



Micro B3 Information System & BioVeL

Resources, Services, Workflows and Interfaces



**Microbial Genomics and
Bioinformatics Research Group**

Renzo Kottmann

rkottman@mpi-bremen.de

Bremen, 2014-06-03



Resources

Databases & Technology

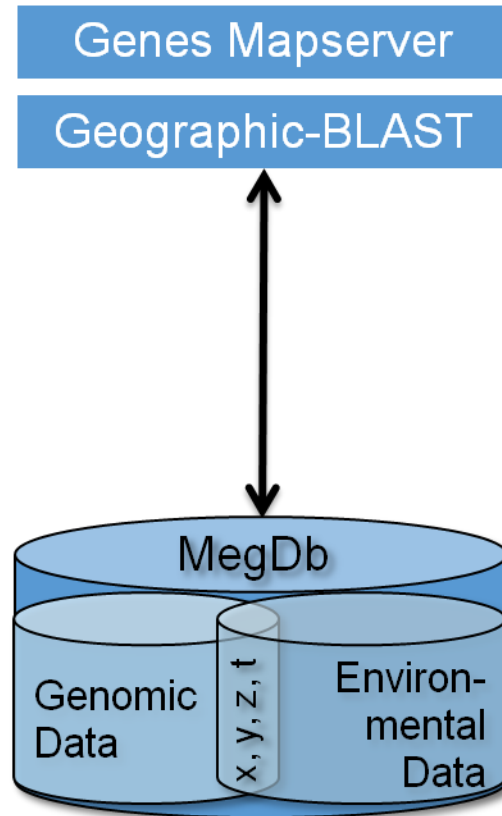


silvangs

www.arb-silva.de/ngs



Megx.net and Micro B3 Information System



Micro B3 Information System

Megx.net



PostBIS: In Database Sequence Data Compression

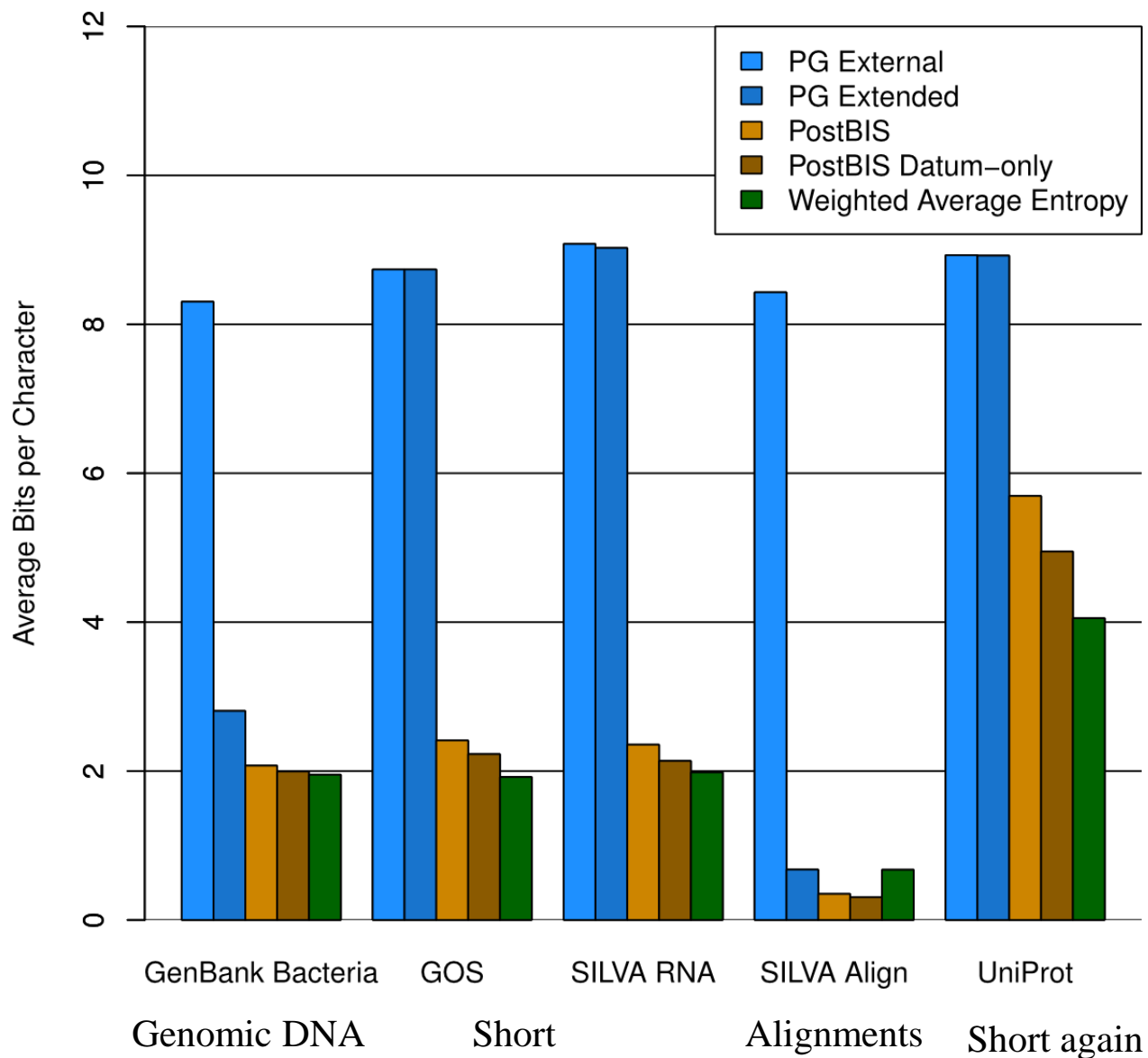


PostBIS (Hamburg University)

- Efficient storage and retrieval of DNA sequence data
- <2 bits per nucleotide base
- 500x faster substring operation
- ▶ **PostgreSQL extension**
 - In-database sequence compression
 - Special Data Types
 - Special Functions

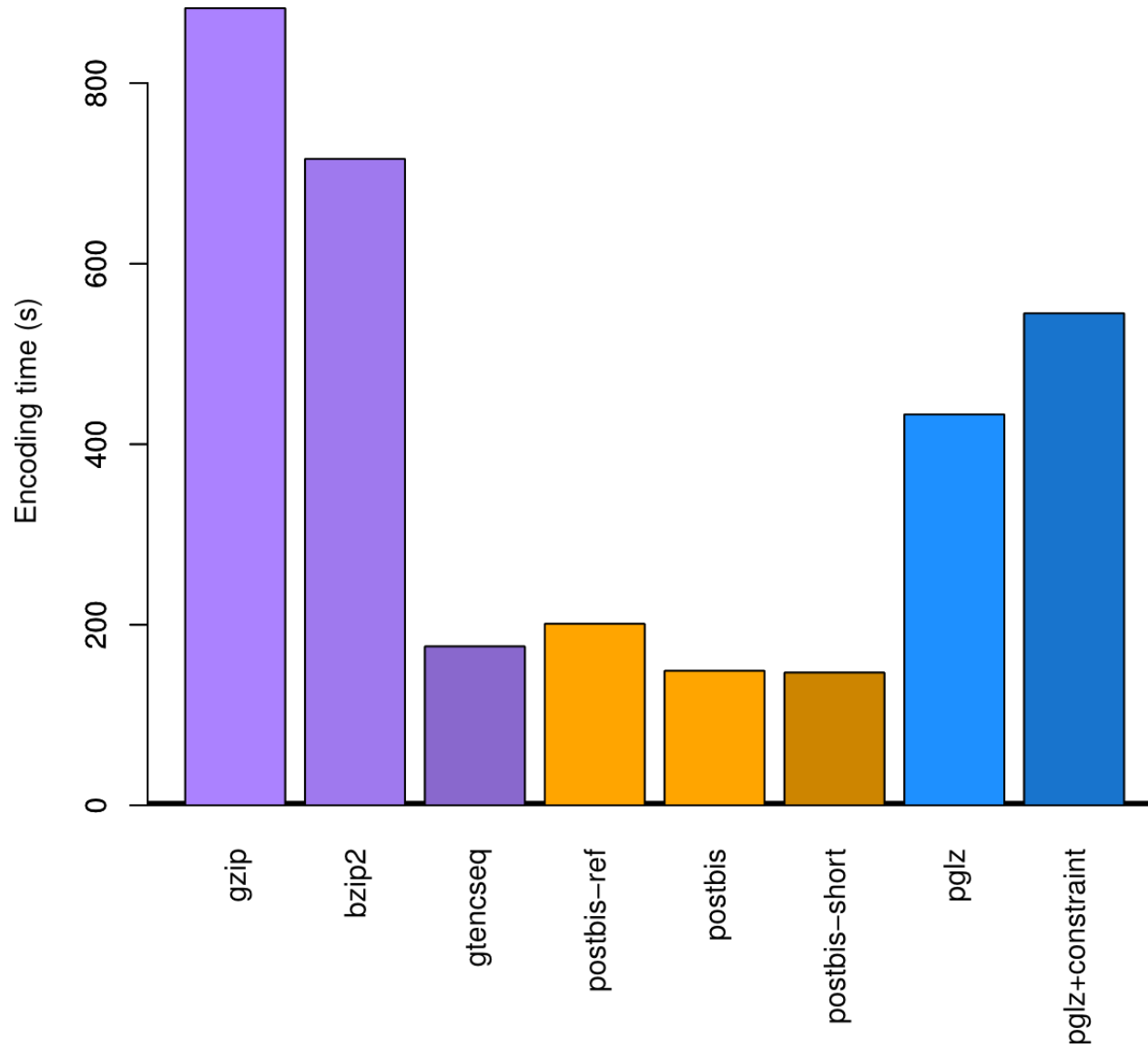


PostBIS: Storage Size





PostBIS: Encoding Performance





Services

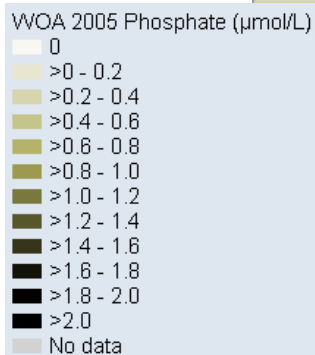
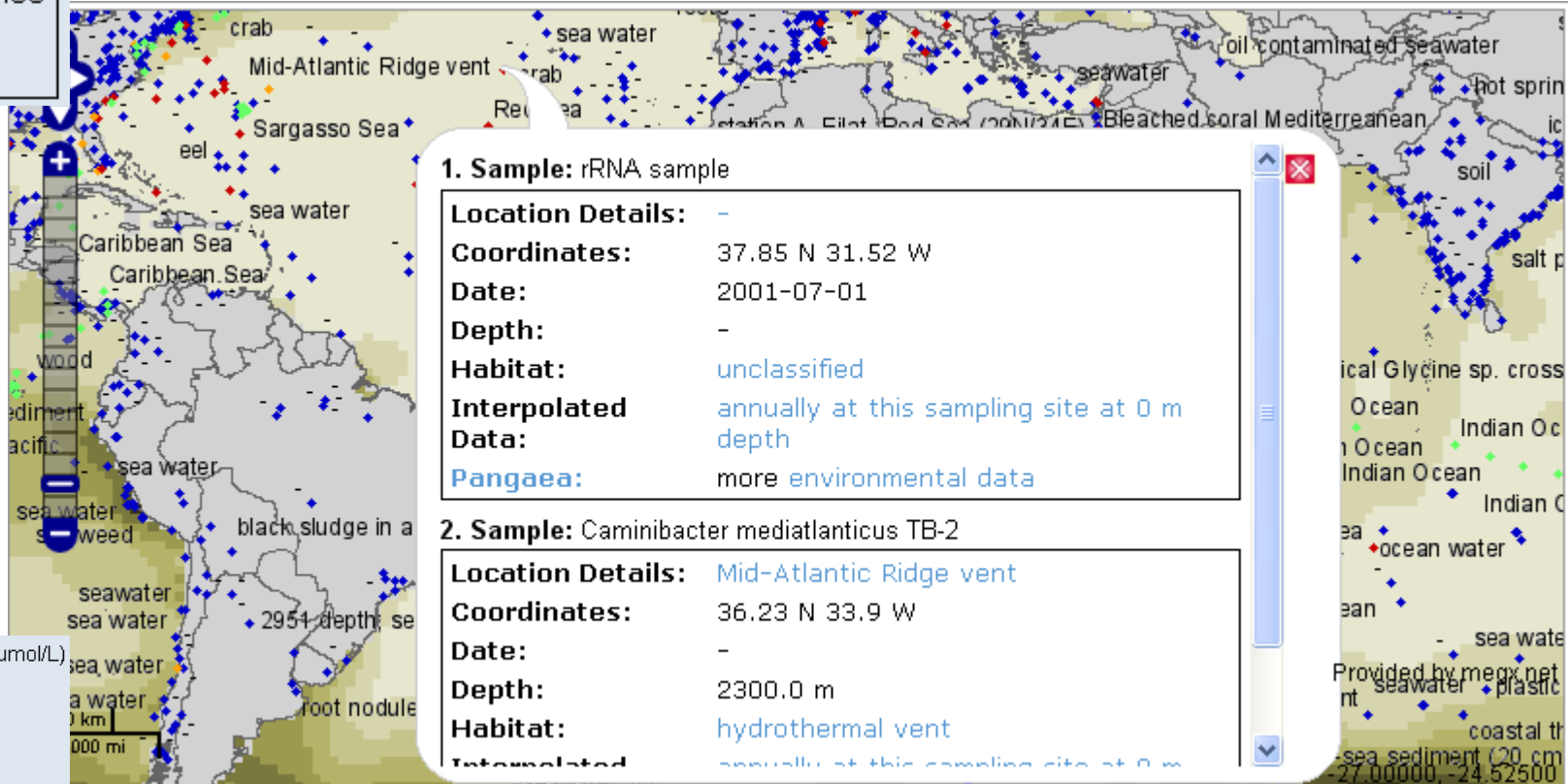
RESTful, WMS/WFS and Co.



Who is out there and where?

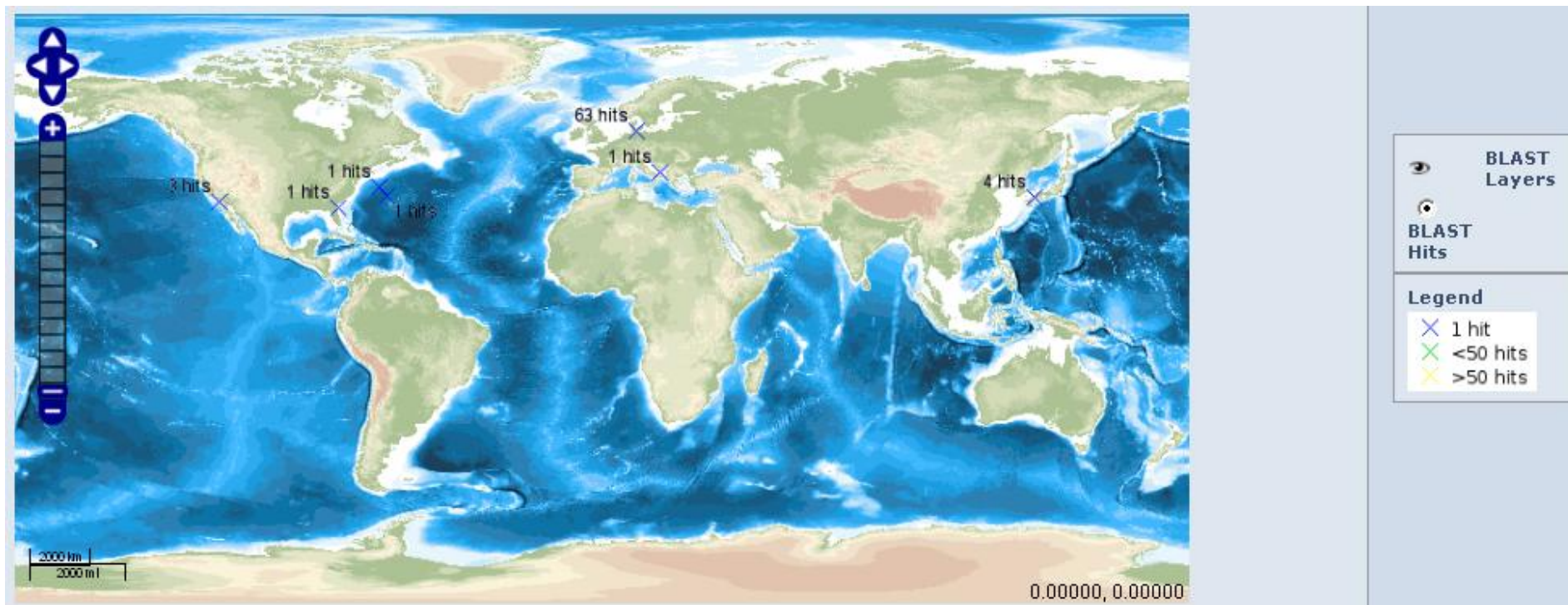
(in terms of sequenced genomes, metagenomes and key genes)

- Sampling Sites
- ◆ Genomes
 - ◆ Metagenomes
 - ◆ rRNAs
 - ◆ Phages





Geographic-BLAST



Target description	Target gi	E-value	Sampling site	Habitat
arylsulfatase [Rhodopirellula baltica SH 1]	32471036	0.00e-1	54.5 N 10.3 E	marine habitat
arylsulfatase [Rhodopirellula baltica SH 1]	32471916	8.99e-85	54.5 N 10.3 E	marine habitat
sulfatase [Sinorhizobium medicae WSM419]	150397268	4.04e-77	-	soil
arylsulfatase [Rhodopirellula baltica SH 1]	32475025	9.96e-76	54.5 N 10.3 E	marine habitat
arylsulfatase (aryl-sulfate sulphohydrolase) [Bacteroides thetaiotaomicron VPI-5482]	29348502	6.04e-73	-	organism-associated habitat



OSD App Server Map



Recent observations



Observer: Osd Summer School
OSD

Sample Label:
Crete_Summer_School_surface_Group-1

Ocean: Aegean Sea

Date: May 27, 2014 5:42:00 PM

[View more](#)



Observer: Julia Schnetzer
Sample Label: crete summer school deep sample

Ocean: Aegean Sea

Date: May 27, 2014 8:41:00 AM

[View more](#)

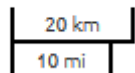
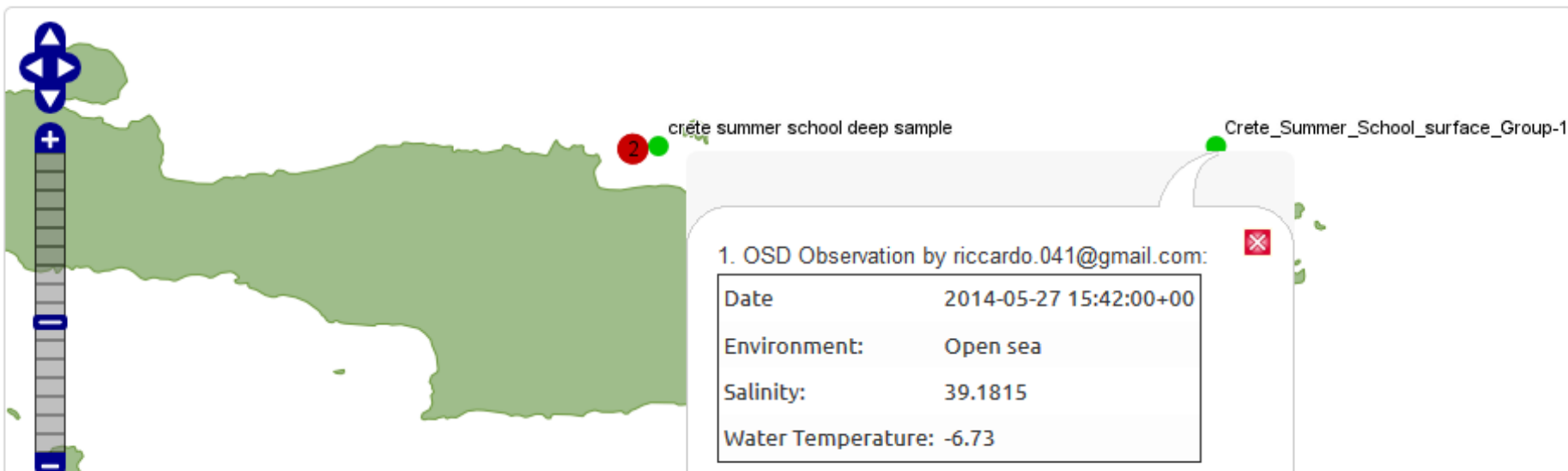


Observer: Julia Schnetzer
Sample Label: Crete Summer School surface

Ocean: Aegean Sea

Date: May 27, 2014 2:16:00 PM

[View more](#)



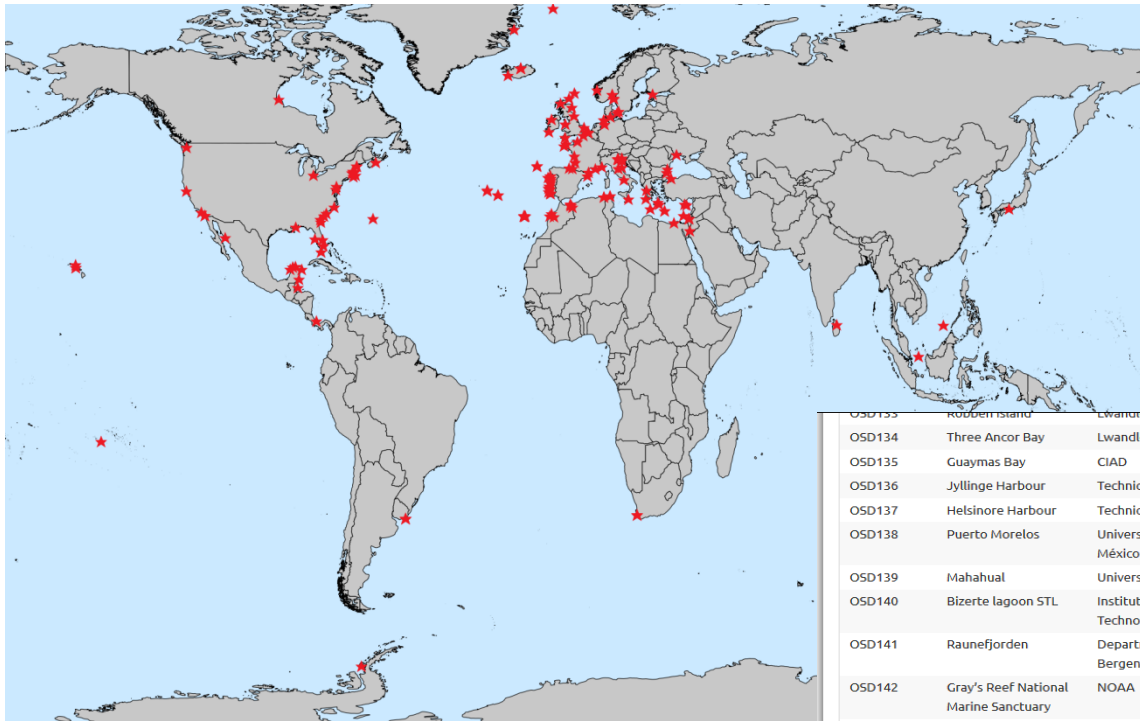
<http://mb3is.megx.net/osd-app/samples>

[Permalink](#)

28 78944 34 94717



OSD Registry



OSD133	Roodepan Island	Lwandle Technologies	Kate Munnik	South Africa	View
OSD134	Three Ancor Bay	Lwandle Technologies	Kate Munnik	South Africa	View
OSD135	Guaymas Bay	CIAD	Jaqueline García	Mexico	View
OSD136	Jyllinge Harbour	Technical University of Denmark	Eva Sonnenschein	Denmark	View
OSD137	Helsingore Harbour	Technical University of Denmark	Eva Sonnenschein	Denmark	View
OSD138	Puerto Morelos	Universidad Nacional Autónoma de México, Unidad Sisal	Alejandra Prieto Davó	Mexico	View
OSD139	Mahahual	Universidad Autónoma de Yucatán	Rafael Rojas Herrera	Mexico	View
OSD140	Bizerte lagoon STL	Institut National Des Sciences et Technologies de la Mer (INSTM)	Monia Elbour	Tunisia	View
OSD141	Raunefjorden	Department of Biology, University of Bergen	Lise Øvreås	Norway	View
OSD142	Gray's Reef National Marine Sanctuary	NOAA	Sarah Fangman & Marc Frischer	USA	View
OSD143	Skidaway Institute of Oceanography	Skidaway Institute of Oceanography, University of Georgia	Marc Frischer	USA	View
OSD144	Maunaloa Bay O'ahu	Hawaiian Island Humpback Whale National Marine Sanctuary	Jonathan Martinez	USA	View
OSD145	North Sea - Blankenberge	Datafable	Bart Aelterman	Belgium	View
OSD146	Fram Strait	Alfred Wegener Institute Helmholtz Center for Polar and Marine Research	Katja Metfies, Christina Bienhold	Germany	View
OSD147	Rajarata	Faculty of Applied Sciences, Rajarata, University of Sri Lanka, Mihintale	Ranjith Eedirisinghe	Sri Lanka	View
OSD148	Wadden Sea	Institut fuer Chemie und Biologie des Meeres (ICBM) at Carl von Ossietzky Universitaet Oldenburg	Julia Busch	Germany	View
OSD149	Laguna Rocha Norte	Centro Universitario de la Región Este, Rocha	Dr. Cecilia Alonso	Uruguay	View
OSD150	Laguna Rocha Sur	Centro Universitario de la Región Este, Rocha	Dr. Cecilia Alonso	Uruguay	View



Ecological Analysis Tools for Microbial Ecology (EATME)



Perform a principal components analysis...

Data upload Transformations PCA parameters

Download results...

Description

This App will perform a PCA using the rda() function from the vegan package for R. Transformations are performed by decostand(), also from vegan

CSV parameters

Note that these parameters apply to all files uploaded. If your files are not correctly formatted, errors will result.

Header

Which column contains row labels (enter "0" if there is no such column)?

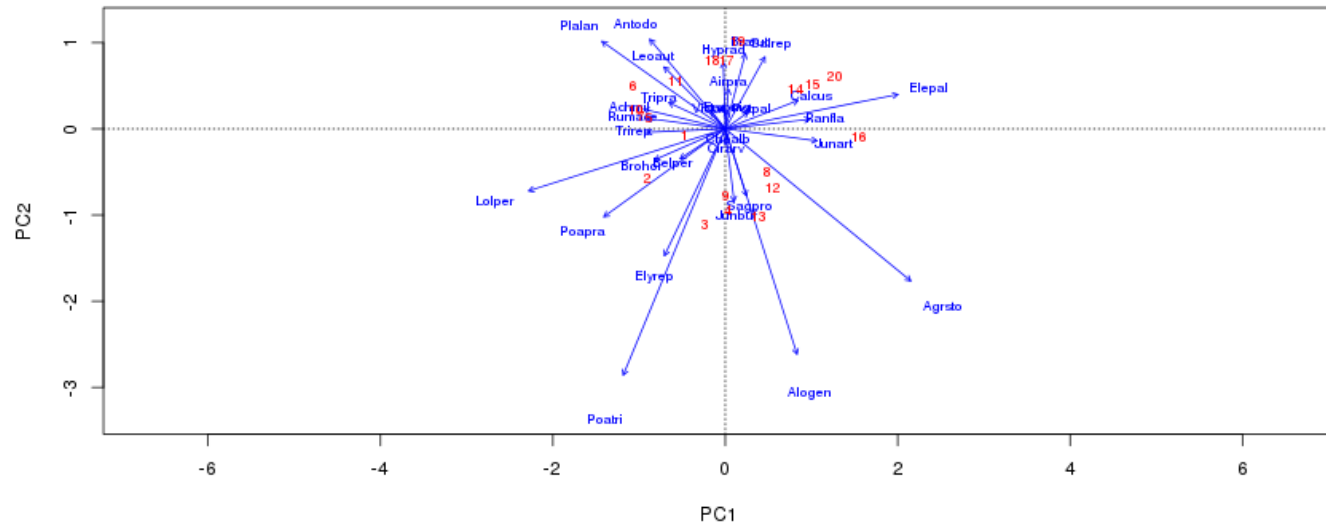
1

Separator

Comma

Semicolon

Plot Summary Eigenvalues Object scores Variable scores



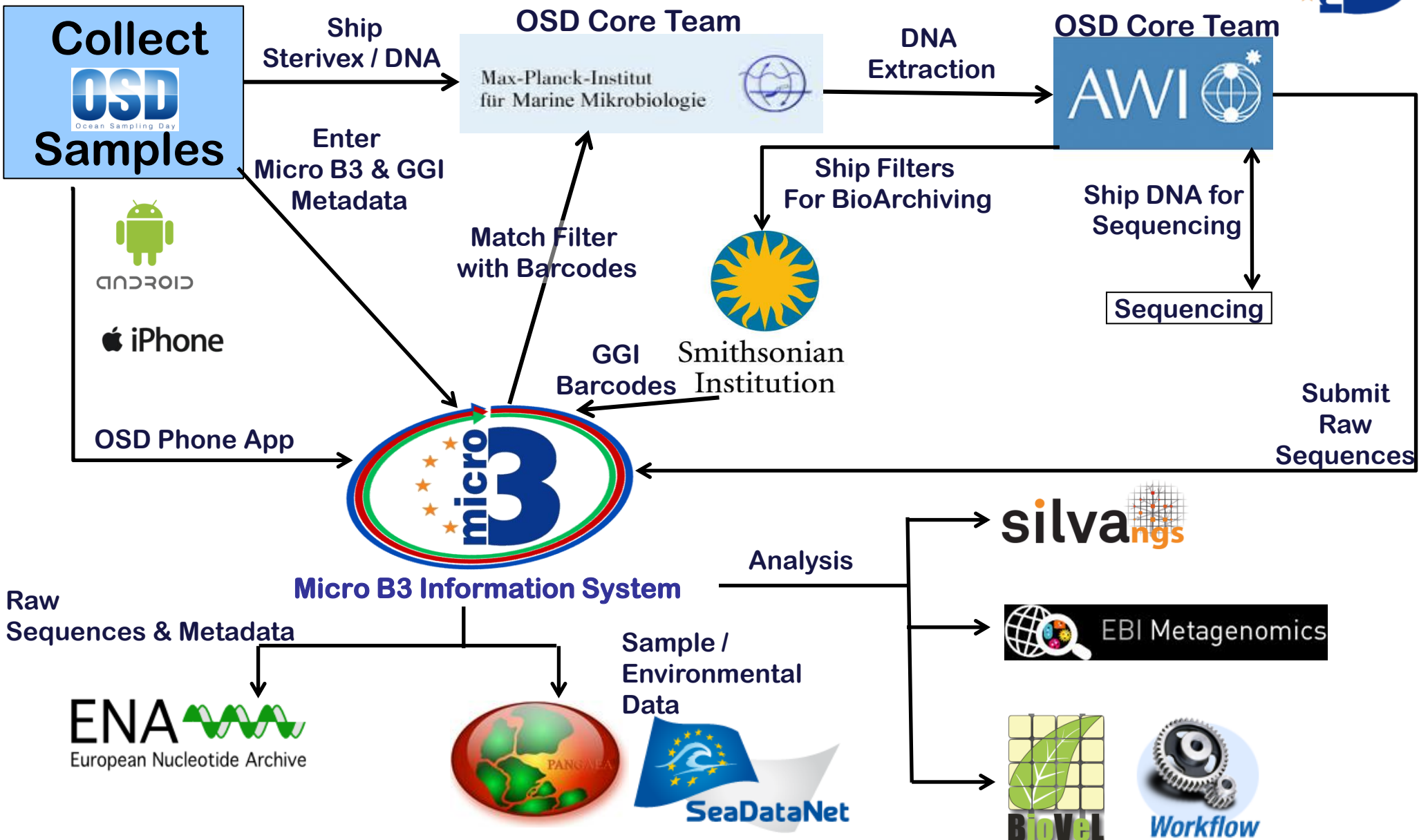
<http://mb3is.megx.net/eatme>



Workflows

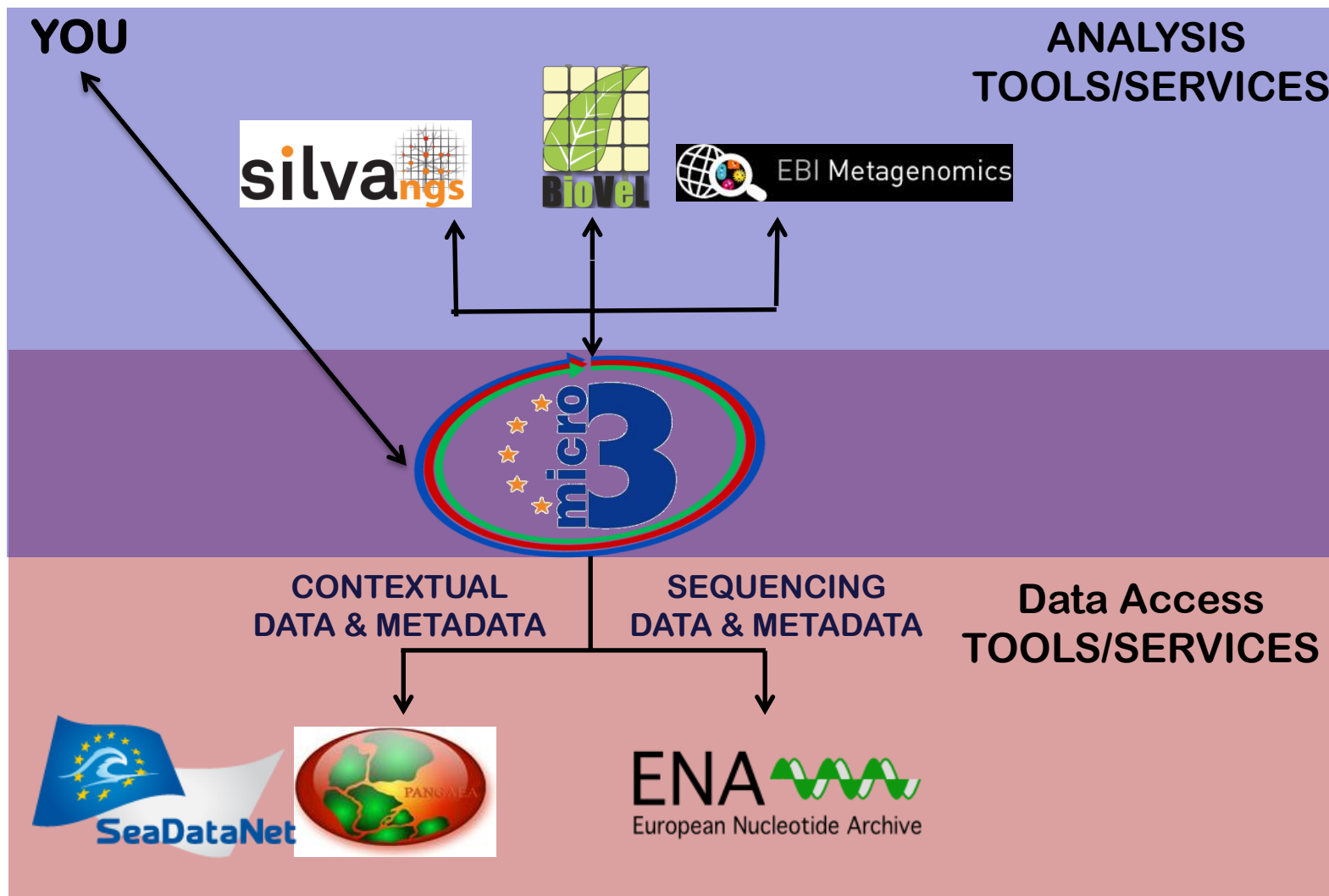


Ocean Sampling Day Workflows



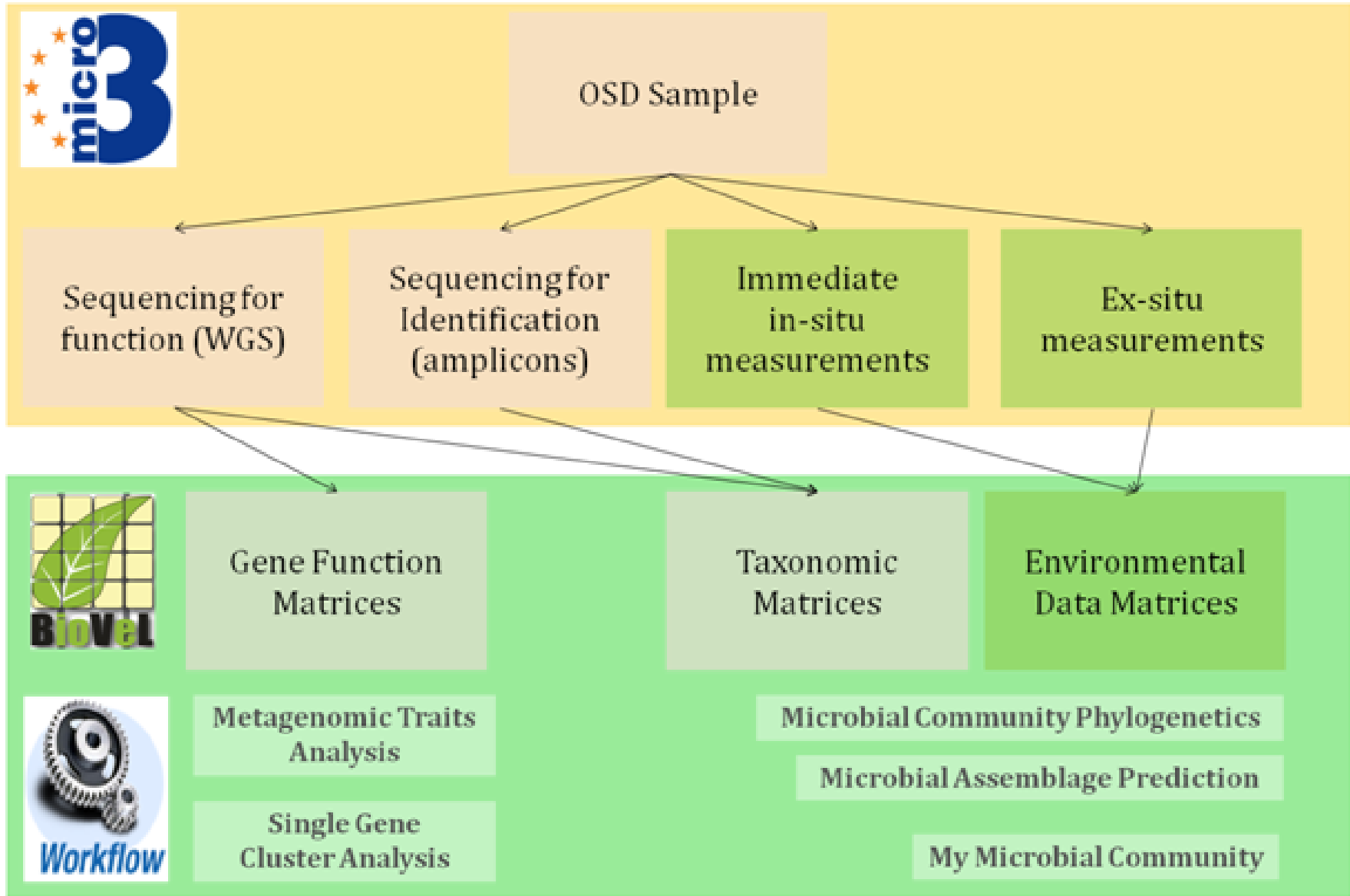


Accessing and Analyzing





Collaboration with BioVeL

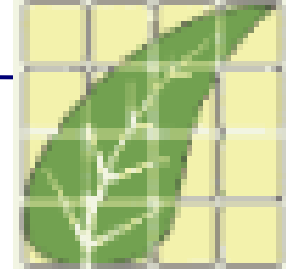




Biodiversity Virtual e-Laboratory

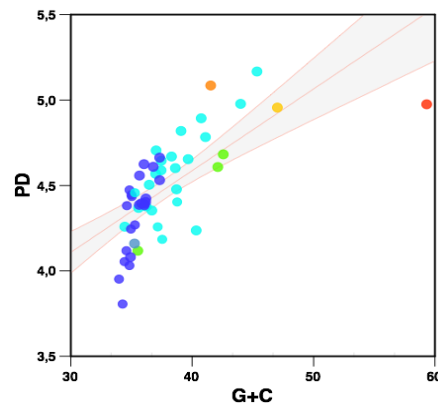
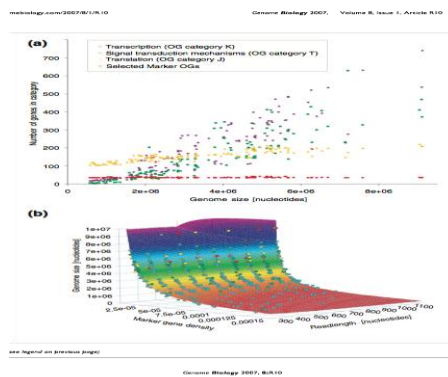
Project Overview

BioVeL is an e-laboratory that supports research on biodiversity to process large amounts of data from cross-disciplinary sources



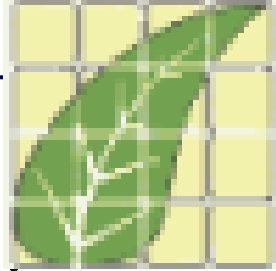
Metagenomic Workflows

FUNCTIONAL TRAI-BASED ANALYSIS OF AQUATIC MICROBIAL COMMUNITIES





Examples of Metagenomic Traits

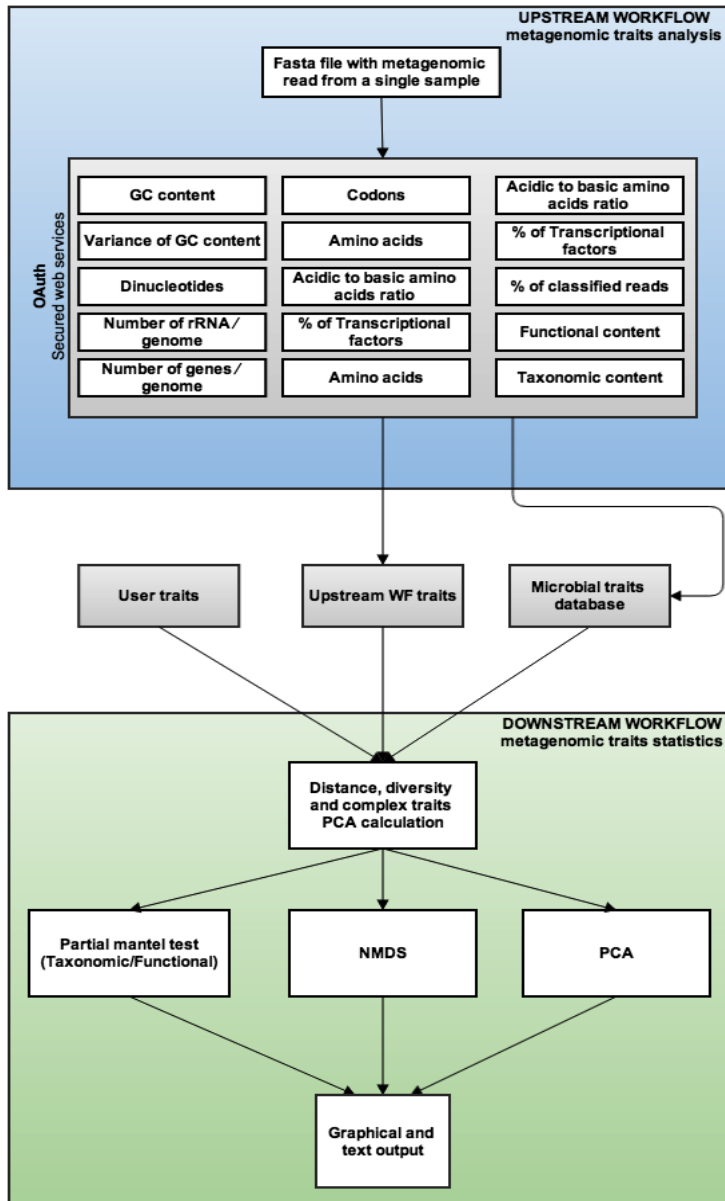
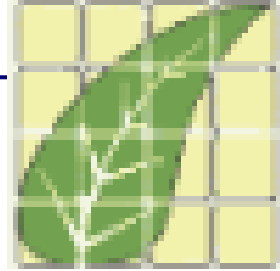


- ▶ **GC (Guanine-Cytosine) content (mean and variance).**
 - Related to genome size, environmental complexity and community composition.
- ▶ **Functional and phylogenetic diversity:**
 - Related to metabolic potential, community composition and environmental biogeochemistry.
- ▶ **Dinucleotide frequency:**
 - Related to phylogenetic composition.

Explore community traits as ecological markers in microbial metagenomes.
(Barberan, Fernandez et al. 2012).



The Metagenomic Trait Workflow(s)



▶ Upstream:

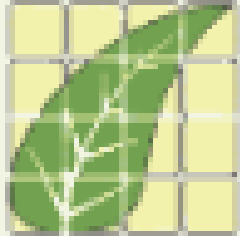
- Calculating traits
 - ◆ (traits-analysis workflow)

▶ Downstream

- Calculating statistics
 - ◆ (traits-statistics workflow)
 - ◆ R scripts perform multivariate statistic analyses using the vegan package and plot the results using ggplot2



Metagenomic Trait Database



► <http://mb3is.megx.net/mg-traits/sampleDetails?id=656>

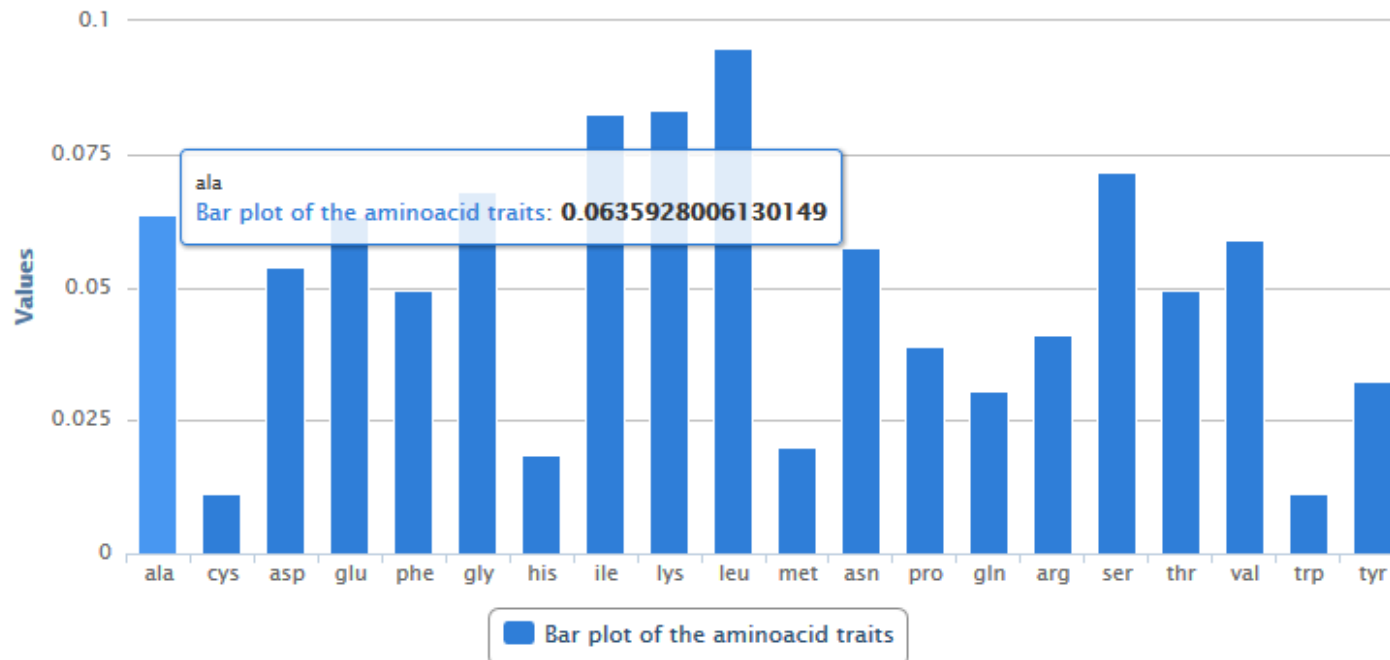
[Download](#)

Sample name: GS000d	Sample description: GS000d - Sargasso Station 13	
Latitude: 31.175	Longitude: -64.32433	Environment: Open Ocean
Environmental ontology: marine	GC content: 36.5384768841801	GC variance: 105.794430441467
Number of reads: 332240	Total nucleotides: 335939509	Number of genes: 500105
AB ratio: 0.817568	%TF: 2.37228237518261	%Classified: 54.651723138141

Aminoacid Trait



[Download](#)





Ecological Analysis Tools for Microbial Ecology (EATME)



Perform a principal components analysis...

Data upload Transformations PCA parameters

Download results...

Description

This App will perform a PCA using the rda() function from the vegan package for R. Transformations are performed by decostand(), also from vegan

CSV parameters

Note that these parameters apply to all files uploaded. If your files are not correctly formatted, errors will result.

Header

Which column contains row labels (enter "0" if there is no such column)?

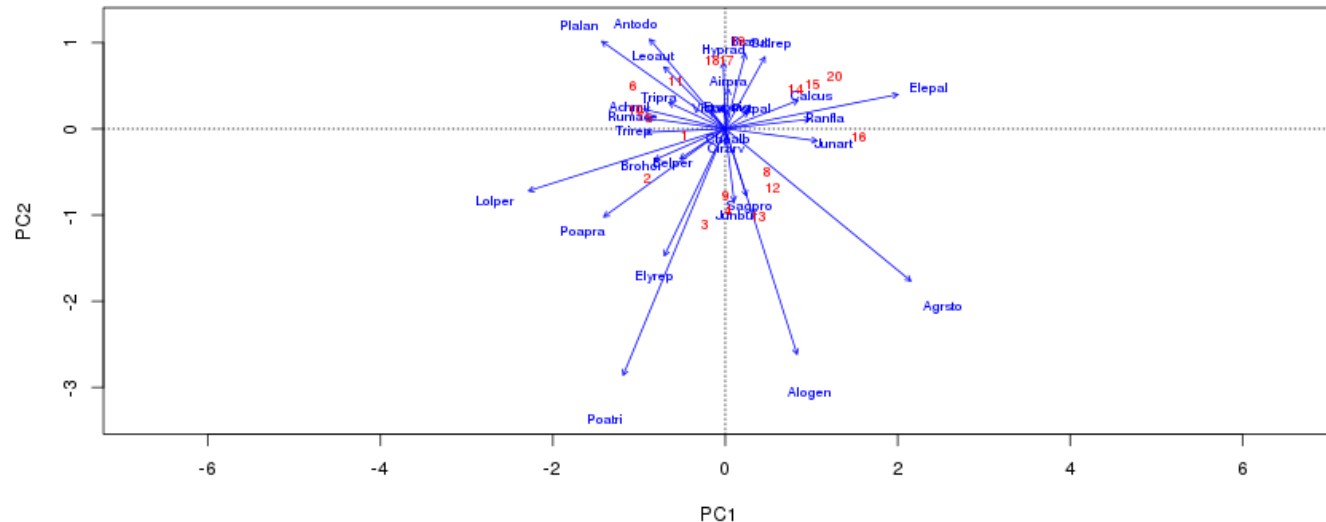
1

Separator

Comma

Semicolon

Plot Summary Eigenvalues Object scores Variable scores



<http://mb3is.megx.net/eatme>



Interfaces



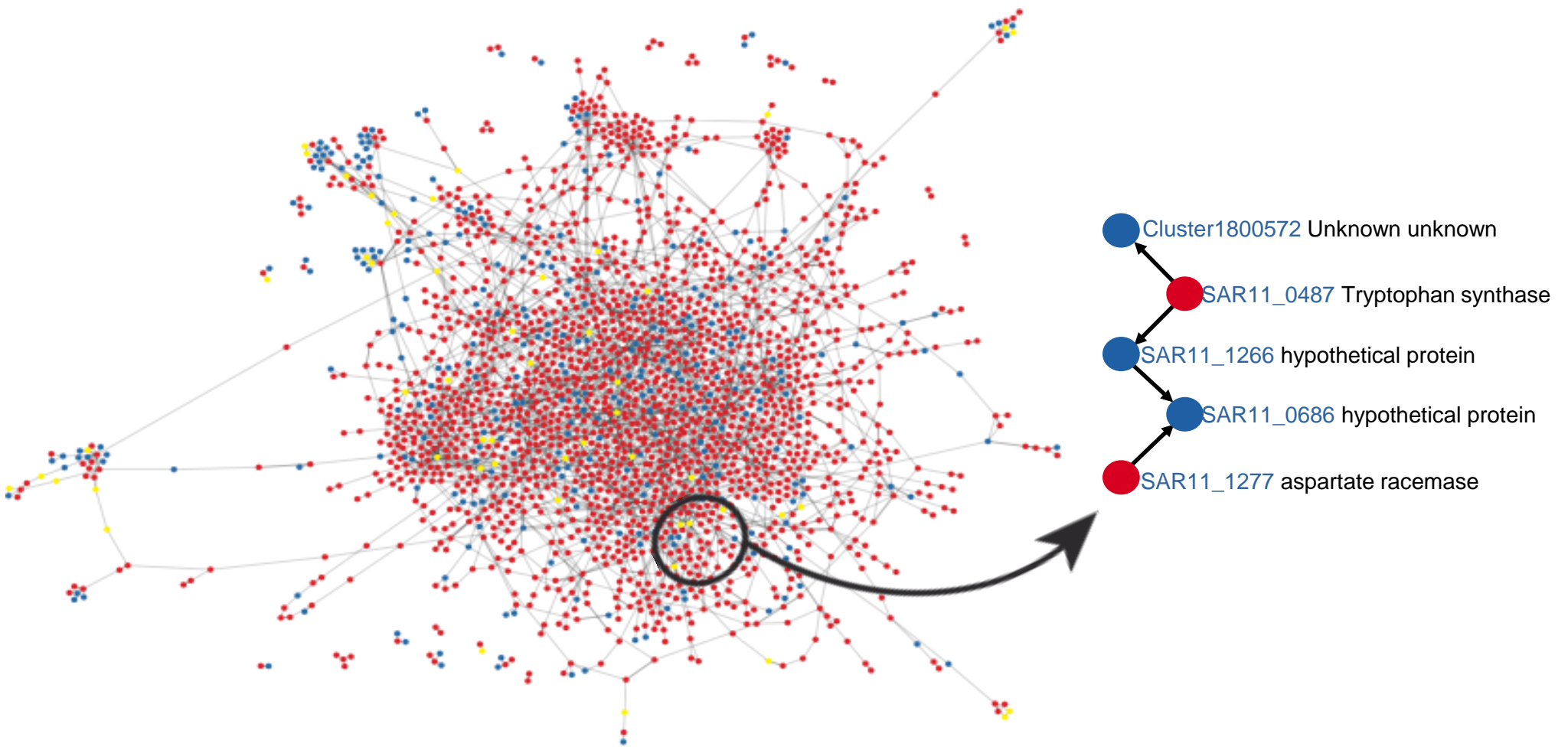
▶ **Web Services are interfaces**

- **Secured**
- **Versioned**

- **Interoperable**



Discovery: knowns, known *unknowns* and *unknown unknowns*



Pelagibacter ubiquus proteome centered subnetwork
Antonio Fernandez, submitted



Data Access: Visualization of *unknown* networks



ProX 0.4

Search term:

▸ Data

▾ Pfam

PF00610.16 Domain found in Dishevelled, Egl-10, and Pleckstrin (DEP)

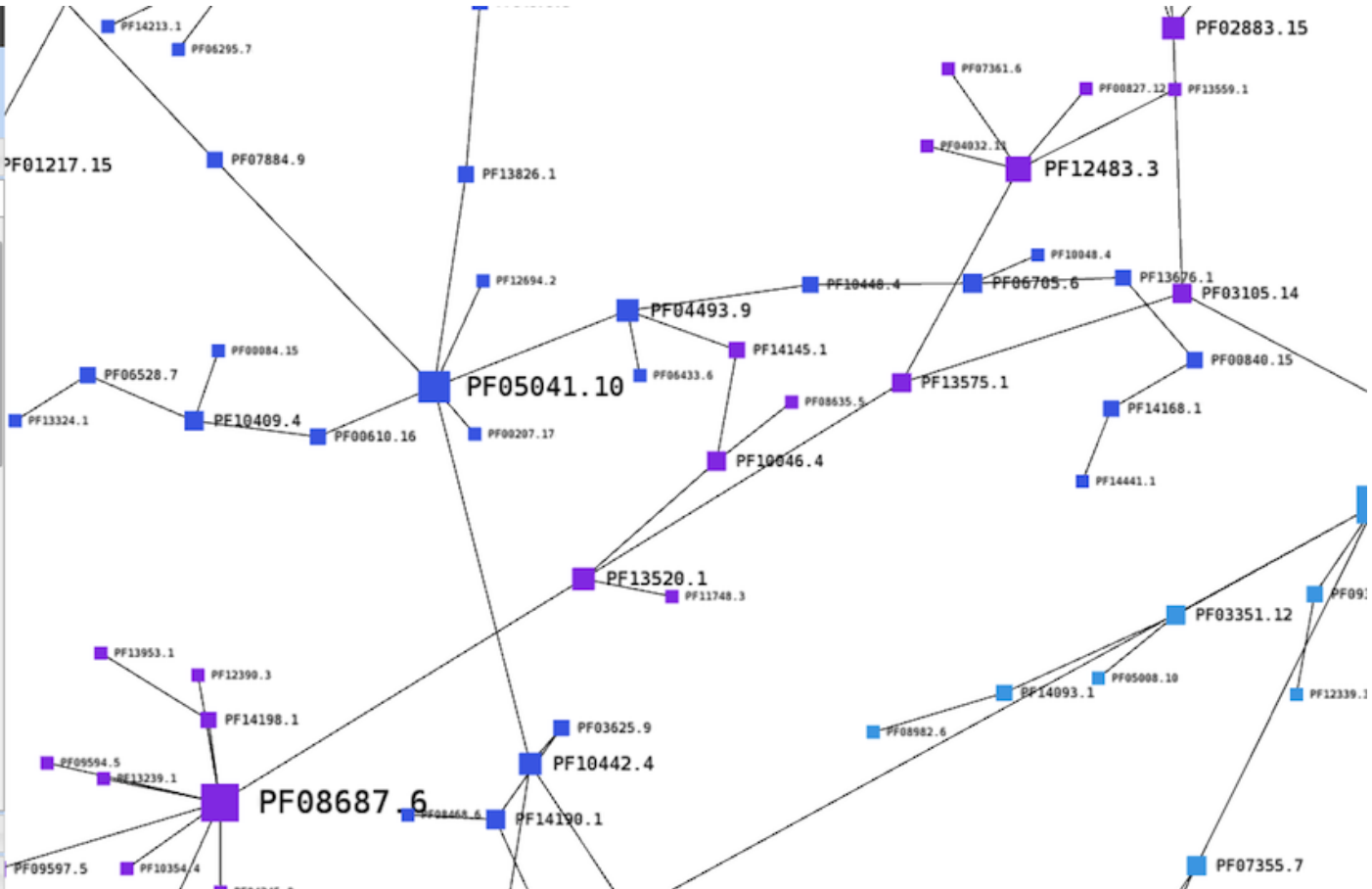
The DEP domain [1] is responsible for mediating intracellular protein targeting and regulation of protein stability in the cell [2-3]. The DEP domain is present in a number of signaling molecules, including Regulator of G protein Signaling (RGS) proteins, and has been implicated in membrane targeting [4-5]. New findings in yeast, however, demonstrate a major role for a DEP domain in mediating the interaction of an RGS protein to the C-terminal tail of a GPCR, thus placing RGS in close proximity with its substrate G protein alpha subunit [6-7].

Ponting CP, Bork P, Trends Biochem Sci 1996;21:245-246.: Pleckstrin's repeat performance: a novel domain in G-protein signaling? [PUBMED:8755244](#)

Martemyanov KA, Lishko PV, Calero N, Keresztes G, Sokolov M, Strissel KJ, Leskov IB, Hopp JA, Kolesnikov AV, Chen CK, Lem J, Heller S, Burns ME, Arshavsky VY, J Neurosci 2003;23:10175-10181.:

▸ Ego Network

▸ Language





- ▶ **Master Thesis: Matthias Stock (Hochschule Bremen)**
- ▶ **Efficient web-based and large-scale visualization of networks**
 - **Outperforms state of the art web tools**



Ocean Sampling Day App

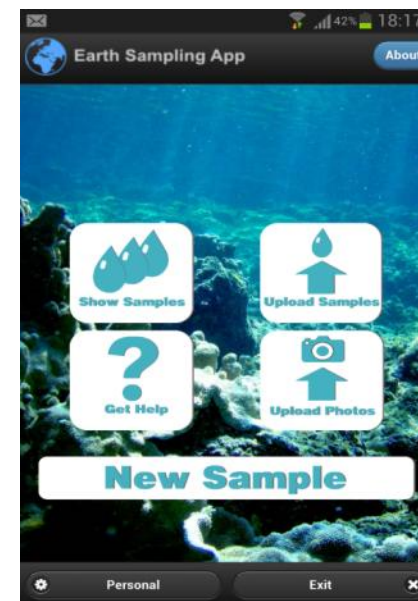
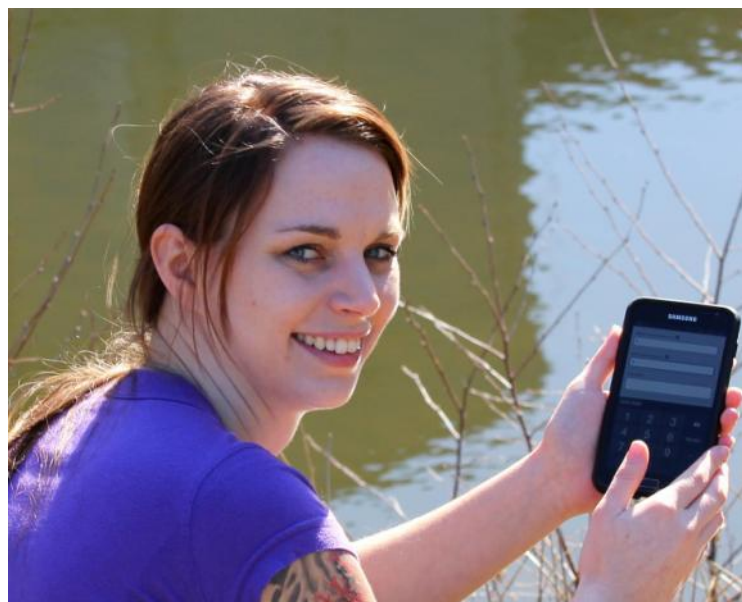


Early, consistent, digital acquisition
of environmental data



ANDROID

iPhone



InterWorks

<https://itunes.apple.com/us/app/osd-citizen/id834353532?mt=8>

<https://play.google.com/store/apps/details?id=com.iw.esa>



Entering Data



o2 - de 18:00

* Time: **i**
2014-03-26 17:49:00

* Latitude: **i**
0

* Longitude: **i**
0

* Accuracy (m): **i**
789654

* Sampling depths (m): **i**
100

Get GPS Location Show on Map

o2 - de 17:48

New Sample (2/4)

* Sample name **i**
Renzo-4

Weather condition: **i**
Clear night

Air temperature °C: **i**

Water temperature °C: **i**

Wind speed (km/h): **i**

Photos(0):
Take photo
From gallery
Preview photos

Back Next

o2 - de 17:48

New Sample (3/4)

Salinity (psu): **i**

Phosphate (mg/l): **i**

Nitrate (mg/l): **i**

Nitrite (mg/l): **i**

pH: **i**

Back Next

o2 - de 17:48

New Sample (4/4)

Biome: **i**
Coastal sea area

Secchi depth (m): **i**

Comment: **i**

Back Done



Features



- ▶ **Allows to take data in the field**
 - **NO internet connection needed**
 - **GSC standards compliant**

o2 - de 18:00

* Time: ⓘ
2014-03-26 17:49:00

* Latitude: ⓘ
0

* Longitude: ⓘ
0

* Accuracy (m): ⓘ
789654

* Sampling depths (m): ⓘ
100

Get GPS Location Show on Map



Megx.net / Micro B3-IS is Open Source



▶ Subversion

- <https://projects.mpi-bremen.de/micro-b3/svn/>

▶ Source Code Browser

- <https://colab.mpi-bremen.de/source/>

▶ Wiki

- <https://colab.mpi-bremen.de/wiki>

▶ Issue Tracker

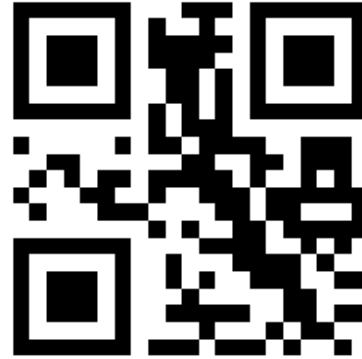
- <https://colab.mpi-bremen.de/its/>



Thanks for your attention



<http://www.microb3.eu>



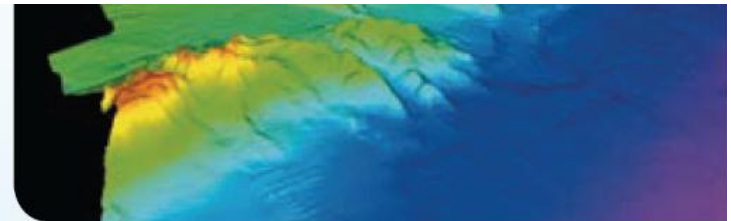
<http://www.oceansamplingday.org>

http://twitter.com/Micro_B3

PROCEEDINGS

Hypotheses come and go but data remain

Santiago Ramón y Cajal,
Nobel Prize 1906



1st Marine Board Forum: Marine data Challenges: from Observation to Information