

# Metabarcoding data from *Tara* expeditions

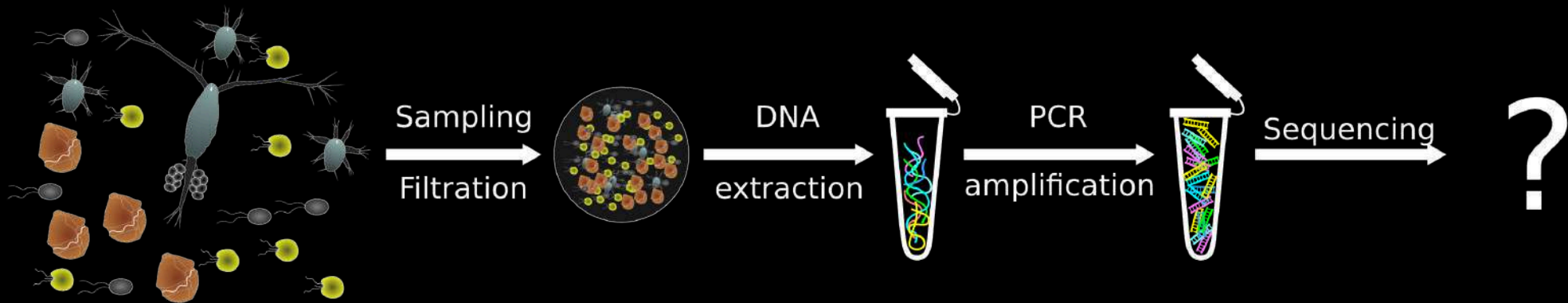
Nicolas Henry 

nicolas.henry@cnr.fr

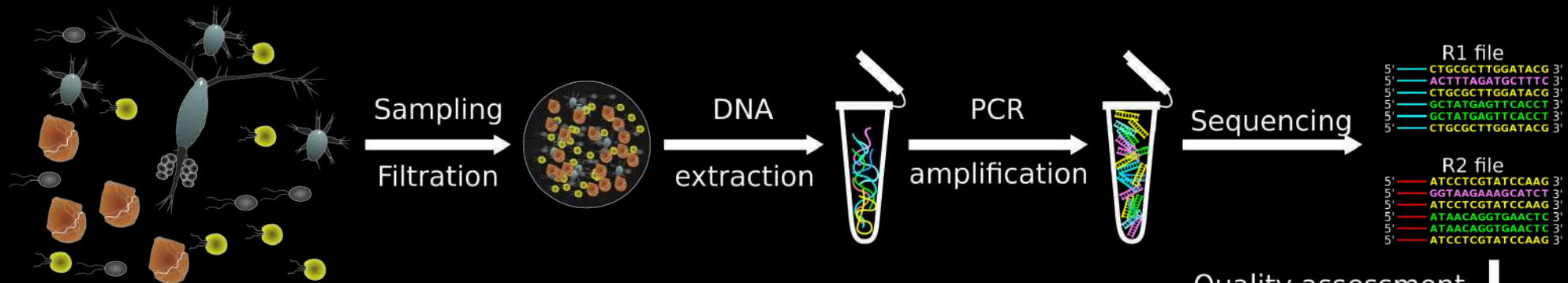
GOSEE, ABiMS, Station Biologique de Roscoff

# Metabarcoding

# Metabarcoding in the lab



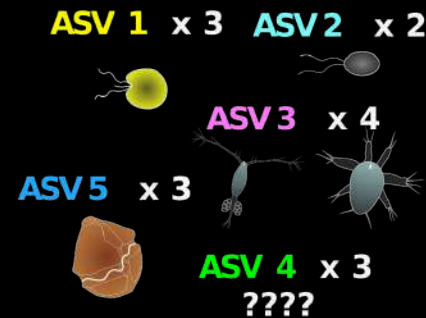
# MetaB bioinformatics



Quality assessment  
Demultiplexing  
Primer trimming  
Quality filtering  
ASVs inferring

taxonomy	sample 1	sample 2	sample 3
ASV 1 <i>Phaeocystis sp.</i>	3	0	10
ASV 2 <i>Telonema sp.</i>	2	5	1
ASV 3 <i>Oithona nana</i>	4	2	1
ASV 4 ????	3	1	1
ASV 5 <i>Blastodinium sp.</i>	3	8	1

ASV table  
← assembly



← Taxonomic  
assignment

ASV 1 x 3  
ASV 2 x 2  
ASV 3 x 4  
ASV 4 x 3  
ASV 5 x 3

# Metabarcoding data

# Contextual data

sample_id	event_label	protocol_label	date	latitude	longitude	temperature_ctd
SAMEA7917593	EVENT_20211207T0936Z_MMA-059_BGC_ROSETTE	gB023	2021-12-07 09:47:00	-40.3926	-55.6750	11.007
SAMEA7917588	EVENT_20211207T0936Z_MMA-059_BGC_ROSETTE	gB320	2021-12-07 09:47:00	-40.3926	-55.6750	11.007
SAMEA7917594	EVENT_20211207T0936Z_MMA-059_BGC_ROSETTE	gB023	2021-12-07 09:47:00	-40.3926	-55.6750	9.701
SAMEA7917589	EVENT_20211207T0936Z_MMA-059_BGC_ROSETTE	gB320	2021-12-07 09:47:00	-40.3926	-55.6750	9.701
SAMEA7917595	EVENT_20211207T0936Z_MMA-059_BGC_ROSETTE	gB023	2021-12-07 09:47:00	-40.3926	-55.6750	8.072
SAMEA7917590	EVENT_20211207T0936Z_MMA-059_BGC_ROSETTE	gB320	2021-12-07 09:47:00	-40.3926	-55.6750	8.072
SAMEA7917596	EVENT_20211207T0936Z_MMA-059_BGC_ROSETTE	gB023	2021-12-07 09:47:00	-40.3926	-55.6750	7.409
SAMEA7917591	EVENT_20211207T0936Z_MMA-059_BGC_ROSETTE	gB320	2021-12-07 09:47:00	-40.3926	-55.6750	7.409
SAMEA7917597	EVENT_20211207T0936Z_MMA-059_BGC_ROSETTE	gB023	2021-12-07 09:47:00	-40.3926	-55.6750	5.872
SAMEA7917592	EVENT_20211207T0936Z_MMA-059_BGC_ROSETTE	gB320	2021-12-07 09:47:00	-40.3926	-55.6750	5.872
SAMEA7905646	EVENT_20211207T1343Z_MMA-059_EHM_NET-20	S20	2021-12-07 13:46:00	-40.3980	-55.6796	9.722
SAMEA7917643	EVENT_20211209T0924Z_MMA-060_BGC_ROSETTE	gB320	2021-12-09 09:32:00	-42.2362	-57.2352	10.329
SAMEA7917646	EVENT_20211209T0924Z_MMA-060_BGC_ROSETTE	gB023	2021-12-09 09:32:00	-42.2362	-57.2352	10.320
SAMEA7917645	EVENT_20211209T0924Z_MMA-060_BGC_ROSETTE	gB320	2021-12-09 09:32:00	-42.2362	-57.2352	10.320
SAMEA7917648	EVENT_20211209T0924Z_MMA-060_BGC_ROSETTE	gB023	2021-12-09 09:32:00	-42.2362	-57.2352	9.529
SAMEA7917647	EVENT_20211209T0924Z_MMA-060_BGC_ROSETTE	gB320	2021-12-09 09:32:00	-42.2362	-57.2352	9.529

# Raw data

@V1:1:HGVH5DRX3:1:2101:10004:14685/2

ACTTTCGTTATTGATTAATGTGAACGTCCTTGTTTCGACGCGTTCGCAGCAGTTCGTCTTCCATAAATCCAAGACTTTCAI

+

F,FFFFFFF,FF,F,F:,FF,,F,,,F,FFFFFF:,,:FFF:FFF,F:,:FF,,FF,,,,FF,F,,FFF:FFF,,FFFFFFI

@V1:1:HGVH5DRX3:1:2101:10004:16939/2

CCAGCAGCTGCGGTAATTCCAGCTCCAATAGCGTATATTAAGTTGTTGCGGTTAAAAGCTCGTAGTTGGATTTCTGCI

+

FFI

@V1:1:HGVH5DRX3:1:2101:10004:18192/2

ACTTTCGTTCTTGATCAATGAAAACATCCTTGGCAAATGCTTTCGCACCTTGTTTGTTCGTAACGCAATCTAAGAATTCCA

+

FF:FFFFFF:FFFFFF:,FFFFFFI

@V1:1:HGVH5DRX3:1:2101:10004:20322/2

CCAGCAGCTGCGGTAATTCCAGCTCCAATAGCGTATATTAAGTTGTTGCGGTTAAAAGCTCGTAGTTGGAGTTCTGCI

+

FFFFFFFFFFFFFF.FFF FFFFFFFFFFFFFFF.FFF.FF.FFFFFFFFFI

# ASV table

asv_id	family	species	confidence	SAMEA7906352	SAMEA7908523
ea64ae1d6dddc2146105a882229c81bb	Maxillopoda	Oithona_similis	0.95	6567	89
bdcfdb40ee97fc620dafa95637fcb13	Phaeocystaceae	Phaeocystis_antarctica	1.00	2142	2
7e3fa204e95db351839d719215fba508	Mamiellaceae	Micromonas_commoda_A2	0.91	127	1419
2ccf1f15767862d5c6819988bf8e33ff	Noelaerhabdaceae	Gephyrocapsa_oceanica	1.00	227	374
186f848bd700e36dc071bc9f97276862	Parodontellaceae	Trieres_sp.	0.57	0	0
aba6dae61f5ddf30d17d63fbb226e526	Phaeocystaceae	Phaeocystis_antarctica	1.00	9	0
4ff66f8696a79c4dca4499d9957449ec	Telonemia-Group-1	Telonemia-Group-1_X_sp.	1.00	5	0
8704308146a1ea2ec2508c4d7f7c7cd3	Corethraceae	Corethron_inerme	0.94	0	0
dd9ff45aad9d327fd16075f1ff362f4c	Dino-Group-I-Clade-1	Dino-Group-I-Clade-1_X_sp.	1.00	36	0
a03e7a6cec32a7f1268aef5cdfef17ce	NA	NA	1.00	0	0
5aeaacb4835e86de247998ffbc34107c	Florenciellales	Pseudochattonella_verruculosa	0.97	3711	298
3bb4c24b69396b8d5cf445e7d2bd580c	Maxillopoda	Calanoides_carinatus	0.50	0	0
2989f5043374984685a0d71bb2292967	NA	NA	1.00	38	1
c836f387c69d729d9e303c2a7a1efd70	MAST-2	MAST-2D_sp.	1.00	1	3
5b2f9bda94dd54330d8de22256613968	Gymnodiniaceae	Gyrodinium_helveticum	0.76	16	0
c453a50385c8dc9dea8d2f59caa98cb0	MAST-3	MAST-3D_sp.	1.00	0	0



# Metabarcoding data from Tara expeditions

# The expeditions

Tara Oceans



P2 pilot project



Tara Microplastics



Tara Europa



Tara Mediterranean



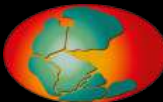

Tara Pacific



Tara Mission Microbiomes



# Contextual data

- First type of data released
- Major asset of Tara expeditions
- In Pangaea  /zenodo/  BioSamples depending on expeditions

Data Descriptor | [Open access](#) | [Published: 26 May 2015](#)

## Open science resources for the discovery and analysis of Tara Oceans data

[Stéphane Pesant](#) , [Fabrice Not](#), [Marc Picheral](#), [Stefanie Kandels-Lewis](#), [Noan Le Bescot](#), [Gabriel Gorsky](#), [Daniele Iudicone](#), [Eric Karsenti](#), [Sabrina Speich](#), [Romain Troublé](#), [Céline Dimier](#), [Sarah Seanson](#) & [Tara Oceans Consortium Coordinators](#)

[Scientific Data](#) **2**, Article number: 150023 (2015) | [Cite this article](#)

27k Accesses | 232 Citations | 48 Altmetric | [Metrics](#)

SARDET Station 860

Tara-OCEANS WATER SAMPLING EVENT LOG-SHEET

TARA OCEANS

SCIENTIST NAME: SARDET / Margaux, Céline, Romain, Noan, Stéphane STB

DATE: UTC/LOCAL: 27/09/2014 TIME: UTC/LOCAL: 06:30  
 LATITUDE: N/S: 38° 00' 22" N Start: 11:30  
 LONGITUDE: E/W: 3° 59' 04" E End: 11:30  
 CAST DEPTH: (m): 0 - 40 m DCM = 41 m

INSTRUMENTS:  Rosette  CTD  FLUO   
 8L  UVP  NEPH   
 12L  CSTAR  NO<sub>3</sub>


ECOTriplet Barcode

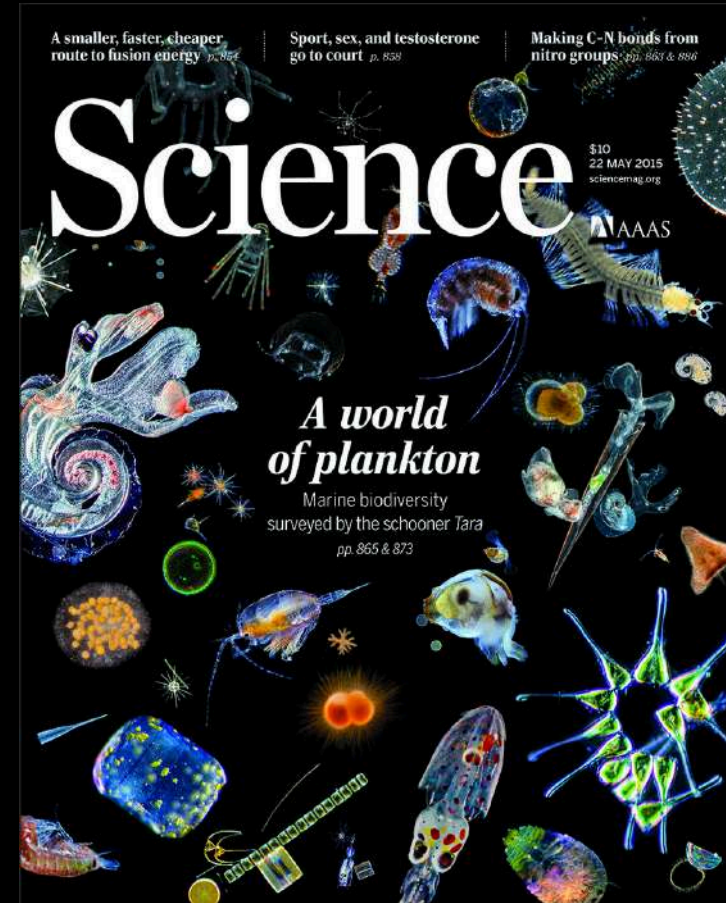
EVENT Barcode	NUTRIENTS (80 mL)	CARBONATE (3x500 mL)	HPLC (2000 mL)	COMMENTS
EVENT Barcode				
SAMPLE Bottle 1 (400 m)	↑	G19 G20 G23	↑	DCM = 41 m Top DCM Bottom DCM Bottom ML
SAMPLE Bottle 2 (300 m)	—	—	—	
SAMPLE Bottle 3 (200 m)	—	—	—	
SAMPLE Bottle 4 (185 m)	—	—	—	
SAMPLE Bottle 5 (140 m)	OK	—	—	
SAMPLE Bottle 6 (85 m)	—	—	—	
SAMPLE Bottle 7 (40 m)	—	—	—	
SAMPLE Bottle 8 (20 m)	—	—	—	
SAMPLE Bottle 9 (15 m)	—	—	—	
SAMPLE Bottle 10 (3 m)	↓	G24 G25 G26	↓	

Double Sided Adhesive

Tara-OCEANS-LogSheets(04092009).doc 04/09/2009  
 A. Van colson, Hervé Bournaud / Céline

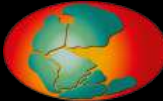
# Raw data

- Sequenced at 
- Raw data submitted at   
European Nucleotide Archive
- When first article using the data is submitted.



First metabarcoding FASTQ files published the 15th of July 2014 (ERA329958) and paper

# OTUs / ASV tables

- First OTU table > 10 years ago by Stéphane Audic
- Multiple tables (marker genes, expeditions)
- Multiple versions (workflows, sequencing batch), at least 4 versions only for Tara Oceans 18S V9
- In Pangaea  /zenodo

# Not that easy to find

The screenshot shows the Zenodo search results for the query 'tara oceans V9 metabarcoding'. The search bar at the top left contains the query, and the results show 7,645 items found. The results are sorted by 'Best match'. The first result is a dataset titled 'Summarized contextual data about metabarcoding Tara Oceans samples (2009-2013)' by Henry, Nicolas and Salazar, Guillem, uploaded on October 20, 2022. The second result is a dataset titled 'ASV Tables inferred by DADA2 from the TARA Oceans v9 metabarcoding dataset' by Benjamin Callahan, uploaded on May 21, 2017. The left sidebar contains filters for 'Versions', 'Access status', and 'Resource types'.

zenodo tara oceans V9 metabarcoding Communities My dashboard nicolas.he...

7,645 result(s) found Sort by Best match

**Versions**

View all versions

**Access status**

Open

Restricted

Embargoed

**Resource types**

Publication

Dataset

October 20, 2022 (v1.0.1) Dataset Open

**Summarized contextual data about metabarcoding Tara Oceans samples (2009-2013)**

Henry, Nicolas ; Salazar, Guillem

Tab-separated values table describing the metabarcoding samples from the expedition Tara Oceans (2009-2013). Information such as depth, time, geographic position, size fraction, collected from Pangaea, are listed in context\_general tables. In context\_stat tables, you will find a selection of physico-chemical parameters. These tables have bee...

Uploaded on October 20, 2022

1 more versions exist for this record

282 123

May 21, 2017 (v1) Dataset Open

**ASV Tables inferred by DADA2 from the TARA Oceans v9 metabarcoding dataset**

Benjamin Callahan


Tables of exact amplicon sequence variants (ASVs) were generated from the TARA Oceans metabarcoding data (~766 million reads from 334 plankton samples, V9 region of the 18S rRNA gene, DOI: 10.5281/zenodo.15600) by DADA2 on a 2016 Macbook Pro. The R script used to process the data is included, alongside 3 ASV tables: the observed ASVs before chim...

Uploaded on May 21, 2017

1348 883

**FAIR with a bigger F**

# Companion website (2015)

DATA CONTACT US

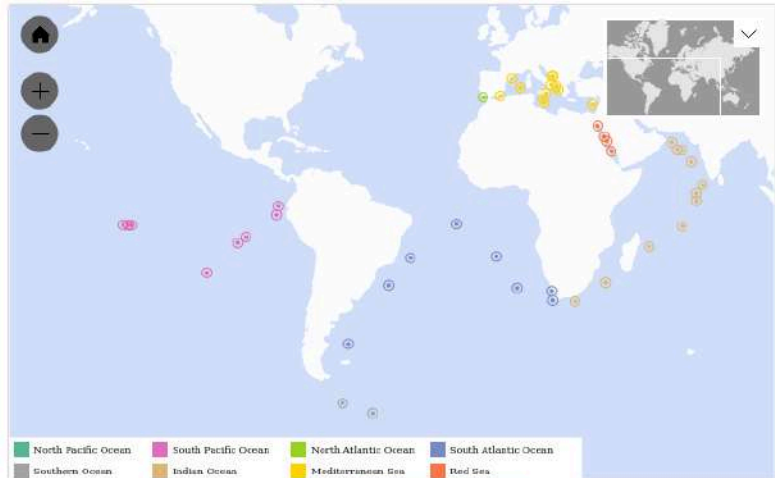
[Companion Website to the article:](#)

## "Eukaryotic plankton diversity in the sunlit ocean".

First Tara Oceans V9 rDNA metabarcoding dataset.

Eukaryotic plankton diversity in the sunlit ocean ; De Vargas, Audic, Henry, et al. DOI: 10.1126/science.1261605

### INTERACTIVE MAP OF SAMPLING STATIONS



Legend:  
North Pacific Ocean (Green), South Pacific Ocean (Pink), North Atlantic Ocean (Light Green), South Atlantic Ocean (Blue)  
Southern Ocean (Grey), Indian Ocean (Orange), Mediterranean Sea (Yellow), Red Sea (Red)

#### Detailed Materials and Methods

- Sampling of size-fractionated eukaryotic plankton communities
- DNA extraction, PCR amplification, and sequencing of 18S-V9 rDNA metabarcodes
- Sequence data cleaning, filtering, and clustering
- The V9\_PR2 reference database and its use for taxonomic assignment
- Phylogenetic analyses of known reference V9 rDNA barcodes and *Tara Oceans* metabarcodes for each major eukaryotic lineage

[Companion Website Texts \(text W1 to W5\)](#)

[Companion Website Figures \(W1 to W14\)](#)

[Companion Website Databases \(W1 to W9\)](#)

Map of 47 *Tara Oceans* sampling stations in (de Vargas et al. 2015). Clicking on a station, a list of genetic samples which were sequenced will pop up, giving access to the characteristics of the sampling, the raw sequence data at the EBI and a Krona interactive pie chart allowing exploration of the sample.

### Companion Website

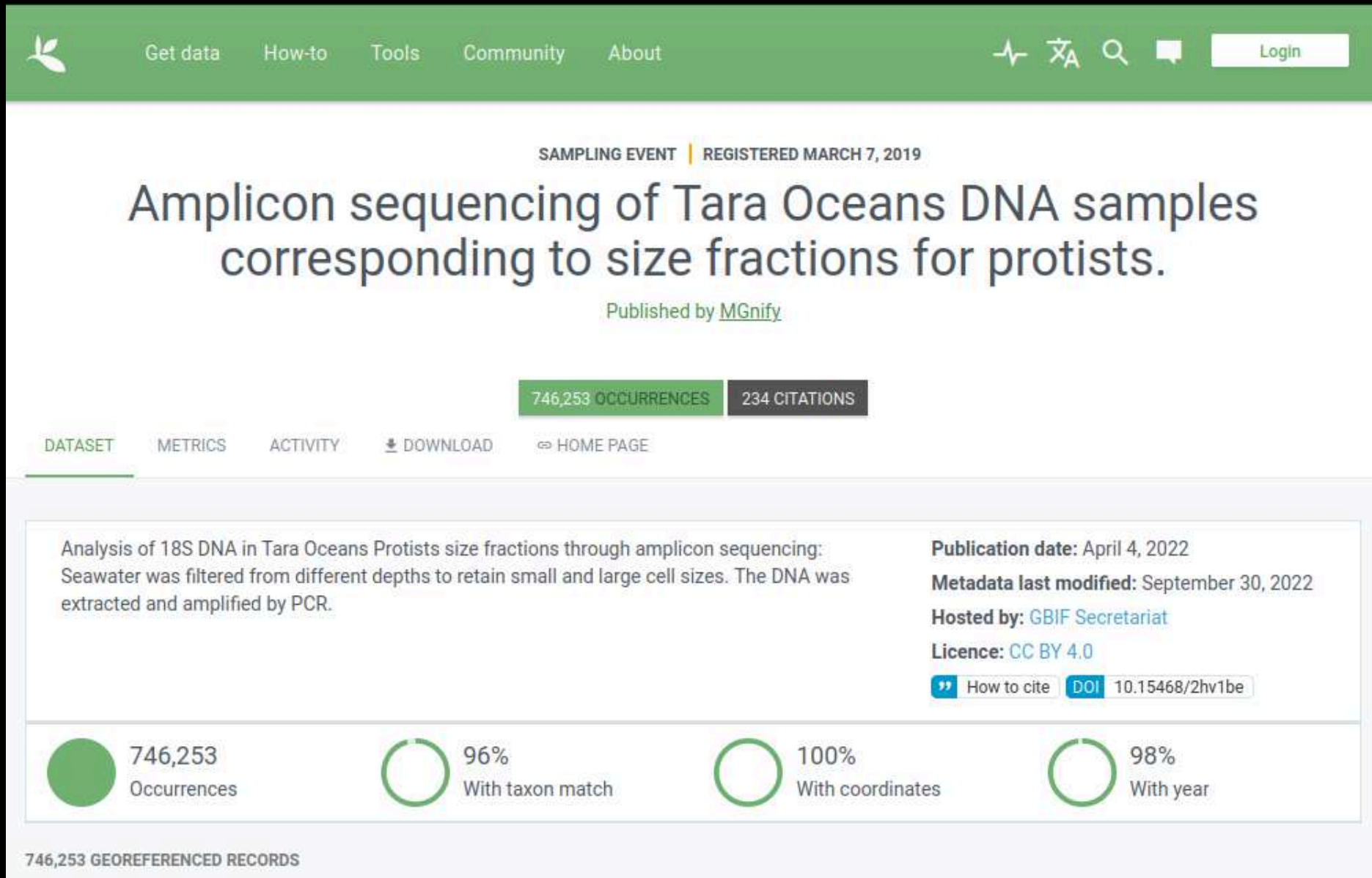
This is accompanying material, text, data and figures for the article "Eukaryotic plankton diversity in the sunlit ocean" (de Vargas et al. 2015). You can also directly download a pdf file with all this material ([SupplementaryMaterial.pdf](#), also archived at zenodo under DOI: 10.5281/zenodo.15600).

- i. Detailed Materials and Methods
- ii. Companion Website Texts (text W1 to W5)
- iii. Companion Website Figures (W1 to W14)
- iv. Companion Website Databases (W1 to W9)
- v. References
- vi. *Tara Oceans* Coordinators
- vii. *Tara Oceans* Contributors
- viii. Contacts







# Ocean Barcode Atlas

The screenshot displays the Ocean Barcode Atlas (OBA) website. At the top left, the OBA logo is shown next to the text "OCEAN BARCODE ATLAS" and "ONE CLICK MARINE BARCODE BIOGEOGRAPHY". A navigation sidebar on the left includes "New job", "Home page", and "Help". The top right features logos for "Fondation taraocéan", "mio", "OCEANOMICS", and "GENOSCOPE". The main content area shows a world map with a "Description" dropdown menu. Below the map are two blue buttons: "Community ecological analysis" and "Sequence based query". The bottom of the page features four species icons: "Species 2" (a blue fish), "Species 1" (an orange coral), "Species 7" (a yellow starfish), and "Species 4" (a green seaweed).



The screenshot shows a GBIF dataset page with a green header. The main title is 'Amplicon sequencing of Tara Oceans DNA samples corresponding to size fractions for protists.' It is published by MGnify and registered on March 7, 2019. The page displays 746,253 occurrences and 234 citations. A navigation bar includes links for DATASET, METRICS, ACTIVITY, DOWNLOAD, and HOME PAGE. A description of the dataset is provided, along with publication details: date (April 4, 2022), last modified (September 30, 2022), hosted by GBIF Secretariat, and licensed under CC BY 4.0. A citation link is also present. At the bottom, four circular progress indicators show: 746,253 Occurrences (100%), 96% With taxon match, 100% With coordinates, and 98% With year. A footer indicates 746,253 GEOREFERENCED RECORDS.



Get data How-to Tools Community About     Login

SAMPLING EVENT | REGISTERED MARCH 7, 2019


## Amplicon sequencing of Tara Oceans DNA samples corresponding to size fractions for protists.





Published by [MGnify](#)

746,253 OCCURRENCES 234 CITATIONS

DATASET METRICS ACTIVITY  DOWNLOAD  HOME PAGE

Analysis of 18S DNA in Tara Oceans Protists size fractions through amplicon sequencing: Seawater was filtered from different depths to retain small and large cell sizes. The DNA was extracted and amplified by PCR.

**Publication date:** April 4, 2022  
**Metadata last modified:** September 30, 2022  
**Hosted by:** [GBIF Secretariat](#)  
**Licence:** [CC BY 4.0](#)  
 [How to cite](#) [DOI](#) [10.15468/2hv1be](#)

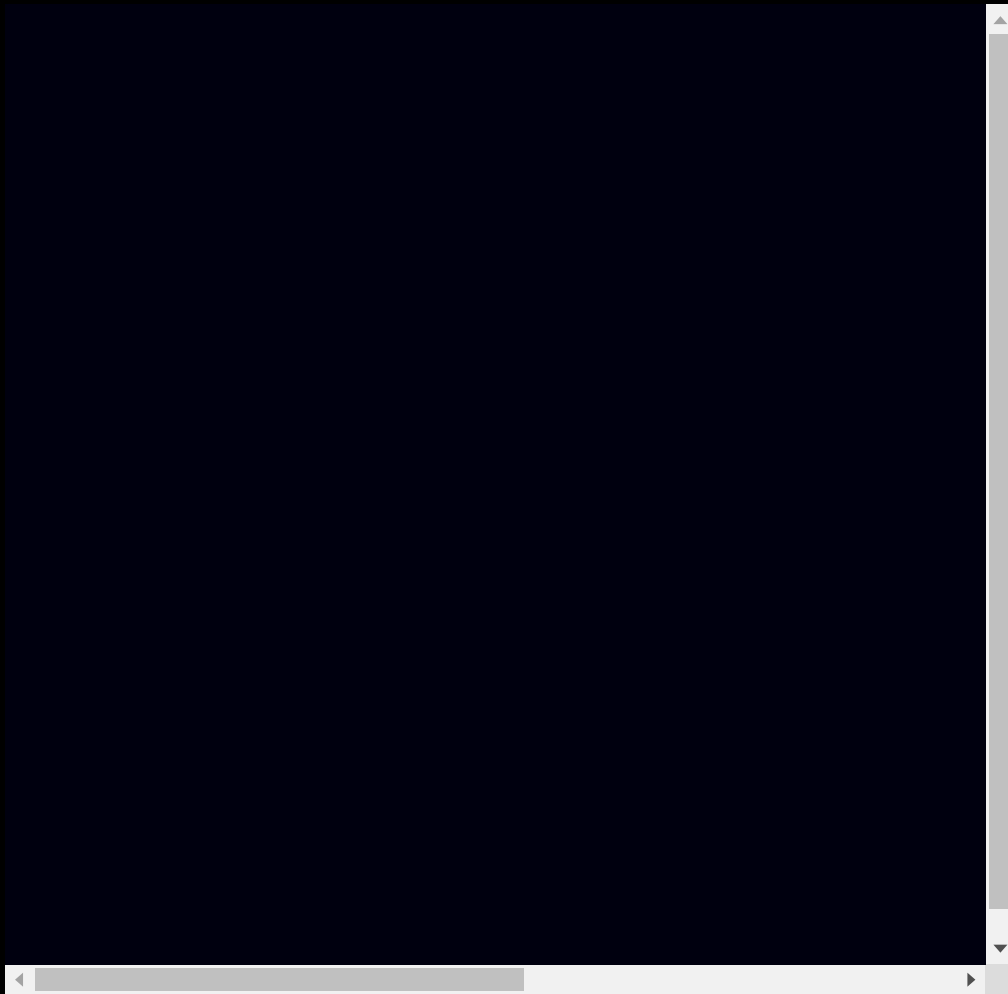
 746,253 Occurrences  96% With taxon match  100% With coordinates  98% With year

746,253 GEOREFERENCED RECORDS

# One place for all metaB data

- A static website hosted at SBR
- For each project, links to:
  - up to date ASV table and other versions
  - contextual data
  - raw data
  - relevant publication

# Landing page



- [Tara Oceans](#) (3214 samples)
- [Plankton Planet Pilot Project](#) (214 samples)
- [Tara Pacific](#) (15417 samples)
- [Tara Mission Microbiome](#) (447 samples)

# Project page

Metabarcoding data from Tara expeditions

[Home](#)

[Tara expeditions](#) ▾

[Related projects](#) ▾

[About](#)

Tara Oceans

Tara Pacific

Tara Mission Microbiome

## Tara Oceans

## Sampling

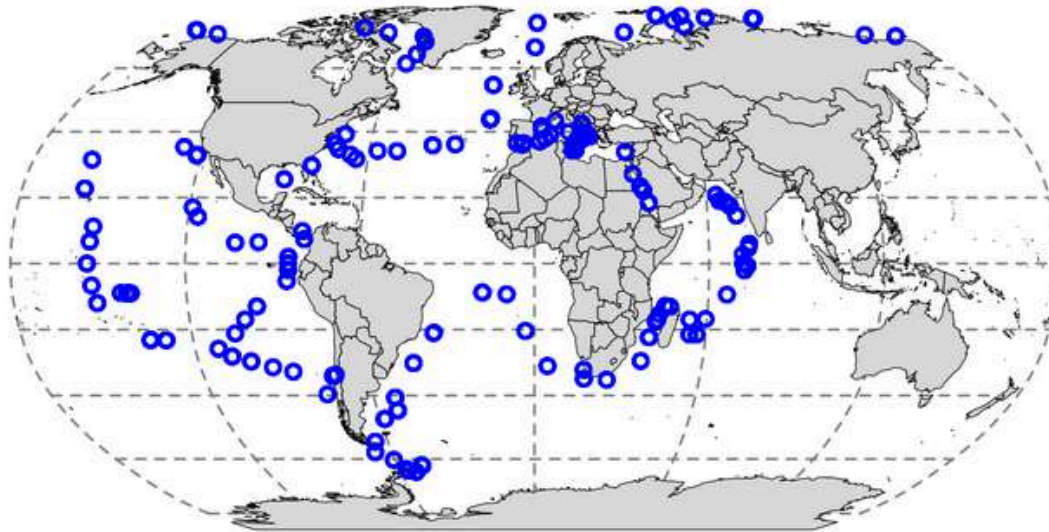
On this page

[Sampling](#)

[Raw data](#)

[ASV tables](#)

[Related project:](#)



**More soon!!**

