

Capturing biodiversity using molecular data

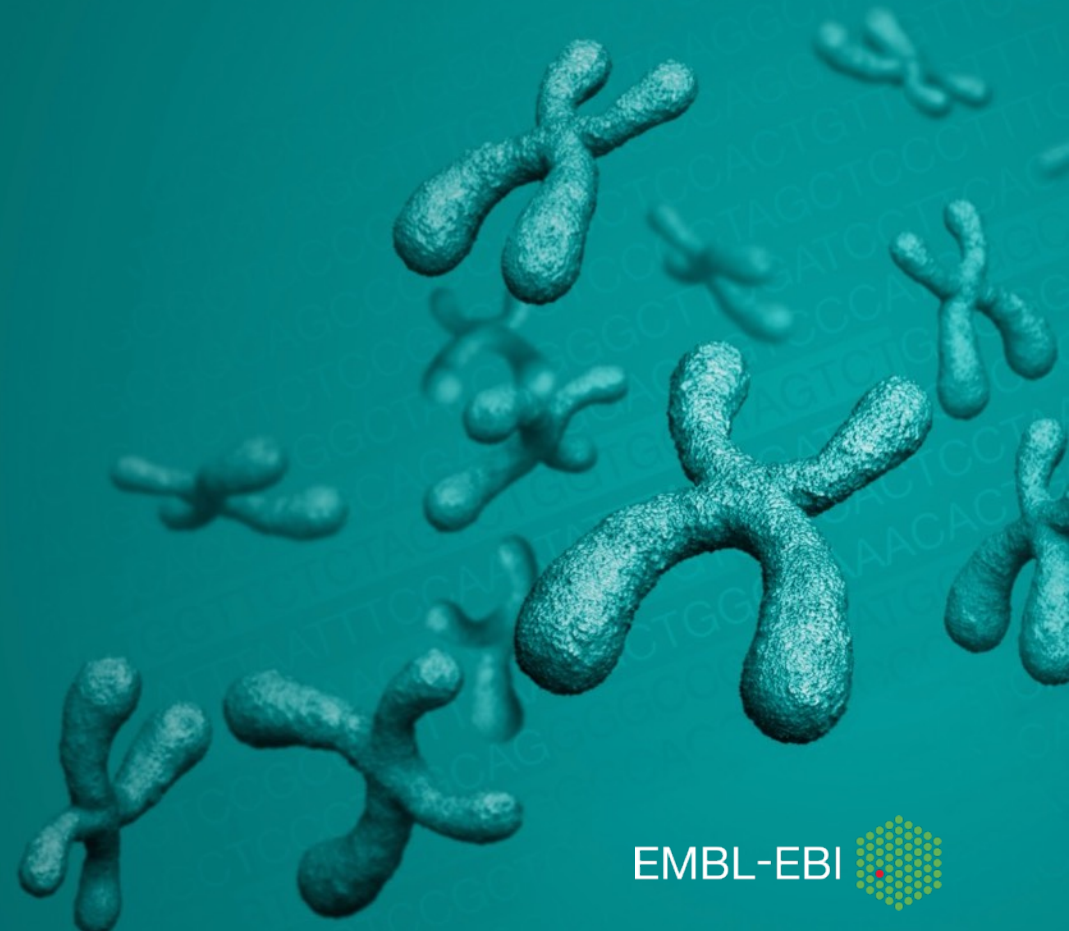
GBIF Impact and Action

9th December 2021

Rob Finn (rdf@ebi.ac.uk, @robdfinn)

European Molecular Biology Laboratory

European Bioinformatics Institute (EMBL-EBI)

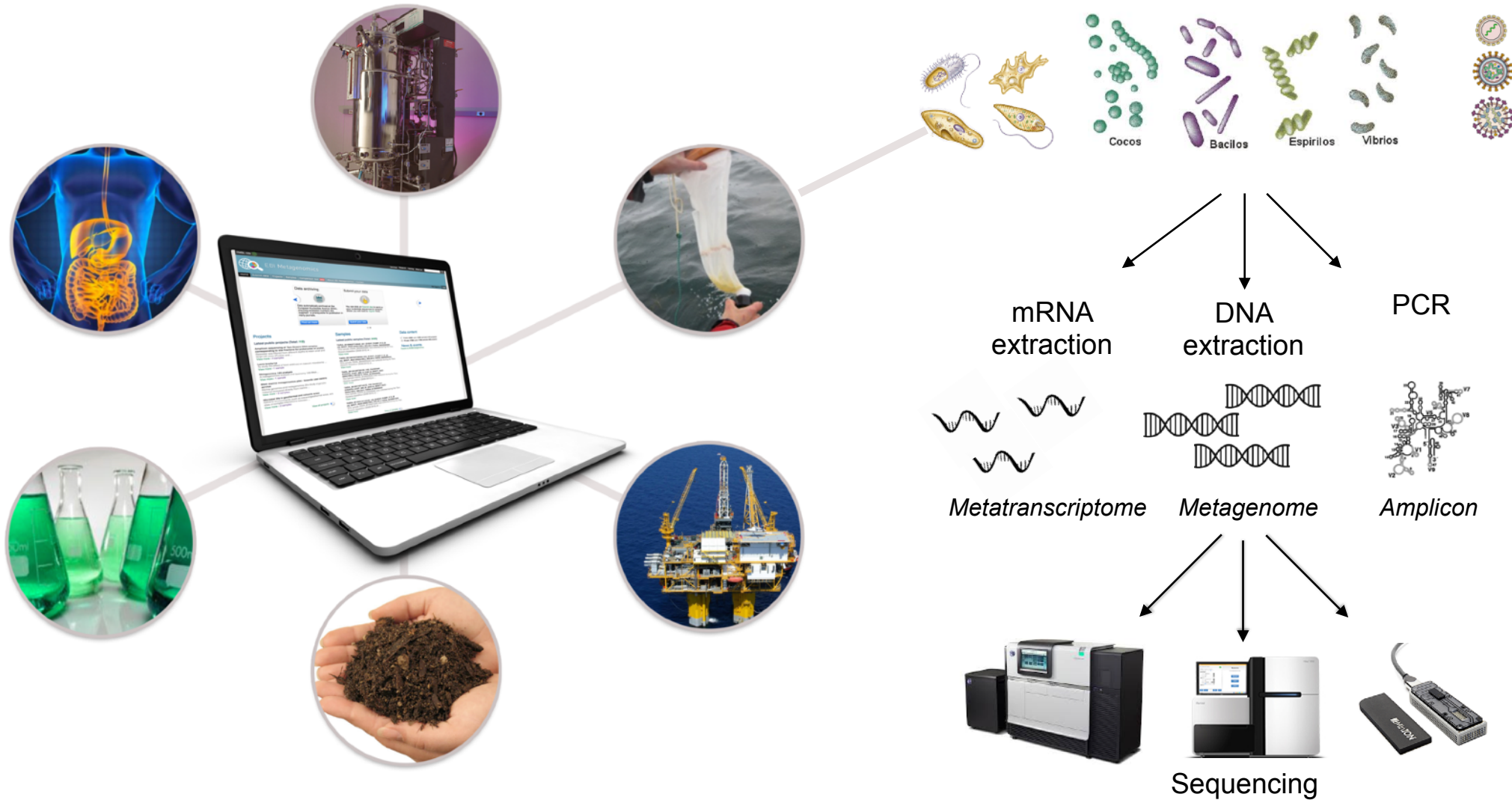


DNA sequencing and biodiversity



- PCR and whole genome sequencing in surveillance of SARS-CoV-2

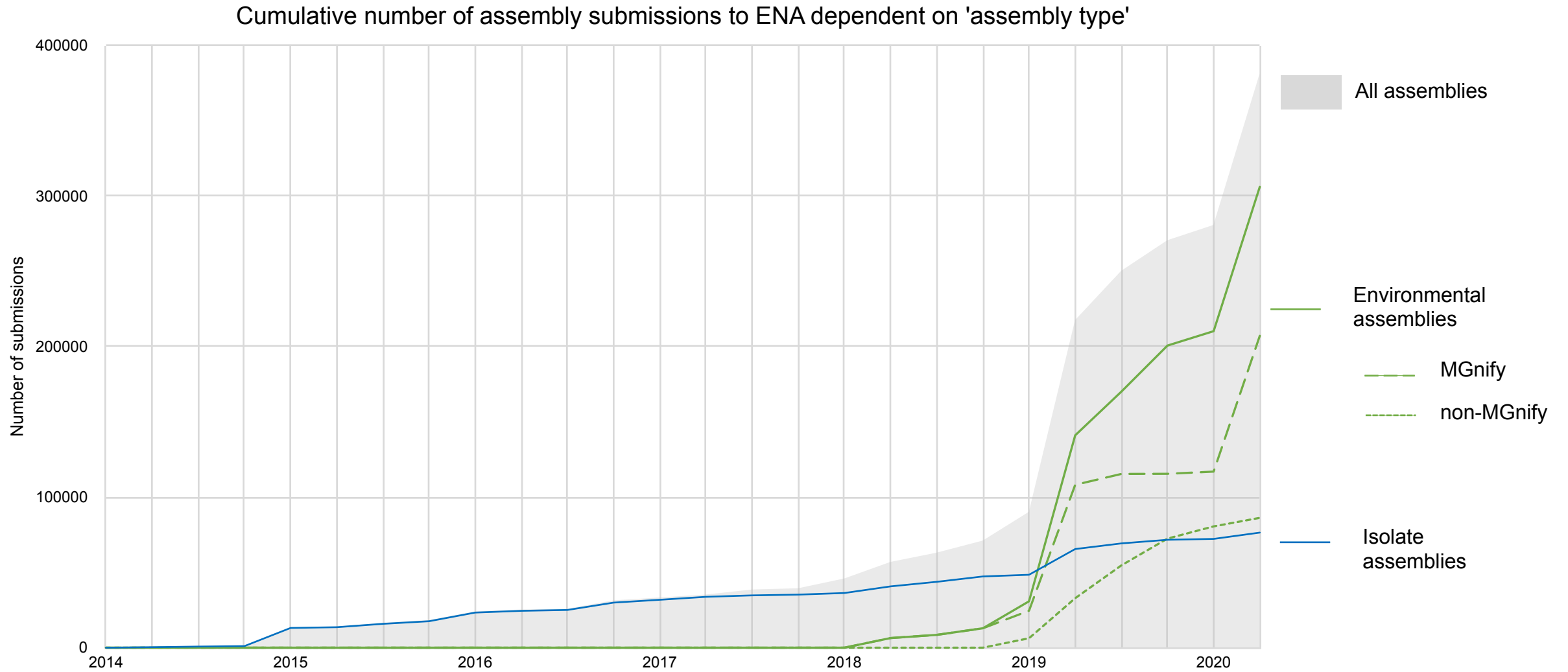
Similar approaches used to study microbiomes

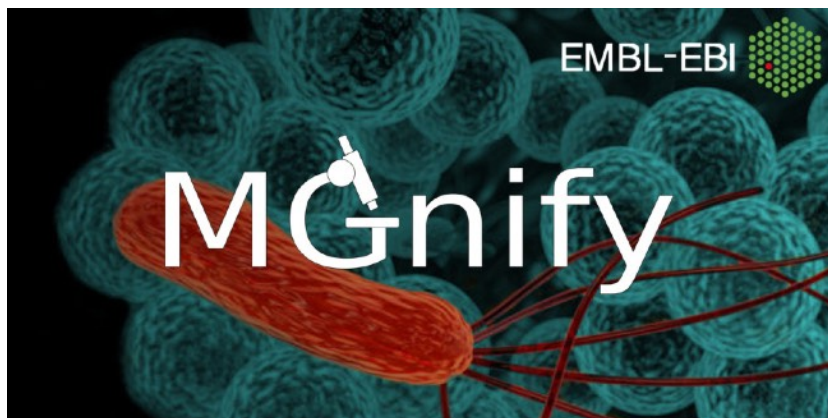


New sequencing technologies facilitating in field sequencing



Growth of environmental assemblies





A **free** to use resource for the archiving, assembly, analysis, & browsing of microbiome data

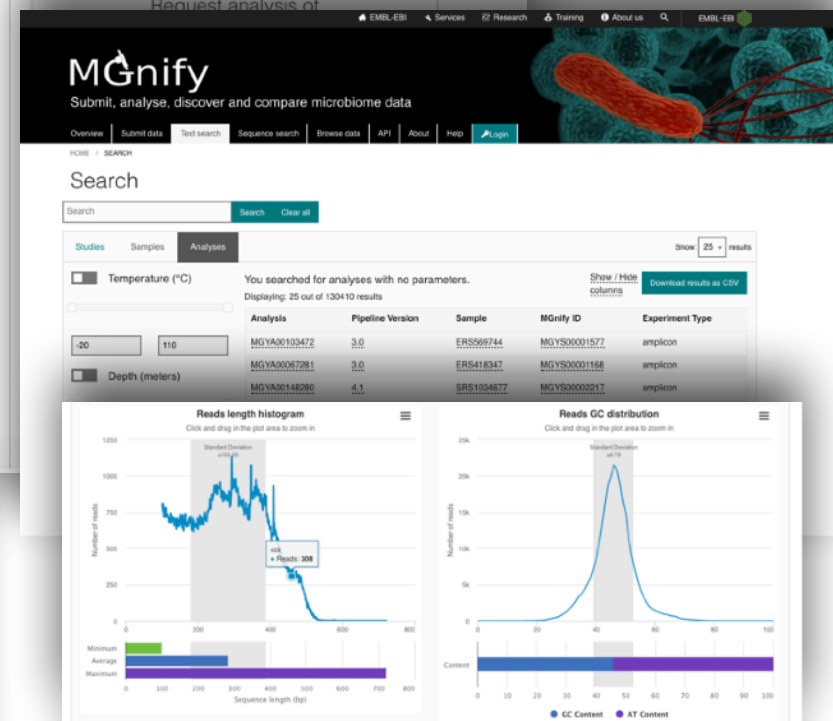
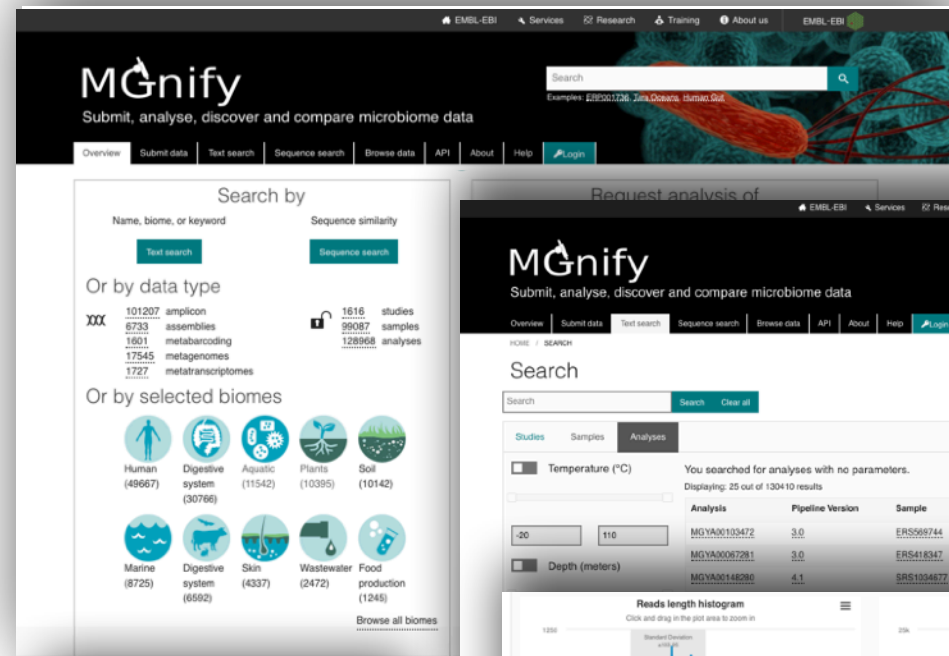
Data archiving



Assembly



Analysis



Tara Oceans in MGnify



EMBL-EBI

Services

Research

Training

About us

EMBL-EBI

MGnify

Submit, analyse, discover and compare microbiome data

Overview | Submit data | Text search | Sequence search | Browse data | API | About | Help | Login

Search

Examples: MGYS0000410, Tara Oceans, Human Gut

HOME / BROWSE STUDIES / STUDY MGYS00002392

Study MGYS00002392


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

Overview

Analysis summary

Last updated: 7-Jun-2018

Classification


 Environmental > Aquatic > Marine

Description

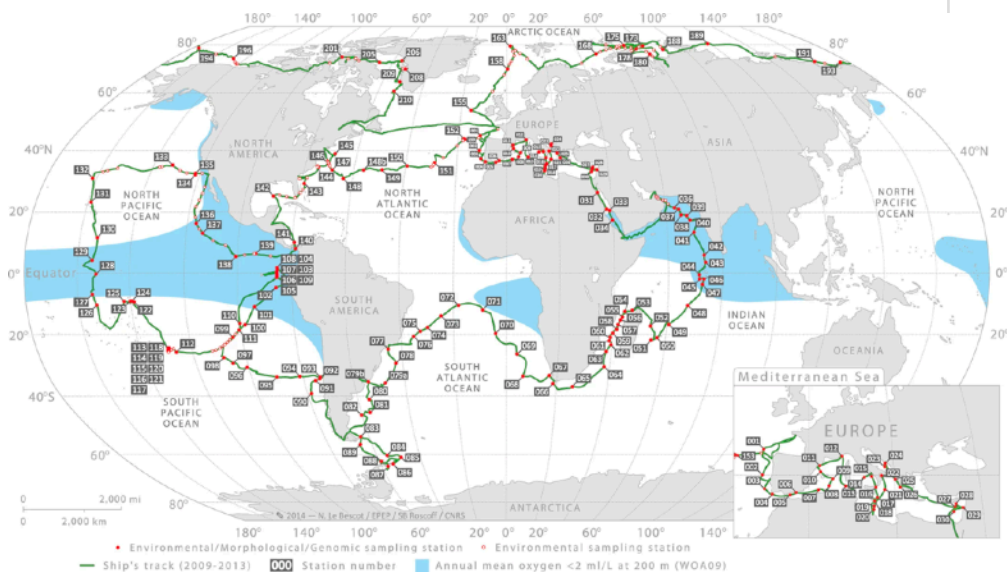
Seawater was filtered from different depths to retain small and large cell sizes (Protists Organisms). The DNA was extracted and amplified by PCR.

Map

Satellite

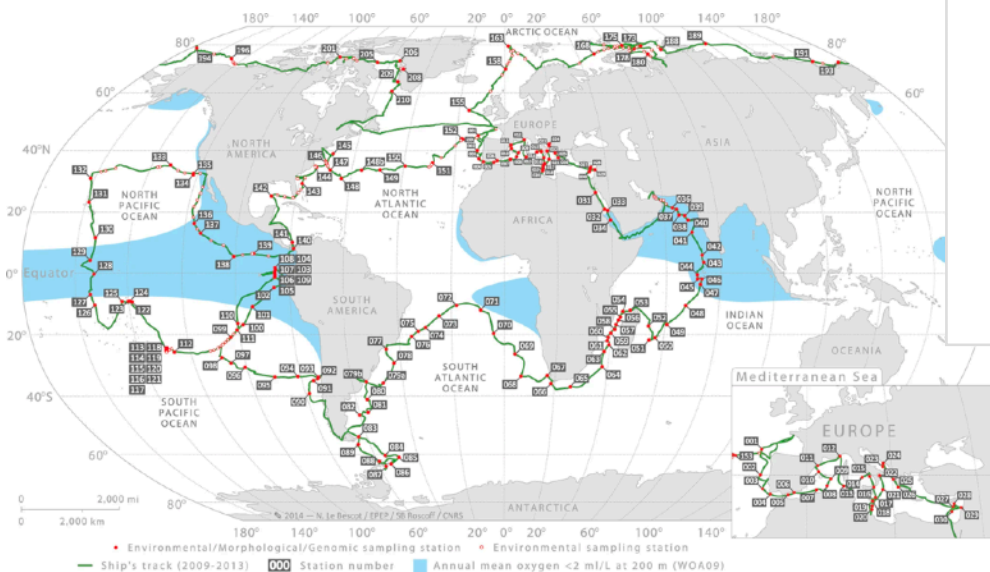


Keyboard shortcuts | Map data ©2021 | Terms of Use

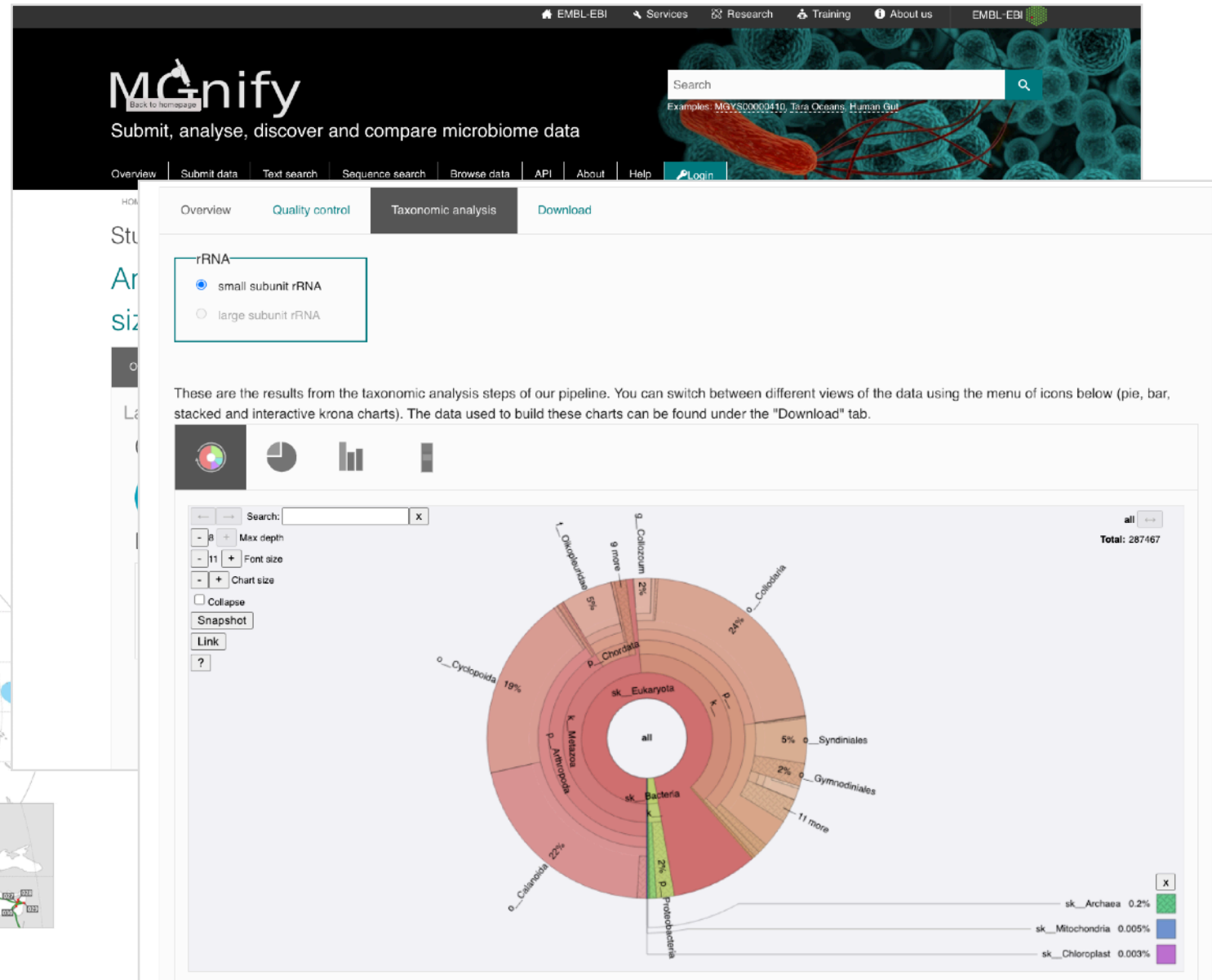


Scientific Data. 2. 10.1038/sdata.2015.23.

Tara Oceans in MGnify

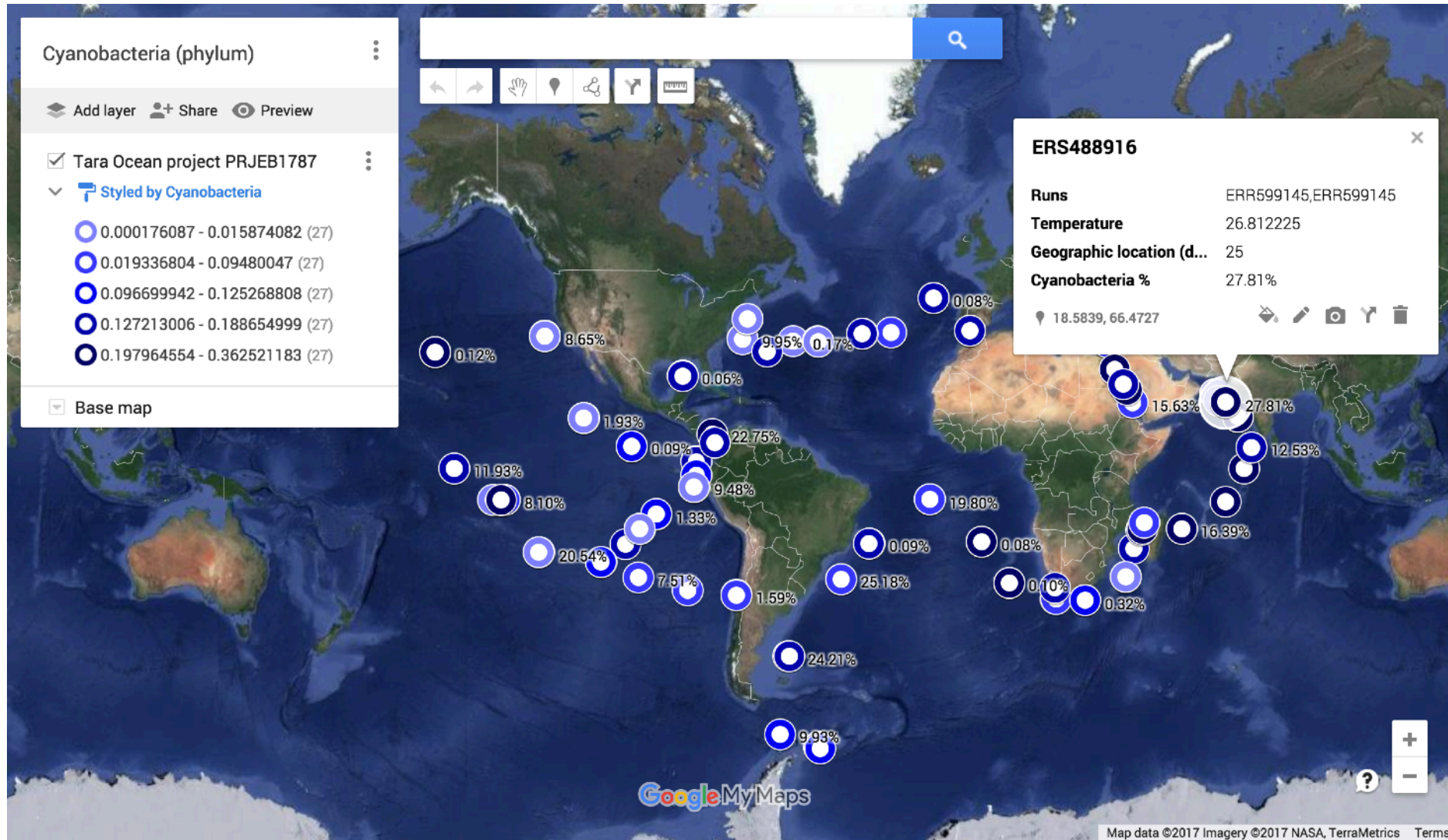


Scientific Data. 2. 10.1038/sdata.2015.23.

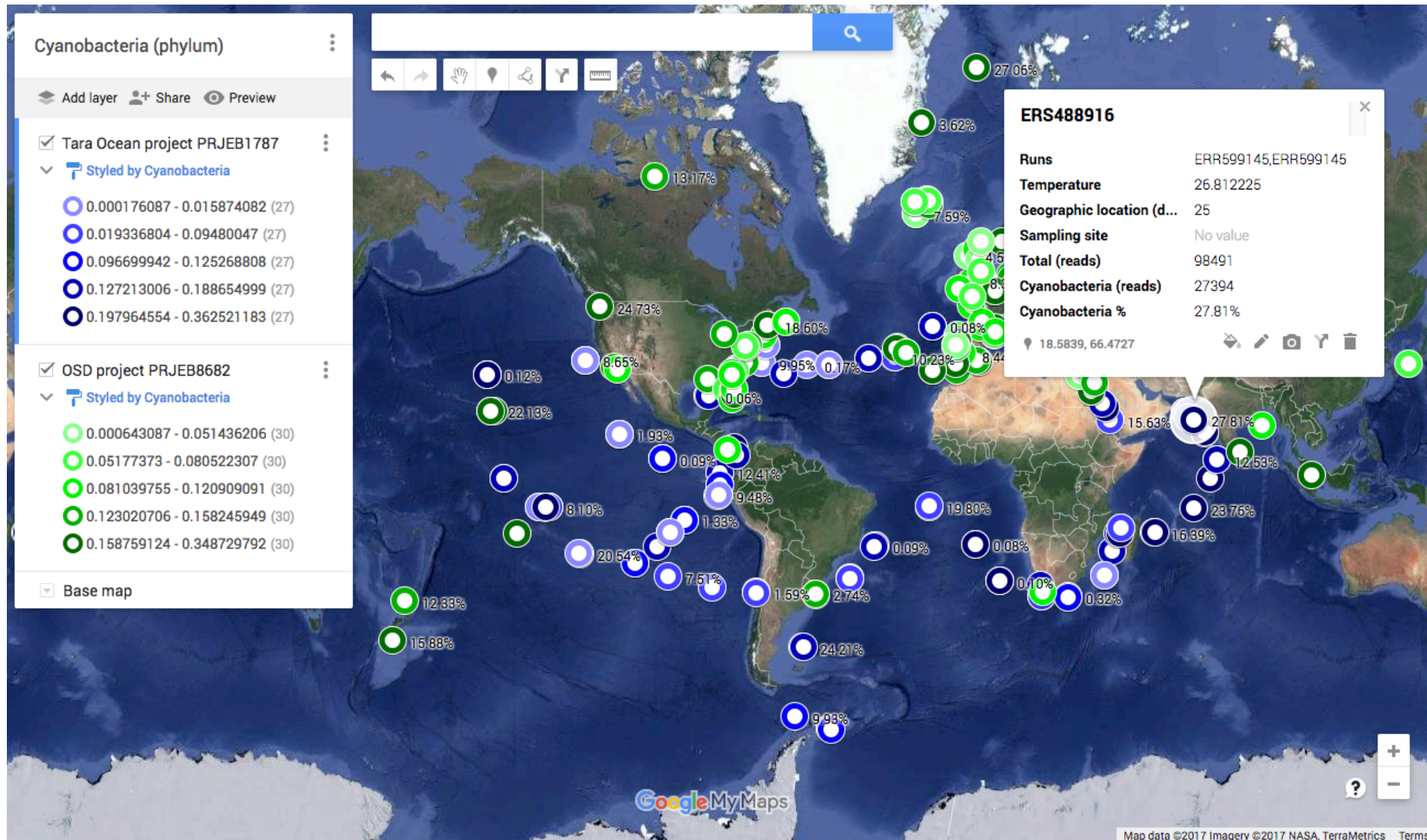


287K, 955 taxonomic assertions, >437 analyses

Geographic distribution of Cyanobacteria



Geographic distribution of Cyanobacteria




```

1  cwlVersion: v1.0
2  class: Workflow
3
4  inputs:
5    input_fasta_file: # input assembly
6      type: File
7    virsorter_virome:
8      type: boolean
9      default: false
10     doc: |
11       Set this parameter if the input fasta is mostly viral.
12       See: https://github.com/simroux/VirSorter/issues/50
13     ...
14

```

```

14
15  steps:
16    fasta_rename:
17      label: Filter contigs
18      run: ./Tools/FastaRename/fastarename.cwl
19      in:
20        input: input_fasta_file
21      out:
22        - renamed_fasta
23        - name_map
24
25    length_filter:
26      label: Filter contigs
27      run: ./Tools/LengthFiltering/length_filtering.cwl
28      doc: Default length 1kb
29      in:
30        fasta_file: fasta_rename/renamed_fasta
31        length:
32          default: 1.0
33      out:
34        - filtered_contigs_fasta
35

```

Tool

Tool

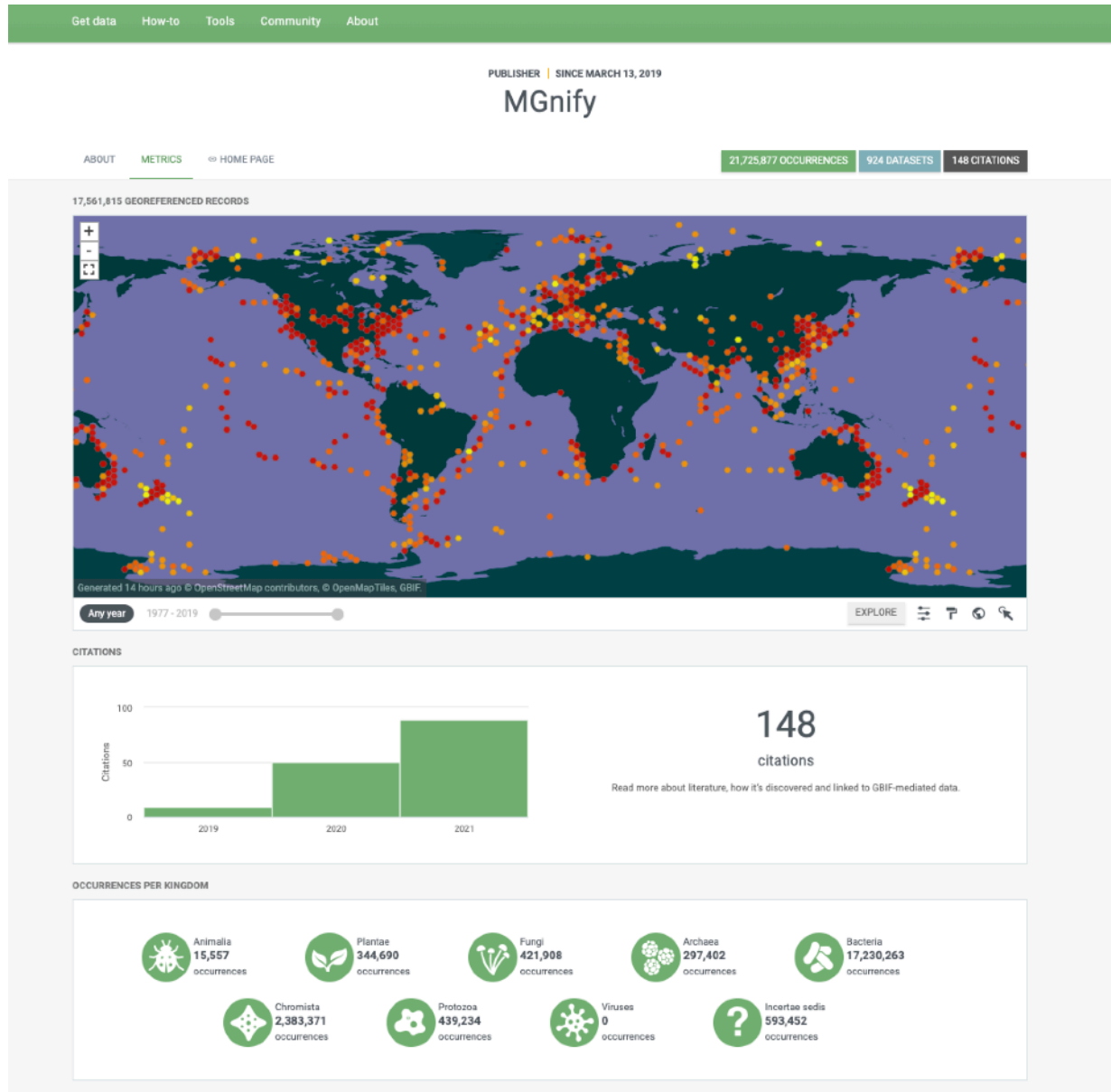
```

37  outputs:
38    filtered_contigs:
39      outputSource: length_filter/filtered_contigs_fasta
40      type: File
41    virfinder_output:
42      outputSource: virfinder/virfinder_output
43      type: File
44    ...

```

- MGnify adherence to the FAIR principles, GBIF values

Disseminating biodiversity information to additional audiences



- GBIF retrieves a subset of MGNify data
- Growing number of citations, both for MGNify and within GBIF
- Different audience
- Solutions allow the data to be consumed in a different way, e.g. taxonomy

Wider Context

- AtlantECO
 - Templating on MGnify-GBIF for OBIS
- ELIXIR
 - Biodiversity community
 - Standards and infrastructure
- Darwin Tree of Life
 - More Eukaryotes
 - Holobionts
- Climate Change - COP26
 - Organisms in context
 - Impact of climate change
 - Natural solutions



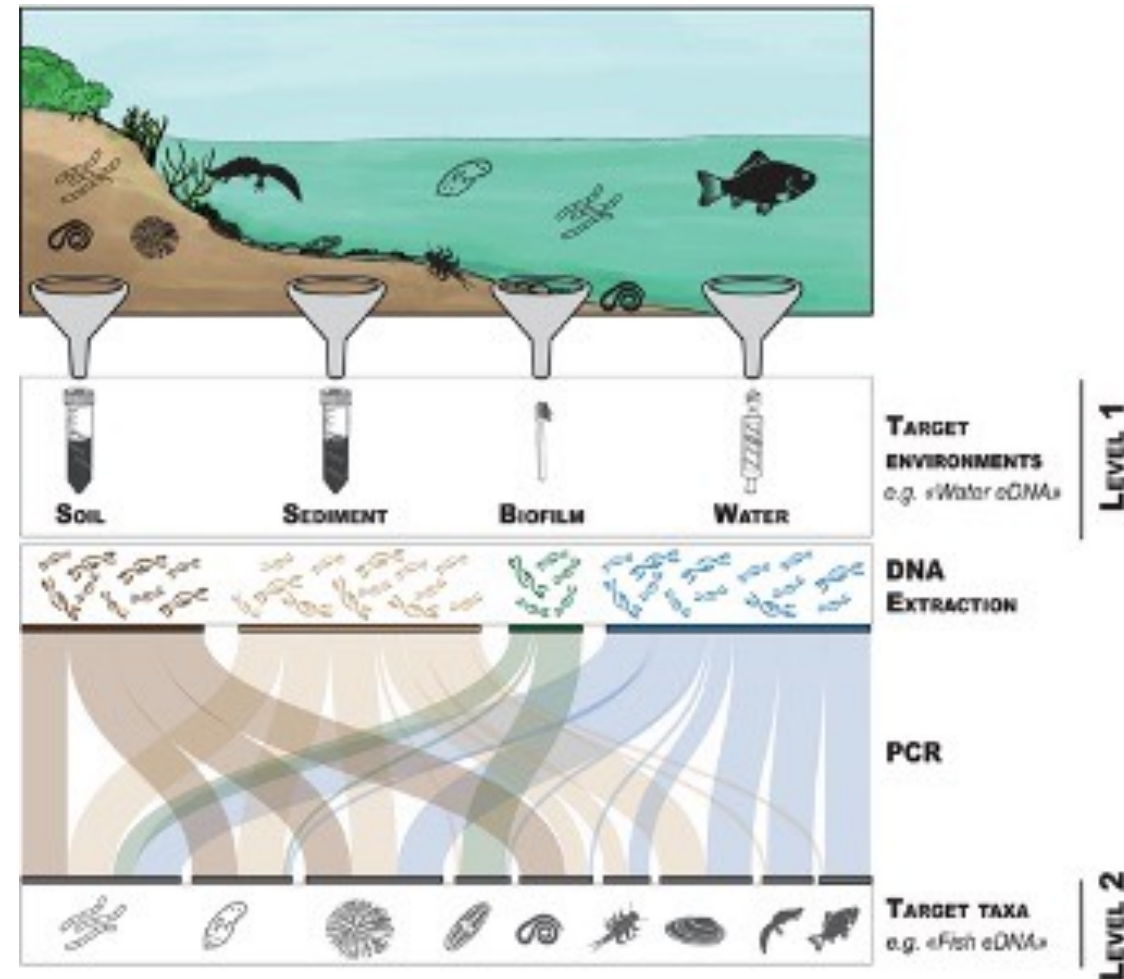
Rob Finn: Microbes can help curb climate change

Microorganisms act as tiny chemical reactors: they can emit greenhouse gases such as methane, but also feed on these gases and convert them into useful molecules. Researchers such as Rob Finn want to use microbes to regulate climate change. He is a microbiologist and team leader at the European Bioinformatics Institute in Hinxton, UK – part of the European Molecular Biology Laboratory.

<https://www.nature.com/articles/d41586-021-03029-w>

The Future?

- eDNA contains signature of every thing!
- Additional marker genes in MGnify
 - COX1, animals and protists
- Scaling to deal with volumes of data
- Linking genomic data to barcodes
- Modelling of whole communities
 - Metabolic
 - Ecological



Molecular Ecology, 2020 29(22): 258-4264, DOI: (10.1111/mec.15643)

Acknowledgements

Alex Almeida

Germana Baldi

Martin Beracochea

Ellen Cameron

Juan Caballero

Ales Escobar

Santiago Fragoso

Tanya Gurbich

Danilo Horta

Varsha Kale

Satwant Kaur

Sara Kashaf

Felix Langer

Khalimat Murtazalieva

Lorna Richardson

Sandy Rogers

Kate Sakharova

Paul Saary

Martin Hoelzer



Guillermo Rangel



Alejandro Reyes



Nikos Kyrpides

Stephen Nayfach



Trevor Lawley

Luis Camarillo-Guerrero

Mark Blaxter



Anton Korobeynikov



Julie Segre



Nicola Segata



Towards amplicon sequence variants (ASVs)

