oceanography, China
Dinoflagellates


Adriana Lopes dos Santos
NTU, Singapore
Marine protists


Ramon Massana
ICM, Spain
Stramenopiles


Kenneth Mertens
IFREMER, France
Dinoflagellates


Sebastian Metz
Station Biologique, Roscoff
Perkinsids


Solenn Mordret
SZN, Italy
Dinoflagellates


Luciana Santoferrara
Hofstra University, USA
Ciliates


Se Hyeon Jang
Chonnam National University, Korea
Dinoflagellates


Алексей Селюк
Saint Petersburg State University
Chytrids


Pavel Škaloud
Charles University, Prague, CZ
Chrysophyceae


Margot Tragin
IFREMER, France
Chlorophyta


Wei-Ting Chen
NTO U., Taiwan
Ciliates

Contributors PR2 primers


Frédéric Mahé
CIRAD, France
Bioinformatics


David Bass
CEFAS, UK
Parasitic protists


Stefan Geisen
Wageningen U., The Netherlands
Soil protists

Core team


Daniel Vaulot
CNRS, France
PR2 coordinator, metaPR2


Javier del Campo
Institute of Evolutionary Biology UPF-CSIC
PR2 reference, EukRef

Scientific Committee


Frédéric Mahé
CIRAD, France
Bioinformatics


Fabien Burki
Uppsala U., Sweden
Tree of eukaryotes


Bente Edvardsen
U. Oslo, Norway
Haptophyta


Stefan Geisen
Wageningen U., The Netherlands
Soil protists


Laure Guillou
CNRS, France
PR2 reference


Martin Kolísko
CAS, Czech Republic
Apicomplexa


Adriana Lopes dos Santos
NTU, Singapore
Marine protists


Ramon Massana
ICM, Spain
Stramenopiles


Fabrice Not
CNRS, France
Radiolaria


Laura Parfrey
UBC, Canada
Protists


Ian Probert
Sorbonne Université, France
Roscoff Culture Collection


Luciana Santoferrara
Hofstra University, USA
Ciliates

G.E.E.K. LAB


Adriana Lopes dos Santos
Assistant Professor
✉ 🐦


Daniel Vaulot
Visiting Professor
✉ 🐦


Fabien Burki
Uppsala U., Sweden

PhD students


Clarence Sim
PhD student

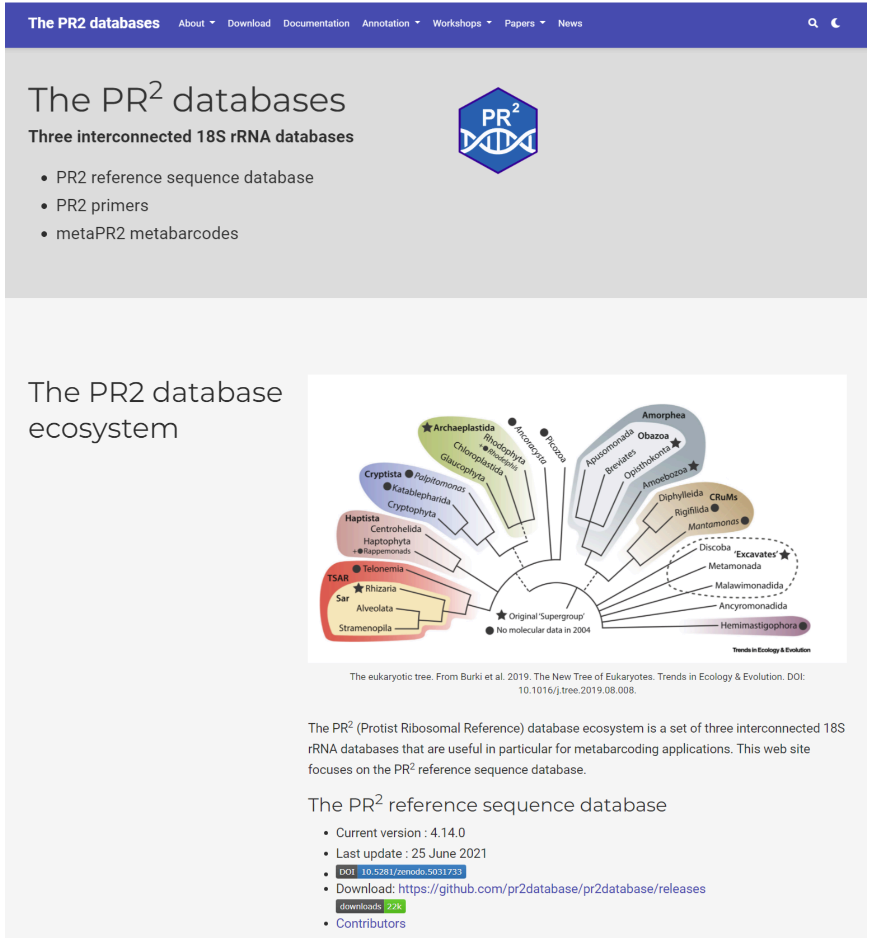

Denise Ong
PhD student


Bryan Teo
URECA Undergraduate student


Mahwash Jamy
Uppsala U., Sweden

Social

Web site: https://pr2-database.org/



The PR^2 databases

Three interconnected 18S rRNA databases

- PR2 reference sequence database
- PR2 primers
- metaPR2 metabarcodes

The PR2 database ecosystem

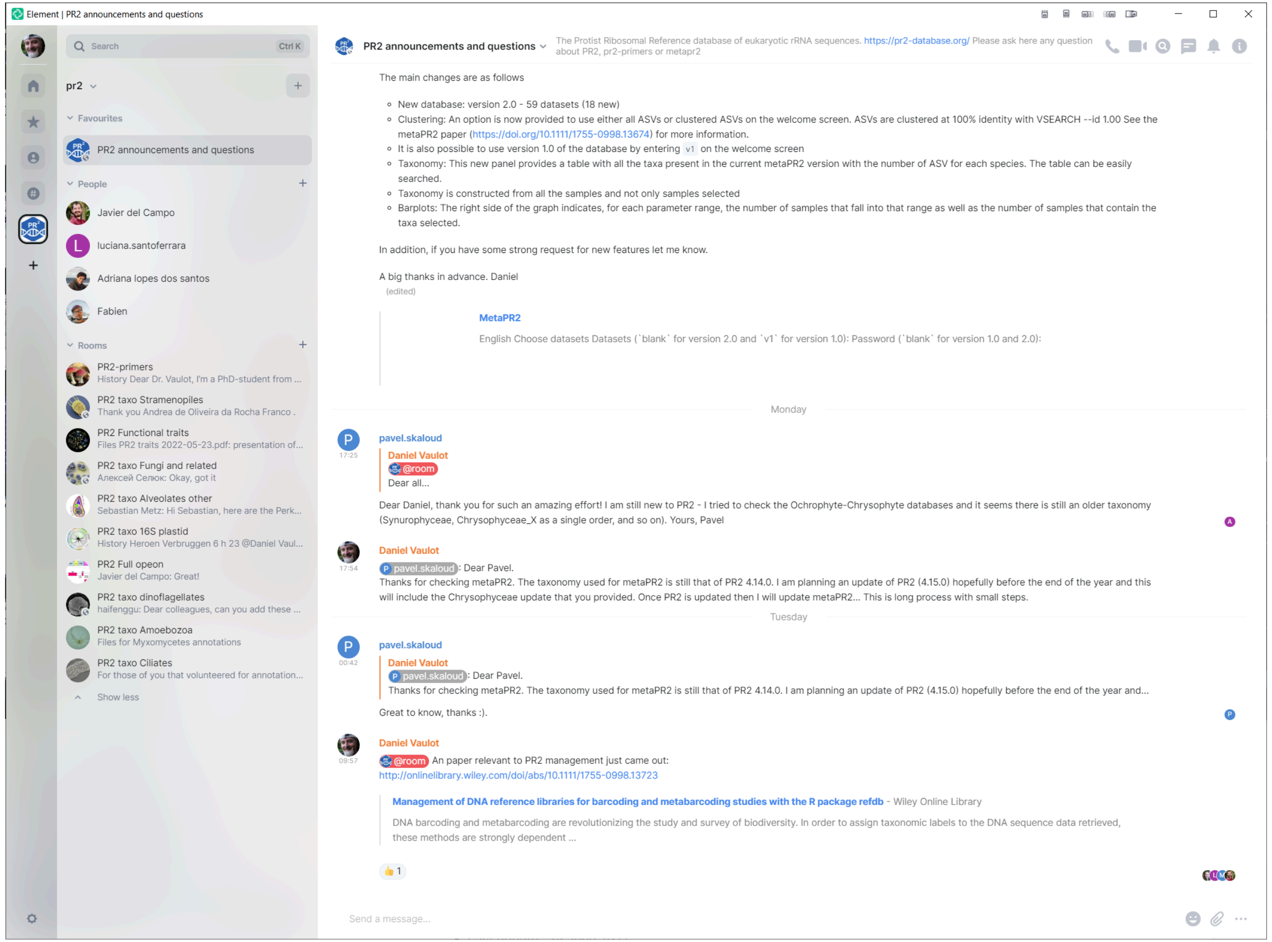
The eukaryotic tree. From Burki et al. 2019. The New Tree of Eukaryotes. Trends in Ecology & Evolution. DOI: 10.1016/j.tree.2019.08.008.

The PR^2 (Protist Ribosomal Reference) database ecosystem is a set of three interconnected 18S rRNA databases that are useful in particular for metabarcoding applications. This web site focuses on the PR^2 reference sequence database.

The PR^2 reference sequence database

- Current version : 4.14.0
- Last update : 25 June 2021
- DOI 10.5281/zenodo.5031733
- Download: https://github.com/pr2database/pr2database/releases
  downloads 22k
- Contributors

Moving from Slack: https://matrix.to/#/#pr2:matrix.org



PR2-primers
History Dear Dr. Vaulot, I'm a PhD-student from ...

PR2 taxo Stramenopiles
Thank you Andrea de Oliveira da Rocha Franco .

PR2 Functional traits
Files PR2 traits 2022-05-23.pdf: presentation of...

PR2 taxo Fungi and related
Алексей Селюк: Okay, got it

PR2 taxo Alveolates other
Sebastian Metz: Hi Sebastian, here are the Perk...

PR2 taxo 16S plastid
History Heroen Verbruggen 6 h 23 @Daniel Vaul...

PR2 Full opeon
Javier del Campo: Great!

PR2 taxo dinoflagellates
haifenggu: Dear colleagues, can you add these ...

PR2 taxo Amoebozoa
Files for Myxomycetes annotations

PR2 taxo Ciliates
For those of you that volunteered for annotation...

PR2 announcements and questions
The Protist Ribosomal Reference database of eukaryotic rRNA sequences. https://pr2-database.org/ Please ask here any question about PR2, pr2-primers or metapr2

The main changes are as follows

- New database: version 2.0 - 59 datasets (18 new)
- Clustering: An option is now provided to use either all ASVs or clustered ASVs on the welcome screen. ASVs are clustered at 100% identity with VSEARCH --id 1.00 See the metaPR2 paper (https://doi.org/10.1111/1755-0998.13674) for more information.
- It is also possible to use version 1.0 of the database by entering `v1` on the welcome screen
- Taxonomy: This new panel provides a table with all the taxa present in the current metaPR2 version with the number of ASV for each species. The table can be easily searched.
- Taxonomy is constructed from all the samples and not only samples selected
- Barplots: The right side of the graph indicates, for each parameter range, the number of samples that fall into that range as well as the number of samples that contain the taxa selected.

In addition, if you have some strong request for new features let me know.

A big thanks in advance. Daniel
(edited)

> **MetaPR2**
> English Choose datasets Datasets (`blank` for version 2.0 and `v1` for version 1.0): Password (`blank` for version 1.0 and 2.0):

Monday

pavel.skaloud 17:25
> **Daniel Vaulot**
> @room
> Dear all...

Dear Daniel, thank you for such an amazing effort! I am still new to PR2 - I tried to check the Ochrophyte-Chrysophyte databases and it seems there is still an older taxonomy (Synurophyceae, Chrysophyceae_X as a single order, and so on). Yours, Pavel

Daniel Vaulot 17:54
pavel.skaloud: Dear Pavel.
Thanks for checking metaPR2. The taxonomy used for metaPR2 is still that of PR2 4.14.0. I am planning an update of PR2 (4.15.0) hopefully before the end of the year and this will include the Chrysophyceae update that you provided. Once PR2 is updated then I will update metaPR2... This is long process with small steps.

Tuesday

pavel.skaloud 00:42
> **Daniel Vaulot**
> pavel.skaloud: Dear Pavel.
> Thanks for checking metaPR2. The taxonomy used for metaPR2 is still that of PR2 4.14.0. I am planning an update of PR2 (4.15.0) hopefully before the end of the year and...

Great to know, thanks :).

Daniel Vaulot 09:57
@room An paper relevant to PR2 management just came out:
http://onlinelibrary.wiley.com/doi/abs/10.1111/1755-0998.13723

> **Management of DNA reference libraries for barcoding and metabarcoding studies with the R package refdb** - Wiley Online Library
> DNA barcoding and metabarcoding are revolutionizing the study and survey of biodiversity. In order to assign taxonomic labels to the DNA sequence data retrieved, these methods are strongly dependent ...

👍 1