

# A cloud-based system to enable streamlined access to, and analysis of, continental-scale environmental metagenomics data by non-genomics researchers

**Jeff Christiansen**, Derek Benson, Grahame Bowland, Samuel Chang, Simon Gladman, Gareth Price, Anna Syme, Tamas Szabo, Mike Thang, Andrew Bissett

---

developed by:



---

funded by:



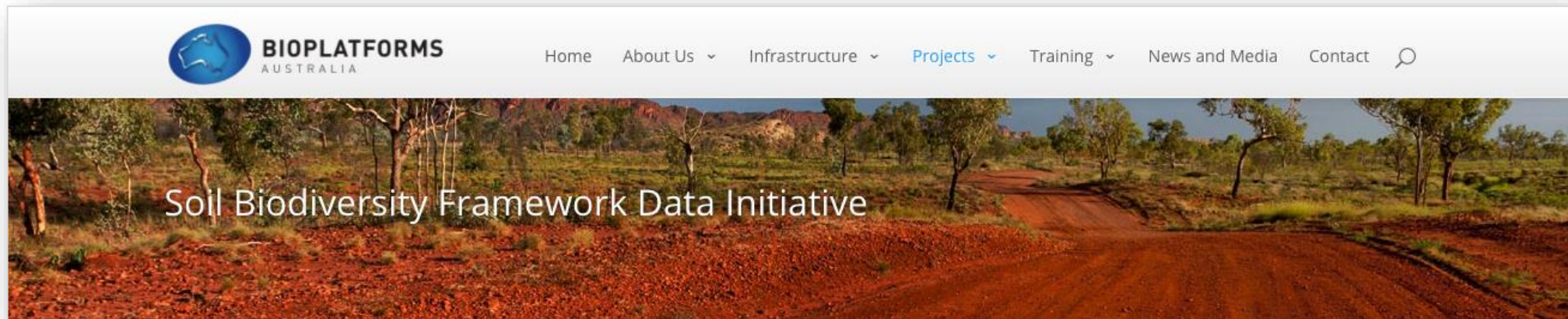
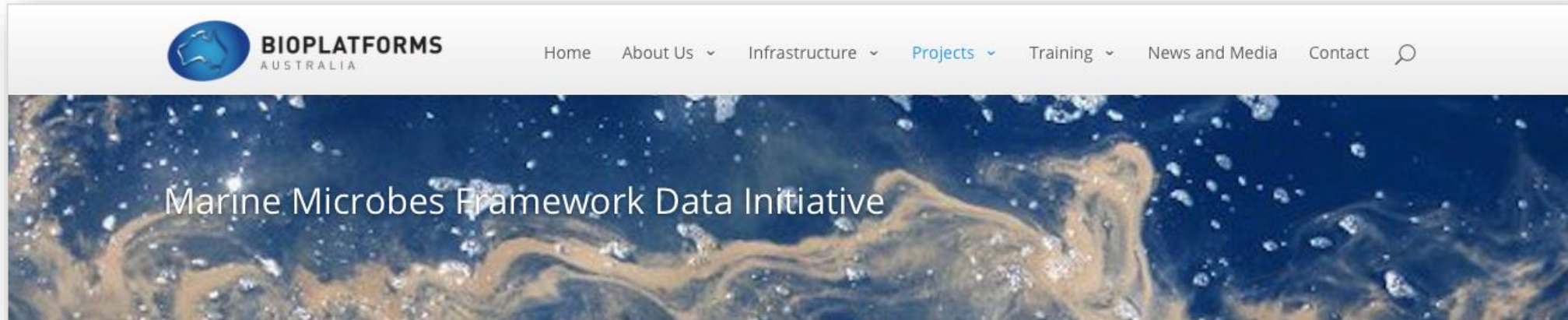
---

supported by:



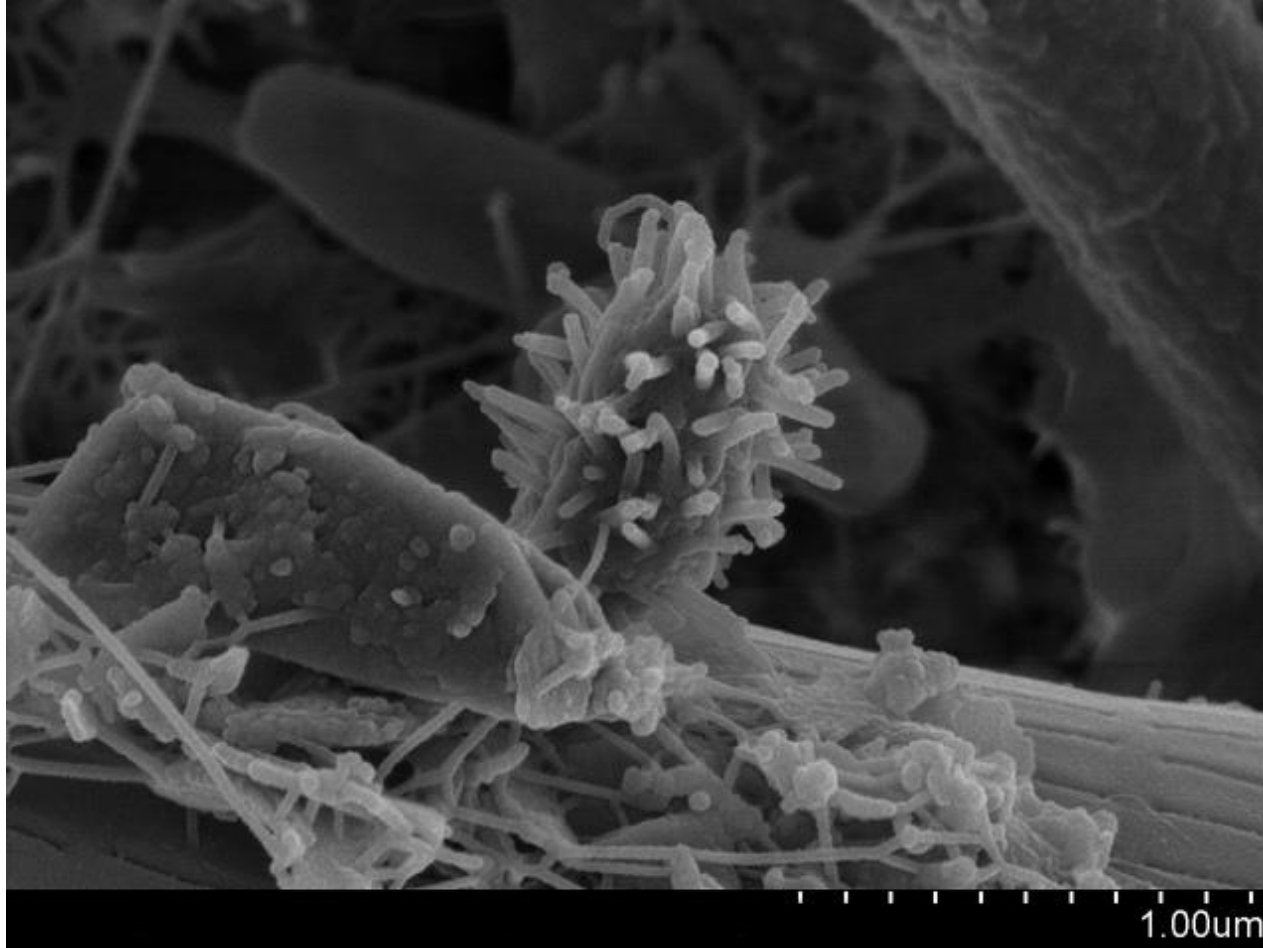
# ARDC funded Bio-RDC project

- Supporting Researchers using Environmental Metagenomics Data



# ARDC funded Bio-RDC project

- The environment is teeming with microorganisms



A community of bacteria growing on a single grain of sand collected from intertidal sediment on a beach near Boston, USA

Image courtesy of the Lewis Lab at Northeastern University. Image created by Anthony D'Onofrio, William H. Fowle, Eric J. Stewart and Kim Lewis. Source: <https://www.flickr.com/photos/adonofrio/4482210139/in/gallery-137201199@N07-72157661521584722/>





## Data format:

**Corresponding biological taxa**  
(Kingdom > Phylum > Class > Order > Family > Genus > Species)

1000s of columns (**samples**)

1000000s of rows  
(sequences)

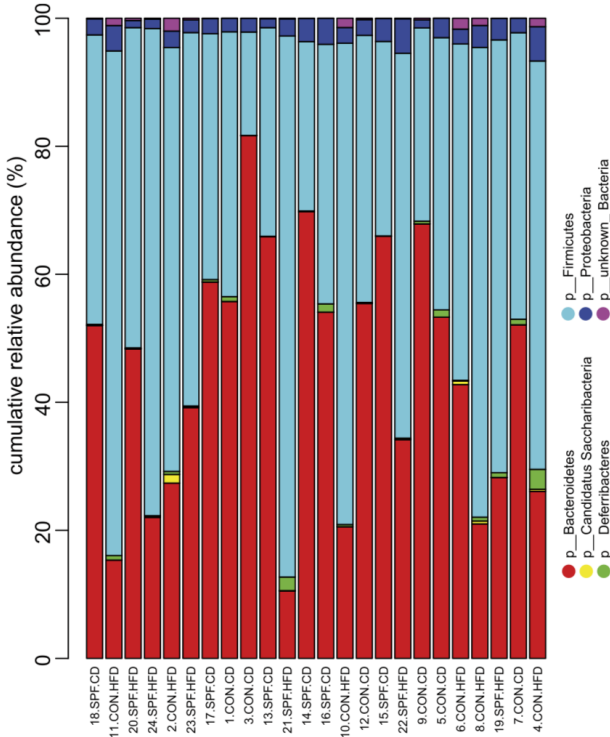
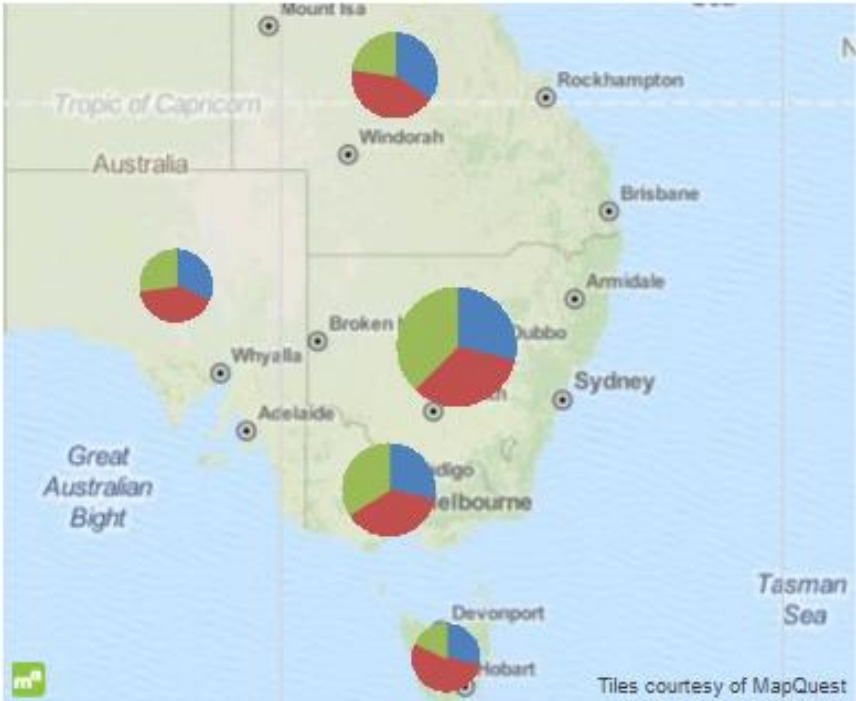
1000000s of rows (sequences)

**Contextual information**  
(e.g. soil, environment features)

## 1000s of samples

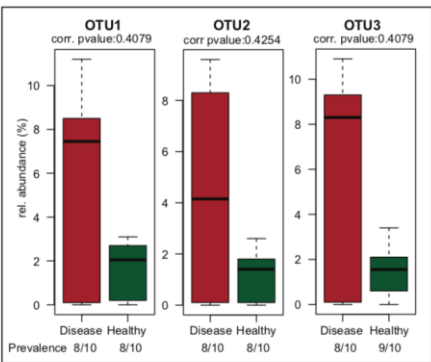
# ARDC funded Bio-RDC project

- Supporting Researchers using Environmental Metagenomics Data
- Streamlining access to data and tools for analysis of that data



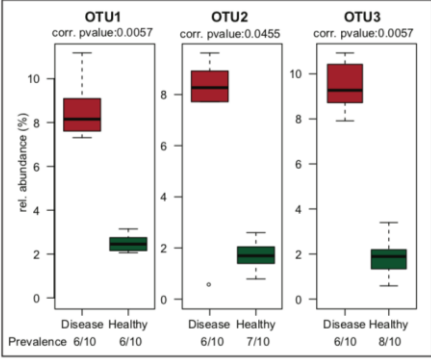
No transformation

	Phenotype	OTU1	OTU2	OTU3
Sample1	Healthy	0	2.3	1.7
Sample2	Healthy	2.1	1.5	0
Sample3	Healthy	0.2	1.7	2.1
Sample4	Healthy	3.1	0.1	1.3
Sample5	Healthy	2.7	1.8	0.6
Sample6	Healthy	2	0	2.1
Sample7	Healthy	0.4	0.8	3.4
Sample8	Healthy	0	1.3	2.3
Sample9	Healthy	2.7	0	1.4
Sample10	Healthy	2.1	2.6	0.1
Sample11	Disease	7.6	0.3	7.9
Sample12	Disease	0	8.3	0.1
Sample13	Disease	8.5	0	10.9
Sample14	Disease	7.3	0.1	10.4
Sample15	Disease	0.3	0	9.3
Sample16	Disease	11.2	8.9	0.3
Sample17	Disease	7.8	0.6	0
Sample18	Disease	0	7.7	9.2
Sample19	Disease	0.1	9.6	8.7
Sample20	Disease	9.1	8.2	0



Not considering near-to-zero values

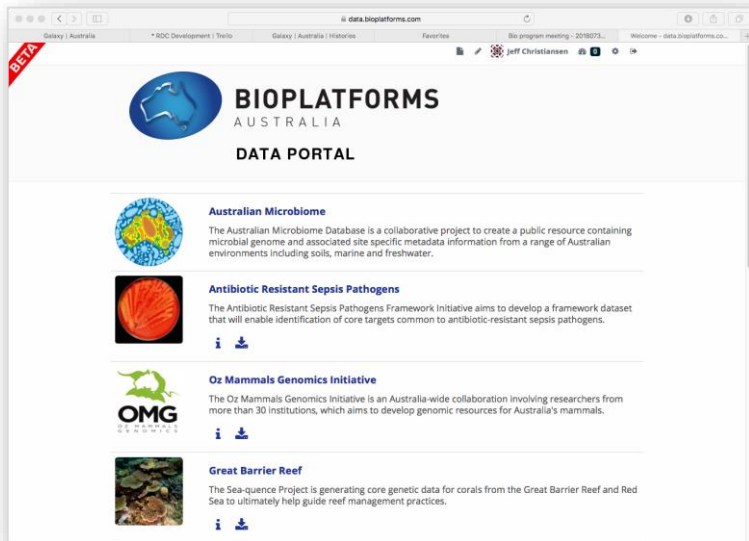
	Phenotype	OTU1	OTU2	OTU3
Sample1	Healthy	NA	2.3	1.7
Sample2	Healthy	2.1	1.5	NA
Sample3	Healthy	NA	1.7	2.1
Sample4	Healthy	3.1	NA	1.3
Sample5	Healthy	2.7	1.8	0.6
Sample6	Healthy	2	NA	2.1
Sample7	Healthy	NA	0.8	3.4
Sample8	Healthy	NA	1.3	2.3
Sample9	Healthy	2.7	NA	1.4
Sample10	Healthy	2.1	2.6	NA
Sample11	Disease	7.6	NA	7.9
Sample12	Disease	NA	8.3	NA
Sample13	Disease	8.5	NA	10.9
Sample14	Disease	7.3	NA	10.4
Sample15	Disease	NA	NA	9.3
Sample16	Disease	11.2	8.9	NA
Sample17	Disease	7.8	0.6	NA
Sample18	Disease	NA	7.7	9.2
Sample19	Disease	NA	9.6	8.7
Sample20	Disease	9.1	8.2	NA



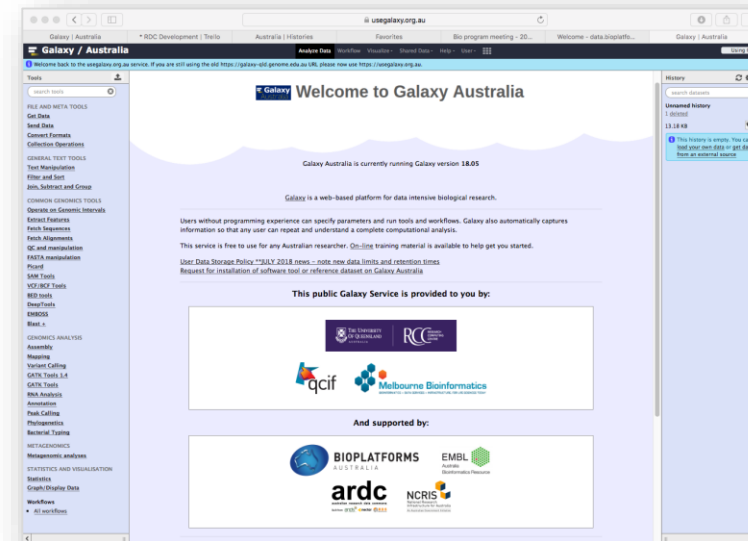
# ARDC funded Bio-RDC project

- Supporting Researchers using Environmental Metagenomics Data
- Streamlining access to data and tools for analysis of that data
- Extending and linking existing data infrastructures:

## Data Storage



## Data Analysis



# ARDC funded Bio–RDC project

- Supporting Researchers using Environmental Metagenomics Data
- Streamlining access to data and tools for analysis of that data
- Extending and linking existing data infrastructures:
- Development Team: CCG, QCIF/Melb Bionf, Andrew Bissett CSIRO, ALA (obs)



# ARDC funded Bio–RDC project

- Supporting Researchers using Environmental Metagenomics Data
- Streamlining access to data and tools for analysis of that data
- Extending and linking existing data infrastructures:
- Development Team: CCG, QCIF/Melb Bionf, Andrew Bissett CSIRO, ALA (obs)
- **Timeline: Feb 2018 – Dec 2018 (10 months)**

# **ARDC funded Bio–RDC project**

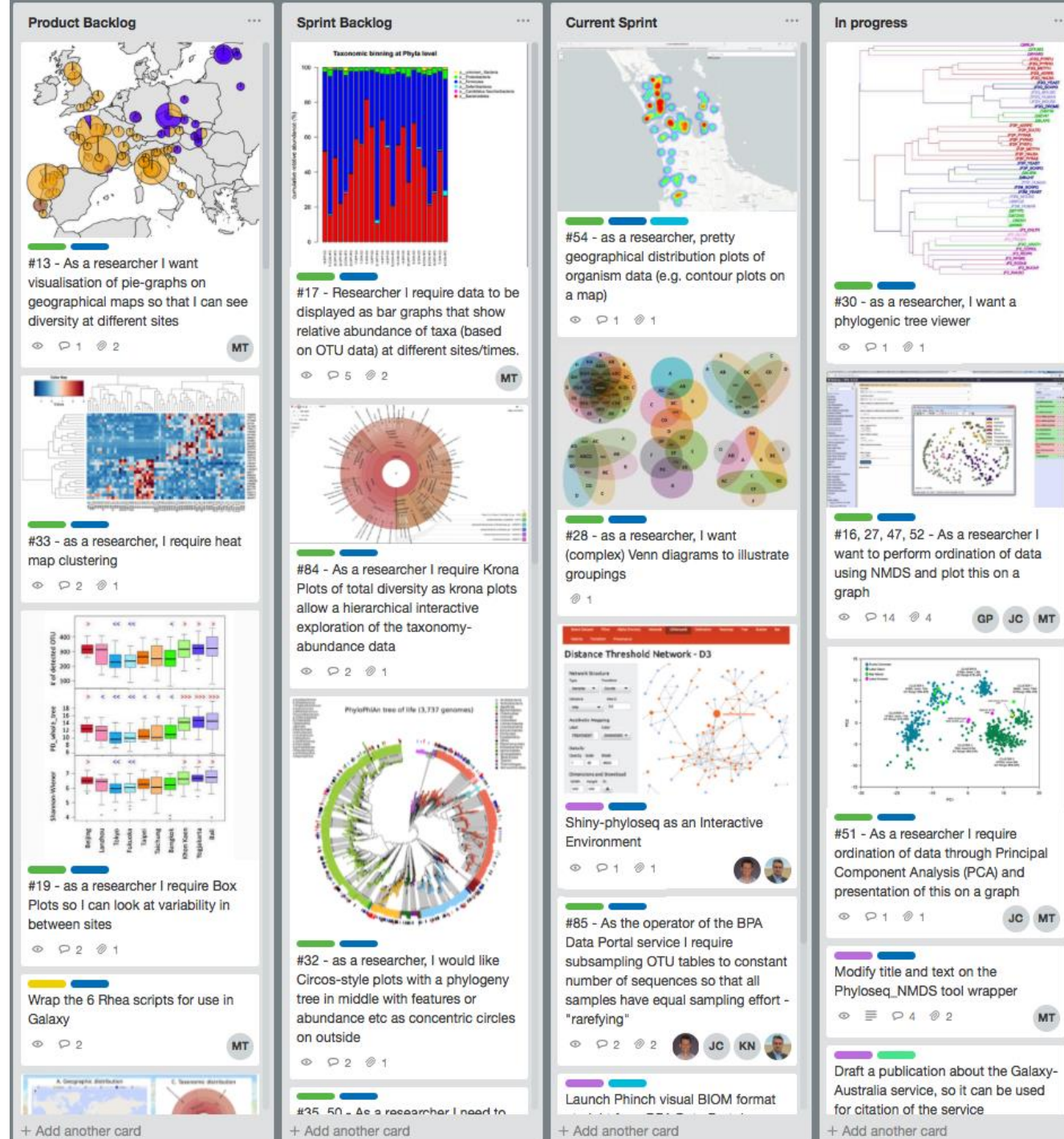
- Researcher's Requirements

# ARDC funded Bio–RDC project

- Researcher's Requirements
- Gathered at BPA coordinated  
Marine Microbes Meeting  
March 2018

# ARDC funded Bio-RDC project

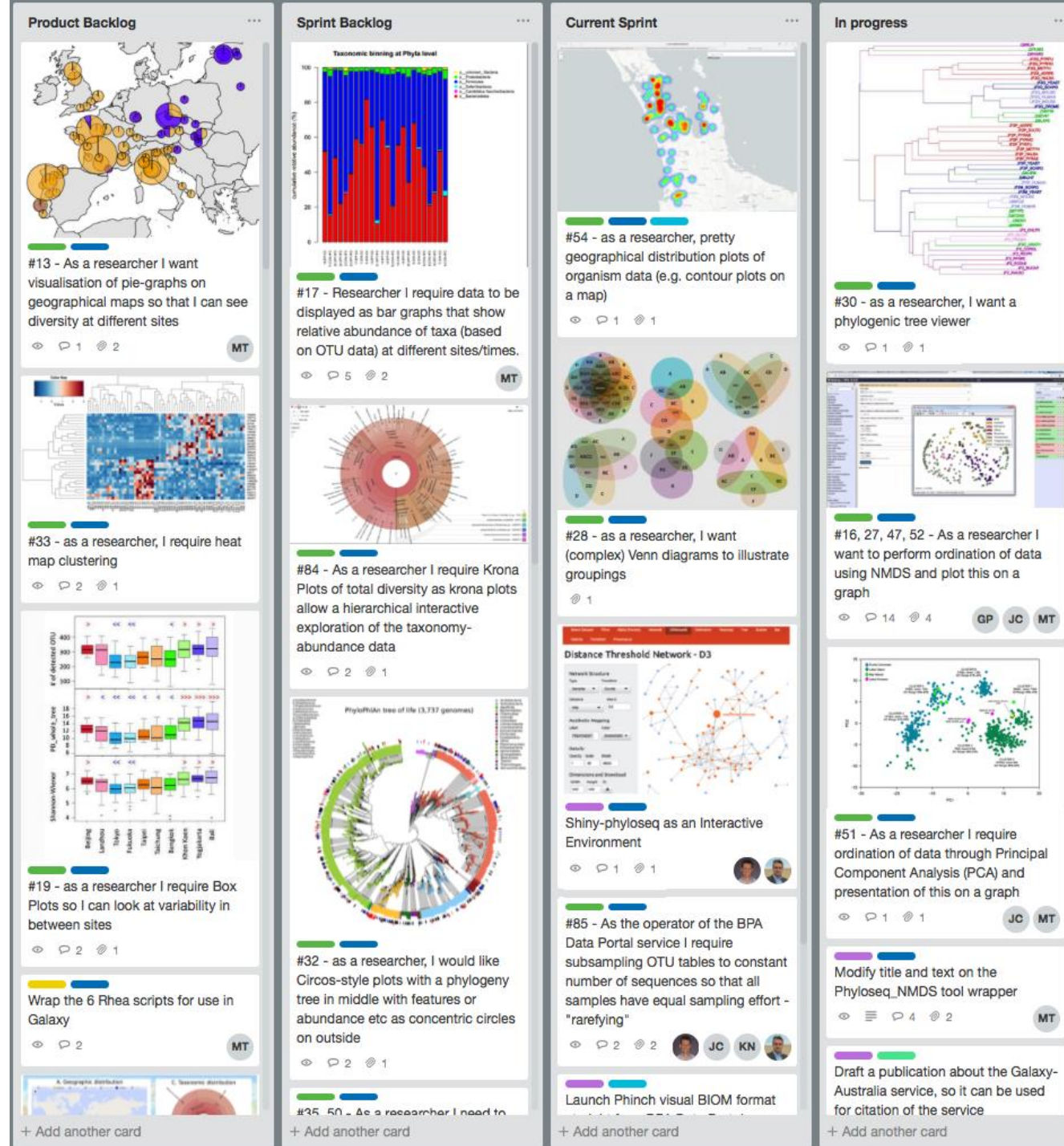
- Researcher's Requirements
- Gathered at BPA coordinated Marine Microbes Meeting March 2018
- Functionality to analyse OTU data across sites/populations:
  - Ordination
  - Clustering
  - Rarefaction
  - Search by Sequence Similarity
  - Plots (Box, Bar, Venn etc)
  - Maps (bubbleplots, heatmaps)
  - Browsing Taxonomic abundance data
  - BYO OTU data



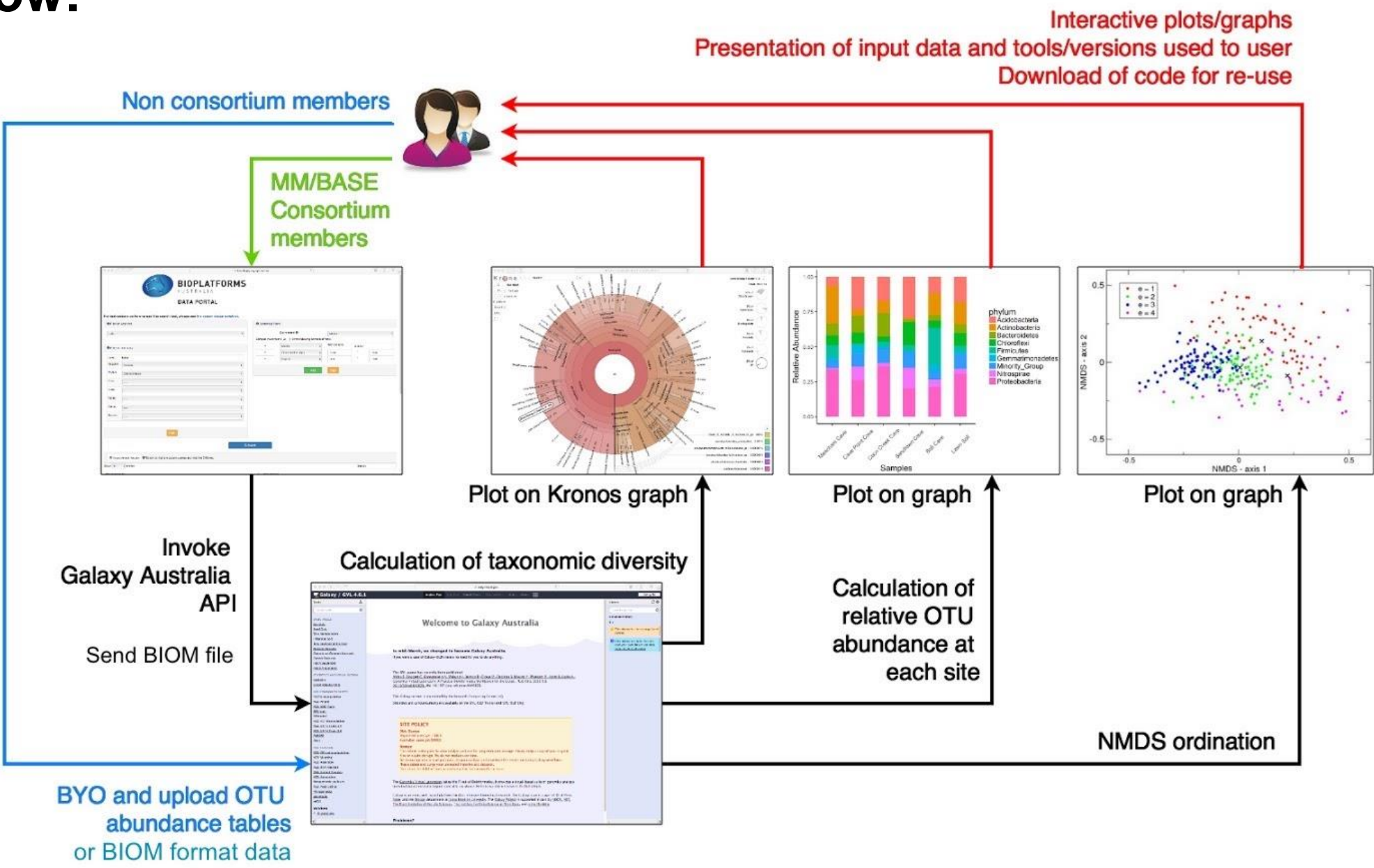


# ARDC funded Bio-RDC project

- Researcher's Requirements
- Gathered at BPA coordinated Marine Microbes Meeting March 2018
- Functionality to analyse OTU data across sites/populations:
  - Ordination
  - Clustering
  - Rarefaction
  - Search by Sequence Similarity
  - Plots (Box, Bar, Venn etc)
  - Maps (bubbleplots, heatmaps)
  - Browsing Taxonomic abundance data
  - BYO OTU data
- Prioritised by Andrew Bissett



# Workflow:



**What have we developed so far?**

# What have we developed so far?

**NOTE: ALL DEMOS ARE DONE ON STAGING SERVERS  
(BPA DATA PORTAL AND GALAXY AUSTRALIA)**

**FUNCTIONALITY WILL BE RELEASED TO PRODUCTION  
SERVERS ONLY AFTER THOROUGH QUALITY ASSURANCE,  
THROUGHOUT Q4 2018**





## DATA PORTAL

Filter by Amplicon ⓘ and Taxonomy ⓘ

Amplicon

is

---

### Taxonomy

Kingdom

is

---

Phylum

is

---

Class

is

---

Order

is

---

Family

is

---

Genus

is

---

Species

is

---

Clear

Contextual Filters ⓘ

Environment ⓘ

is

---

### Contextual Filters

Add

Clear

Search

Show results on Map

Submit to Galaxy

Export Search Results (CSV)

Export Search Results (BIOM)

Account made for BPA Data Portal users  
on Galaxy Australia

Data sent via Galaxy API

Currently this takes a little while

~3mins later...

Species

is

---

Clear

Search

Show results on Map

Submit to Galaxy

Export Search Results (CSV)

Export Search Results (BIOM)

Submission to Galaxy in Progress ...

If you are new to Galaxy Australia, please see this [Getting started guide](#)

BPA Sample ID	BPA Project
<a href="#">7035</a>	Soil
<a href="#">7036</a>	Soil
<a href="#">7049</a>	Soil
<a href="#">7050</a>	Soil
<a href="#">7051</a>	Soil
<a href="#">7052</a>	Soil
<a href="#">7061</a>	Soil
<a href="#">7062</a>	Soil
<a href="#">7067</a>	Soil
<a href="#">7068</a>	Soil

Previous

Page 1 of 31

10 rows

Next

Currently this also takes a little while  
~4mins later...

Plenty of options for increasing speed  
e.g. farming out jobs from specific tools  
to various backends (cloud, HPC etc)



galaxy-aust-dev.genome.edu.au

Bioplatforms OTU Search Facility

Galaxy | Australia

Galaxy / Australia

Analyze Data

Workflow

Visualize

Shared Data

Help

User

Using 2%

✖ This is the galaxy.au staging server. Please email d@uq.edu.au if the are problems with the service.

Tools

phyloseq

Metagenomic analyses

Phyloseq Biom Filtering

biom file filter

Phyloseq Ordination Plot

ordination plotting

FROGSSTAT Phyloseq Multivariate Analysis Of Variance

FROGSSTAT Phyloseq Sample Clustering

of samples using different linkage methods

FROGSSTAT Phyloseq Import Data

from 3 files: biomfile, samplefile, treefile

FROGSSTAT Phyloseq Composition Visualisation

with bar plot and composition plot

FROGSSTAT Phyloseq Alpha Diversity

with richness plot

FROGSSTAT Phyloseq Beta Diversity

distance matrix

FROGSSTAT Phyloseq Structure Visualisation

with heatmap plot and ordination plot

Workflows

All workflows

✓ 1 job has been successfully added to the queue – resulting in the following datasets:

2: Phyloseq Ordination Plot – PCoA (Principal Coordinate Analysis)\_bray (Bray–Curtis distance).html

You can check the status of queued jobs and view the resulting data by refreshing the History pane. When the job has been run the status will change from 'running' to 'finished' if completed successfully or 'error' if problems were encountered.

History

search datasets

Australian Microbiome: amplicon is '27f519r\_bacteria'; phylum is 'Bacteroidetes'; environment is 'Soil'; Vegetation Type is 'Grassland'; Vegetation Type is 'Forest'

2 shown

20.87 MB

2: Phyloseq Ordination Plot – PCoA (Principal Coordinate Analysis)\_bray (Bray–Curtis distance).html

1: AustralianMicrobiome-2018-10-09T081822.biom.zip

# Ordination Plots

Utilises..

- R script from the Phyloseq package...

OPEN ACCESS Freely available online



## phyloseq: An R Package for Reproducible Interactive Analysis and Graphics of Microbiome Census Data

Paul J. McMurdie, Susan Holmes\*

Department of Statistics, Stanford University, Stanford, California, United States of America

### Abstract

**Background:** The analysis of microbial communities through DNA sequencing brings many challenges: the integration of different types of data with methods from ecology, genetics, phylogenetics, multivariate statistics, visualization and testing. With the increased breadth of experimental designs now being pursued, project-specific statistical analyses are often needed, and these analyses are often difficult (or impossible) for peer researchers to independently reproduce. The vast majority of the requisite tools for performing these analyses reproducibly are already implemented in R and its extensions (packages), but with limited support for high throughput microbiome census data.

**Results:** Here we describe a software project, phyloseq, dedicated to the object-oriented representation and analysis of microbiome census data in R. It supports importing data from a variety of common formats, as well as many analysis techniques. These include calibration, filtering, subsetting, agglomeration, multi-table comparisons, diversity analysis, parallelized Fast UniFrac, ordination methods, and production of publication-quality graphics; all in a manner that is easy to document, share, and modify. We show how to apply functions from other R packages to phyloseq-represented data, illustrating the availability of a large number of open source analysis techniques. We discuss the use of phyloseq with tools for reproducible research, a practice common in other fields but still rare in the analysis of highly parallel microbiome census data. We have made available all of the materials necessary to completely reproduce the analysis and figures included in this article, an example of best practices for reproducible research.

**Conclusions:** The phyloseq project for R is a new open-source software package, freely available on the web from both GitHub and Bioconductor.

**Citation:** McMurdie PJ, Holmes S (2013) phyloseq: An R Package for Reproducible Interactive Analysis and Graphics of Microbiome Census Data. PLoS ONE 8(4): e61217. doi:10.1371/journal.pone.0061217

**Editor:** Michael Watson, The Roslin Institute, University of Edinburgh, United Kingdom

**Received:** October 17, 2012; **Accepted:** March 6, 2013; **Published:** April 22, 2013

**Copyright:** © 2013 McMurdie, Holmes. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

**Funding:** This work was supported by grant NIH-R01GM086884. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

**Competing Interests:** The authors have declared that no competing interests exist.

\* E-mail: susan@stat.stanford.edu

# Ordination Plots

Utilises..

- R script from the Phyloseq package

‘wrapped’ for use in the Galaxy GUI

```
library("phyloseq"); packageVersion("phyloseq")
```

```
## [1] '1.22.3'
```

```
data(GlobalPatterns)
library("ggplot2"); packageVersion("ggplot2")
```

```
## [1] '2.2.1'
```

```
library("plyr"); packageVersion("plyr")
```

```
## [1] '1.8.4'
```

```
GP.ord <- ordinate(GP1, "NMDS", "bray")
p1 = plot_ordination(GP1, GP.ord, type="taxa", color="Phylum", title="taxa")
print(p1)
```

```
p2 = plot_ordination(GP1, GP.ord, type="samples", color="SampleType", shape="human")
p2 + geom_polygon(aes(fill=SampleType)) + geom_point(size=5) + ggtitle("samples")
```

```
dist = "bray"
ord_meths = c("DCA", "CCA", "RDA", "DPCoA", "NMDS", "MDS", "PCoA")
plist = llply(as.list(ord_meths), function(i, physeq, dist){
  ordi = ordinate(physeq, method=i, distance=dist)
  plot_ordination(physeq, ordi, "samples", color="SampleType")
}, GP1, dist)
```

## Phyloseq Ordination Plot Ordination Plotting (Galaxy Version 1.22.3.2)

### Choose an input file type

BIOM File

### Input File



1: BiomExport-2018-07-31T135720.biom.zip

### Select variable for ordination

agrochemical\_additions

### Select Ordination method

- ☒ NMDS(Non-metric Multidimensional Scalling)
- ☐ DCA (Detrended Correspondence Analysis)
- ☐ CCA (Constrained Correspondence Analysis)
- ☐ RDA (Redundancy Analysis)
- ☐ DPCoA (Double Principle Coordinate Analysis)
- ☐ MDS (Multidimensional Scalling)
- ☐ PCoA (Principal Coordinate Analysis)

### Select Distance method

- ☒ bray (Bray-Curtis distance)
- ☐ gower (Gower's distance)
- ☐ jds (Jensen-Shannon Divergence)
- ☐ unifrac (unweighted UniFrac distance)
- ☐ wunifrac (weighted-UniFrac distance)

**Other tools to include:**



## **Other tools to include:**

Many are available in the Galaxy Toolshed “App Store”, so trivial to install locally:

# Other tools to include:

Many are available in the Galaxy Toolshed “App Store”, so trivial to install locally:

e.g. Mothur, QIIME, Krona have been installed

The image displays three overlapping screenshots of the Galaxy Australia web interface, showcasing various bioinformatics tools available for use.

**Left Window (Mothur):** The interface shows the "Tools" sidebar on the left with a search bar containing "mothur". The main panel displays the "Libshuff Cramer-von Mises tests communities for the same" tool. The "Tools" sidebar lists various Mothur tools like "Get.otulist", "Get.group", "Summary.single", "Summary.seqs", "Make.fastq", "Get.rabund", "Primer.design", "Cluster.fragments", "List.otulabels", "Pairwise.seqs", "Taxonomy-to-Krona", "Get.sabund", "Libshuff Cramer-von Mises", and "Make.Design". The main panel shows the "phylip - Distance Matrix" tool with input fields for "group - Groups", "groups - Groups to consider", "Iterations - Number of iterations to try (default 10000)", "sim - Matrix values are Similary instead of Distance", "form - Select a Clustering Method", "Output logfile?", and "Execute" button.

**Middle Window (QIIME):** The interface shows the "Tools" sidebar on the left with a search bar containing "qiime". The main panel displays the "Filter OTUs from an OTU table" tool. The "Tools" sidebar lists various QIIME tools like "Run a core set of QIIME diversity analyses", "Compute beta diversity distance matrices and generate PCoA plots", "Make phylogeny", "Filter OTUs from an OTU table", "Calculate alpha diversity on each sample in an OTU table", "Create three-dimensional PCoA plots", "Count the sequences in a fasta file", "Check user's metadata mapping file", "Split fastq libraries", and "Summarize taxa and store results". The main panel shows the "Filter OTUs from an OTU table" tool with input fields for "Input OTU table", "Minimum total observation count", "Fraction of the total observation", "Maximum total observation count", "Minimum number of samples an OTU", "Maximum number of samples an OTU", "Do you want to exclude a predefined OTU", and "Execute" button.

**Right Window (Krona):** The interface shows the "Tools" sidebar on the left with a search bar containing "krona". The main panel displays the "Krona" tool, which is a circular phylogenetic tree visualization of fungal data. The "Tools" sidebar lists various Krona tools like "FILE AND META TOOLS", "GENERAL TEXT TOOLS", "COMMON GENOMICS TOOLS", "QC and manipulation", "FASTA manipulation", "Picard", "SAM Tools", "VCF/BCF Tools", "BED tools", "DEEPTools", "EMBOSS", "Blast +", "GENOMICS ANALYSIS", "Assembly", "Mapping", "Variant Detection", "Variant Calling", "GATK Tools", "RNA Analysis", "Annotation", and "Peak Calling". The main panel shows the "Krona" tool with a large circular phylogenetic tree visualization of fungal data. The "Tools" sidebar lists various Krona tools like "FILE AND META TOOLS", "GENERAL TEXT TOOLS", "COMMON GENOMICS TOOLS", "QC and manipulation", "FASTA manipulation", "Picard", "SAM Tools", "VCF/BCF Tools", "BED tools", "DEEPTools", "EMBOSS", "Blast +", "GENOMICS ANALYSIS", "Assembly", "Mapping", "Variant Detection", "Variant Calling", "GATK Tools", "RNA Analysis", "Annotation", and "Peak Calling". The main panel shows the "Krona" tool with a large circular phylogenetic tree visualization of fungal data. The "Tools" sidebar lists various Krona tools like "FILE AND META TOOLS", "GENERAL TEXT TOOLS", "COMMON GENOMICS TOOLS", "QC and manipulation", "FASTA manipulation", "Picard", "SAM Tools", "VCF/BCF Tools", "BED tools", "DEEPTools", "EMBOSS", "Blast +", "GENOMICS ANALYSIS", "Assembly", "Mapping", "Variant Detection", "Variant Calling", "GATK Tools", "RNA Analysis", "Annotation", and "Peak Calling". The main panel shows the "Krona" tool with a large circular phylogenetic tree visualization of fungal data.

## **Other tools to include:**

We can also 'wrap' any R, python, or other script/command:

# Other tools to include:

We can also ‘wrap’ any R, python, or other script/command:

- e.g.
- More Phyloseq functions
  - Rhea pipeline (for normalization, alpha diversity, beta diversity, Taxonomic binning, Serial Group Comparisons, Correlations):

PeerJ

Submitted 29 July 2016  
Accepted 28 November 2016  
Published 11 January 2017

Corresponding authors  
Ilias Lagkovidatos,  
ili.lagkovidatos@gmail.com  
Thomas Clavel,  
clavetom@gmail.com

Academic editor  
Bianca R. Landa

Additional Information and  
Declarations can be found on  
page 14

DOI 10.7717/peerj.2836

© Copyright  
2017 Lagkovidatos et al.

Distributed under  
Creative Commons CC-BY 4.0

OPEN ACCESS

### Rhea: a transparent and modular R pipeline for microbial profiling based on 16S rRNA gene amplicons

Ilias Lagkovidatos, Sandra Fischer, Neeraj Kumar and Thomas Clavel  
ZIEL—Core Facility Microbiome/NGS, Technical University of Munich, Freising, Germany

**ABSTRACT**

The importance of 16S rRNA gene amplicon profiles for understanding the influence of microbes in a variety of environments coupled with the steep reduction in sequencing costs led to a surge of microbial sequencing projects. The expanding crowd of scientists and clinicians wanting to make use of sequencing datasets can choose among a range of multipurpose software platforms, the use of which can be intimidating for non-expert users. Among available pipeline options for high-throughput 16S rRNA gene analysis, the R programming language and software environment for statistical computing stands out for its power and increased flexibility, and the possibility to adhere to most recent best practices and to adjust to individual project needs. Here we present the Rhea pipeline, a set of R scripts that encode a series of well-documented choices for the downstream analysis of Operational Taxonomic Units (OTUs) tables, including normalization steps, alpha- and beta-diversity analysis, taxonomic composition, statistical comparisons, and calculation of correlations. Rhea is primarily a straightforward starting point for beginners, but can also be a framework for advanced users who can modify and expand the tool. As the community standards evolve, Rhea will adapt to always represent the current state-of-the-art in microbial profiles analysis in the clear and comprehensive way allowed by the R language. Rhea scripts and documentation are freely available at <https://lagkovidatos.github.io/Rhea>.

**Subjects** Bioinformatics, Ecology, Microbiology

**Keywords** 16S rRNA gene, Microbiome analysis, R script, Microbial profile, Host-microbe interactions, Statistical analysis, Alpha-diversity, Beta-diversity, Correlations

**INTRODUCTION**

The analysis of complex microbial communities by high-throughput sequencing of 16S rRNA gene amplicons has become very popular. However, this rapid popularization is in stark contrast to the slow transfer of knowledge about data analysis. The hope generated by pioneering next-generation sequencing studies community was high, but recent findings have highlighted the needs for standardization in the field (Clavel, Lagkovidatos & Hergst, 2016; Hergst, Reichel & Gesner, 2016; Sinha et al., 2015; Walker et al., 2015). With respect to data processing, important parameters such as the choice of clustering methods and thresholds of relative abundances must be considered for proper analysis of high-throughput 16S rRNA gene amplicon datasets (Edgar, 2013; Martinez, Muller & Walter, 2013). The common output of raw sequence processing is a contingency table with

**A**

**B**

**Coefficient of Variation (%)**

**A**

**B**

**C**

Phenotype	OTU1	OTU2	OTU3
Sample1	Healthy	0.23	1.7
Sample2	Healthy	2.1	1.5
Sample3	Healthy	0.2	0.7
Sample4	Healthy	3.1	0.1
Sample5	Healthy	2.7	1.8
Sample6	Healthy	2	0
Sample7	Healthy	0.4	0.8
Sample8	Healthy	0	1.3
Sample9	Healthy	2.7	0
Sample10	Healthy	7.1	2.6
Sample11	Disease	7.6	0.3
Sample12	Disease	0	8.3
Sample13	Disease	8.5	0
Sample14	Disease	7.3	0.1
Sample15	Disease	0.3	0
Sample16	Disease	11.2	0.9
Sample17	Disease	7.8	0.6
Sample18	Disease	0	7.7
Sample19	Disease	11.1	0.6
Sample20	Disease	9.1	8.2

**OTU1**  
corr. p-value 0.4079

**OTU2**  
corr. p-value 0.4254


**OTU3**  
corr. p-value 0.4079



**OTU1**  
corr. p-value 0.0057



**OTU2**  
corr. p-value 0.0465

**OTU3**  
corr. p-value 0.0057


# Three user types to support


 #35, 50 - As a researcher I need to export or be able to re-access the code I used in an analysis in this system, so I can re-deploy/re-use the code in future analyses - on this or other systems

  1


  N




R/python code users

 The system to be deployed must support users who require a Graphical User Interface to conduct complex analyses

 1

GUI users

 #93 - As a researcher, I require a 'one click biogeography' service where I can upload a sequence, and retrieve information about samples where species correspond to that sequence are found, as well as information on the taxonomic diversity at those sites, so I can quickly assess environmental metagenomics survey data that is relevant to my own species of interest


  2  2

One click

# Three user types to support


#35, 50 - As a researcher I need to export or be able to re-access the code I used in an analysis in this system, so I can re-deploy/re-use the code in future analyses - on this or other systems

1



N

R/python code users



The system to be deployed must support users who require a Graphical User Interface to conduct complex analyses

1

GUI users



#93 - As a researcher, I require a 'one click biogeography' service where I can upload a sequence, and retrieve information about samples where species correspond to that sequence are found, as well as information on the taxonomic diversity at those sites, so I can quickly assess environmental metagenomics survey data that is relevant to my own species of interest


2





2

One click



# Three user types to support

 #35, 50 - As a researcher I need to export or be able to re-access the code I used in an analysis in this system, so I can re-deploy/re-use the code in future analyses - on this or other systems

  1  

R/python code users

The system to be deployed must support users who require a Graphical User Interface to conduct complex analyses

 1

GUI users

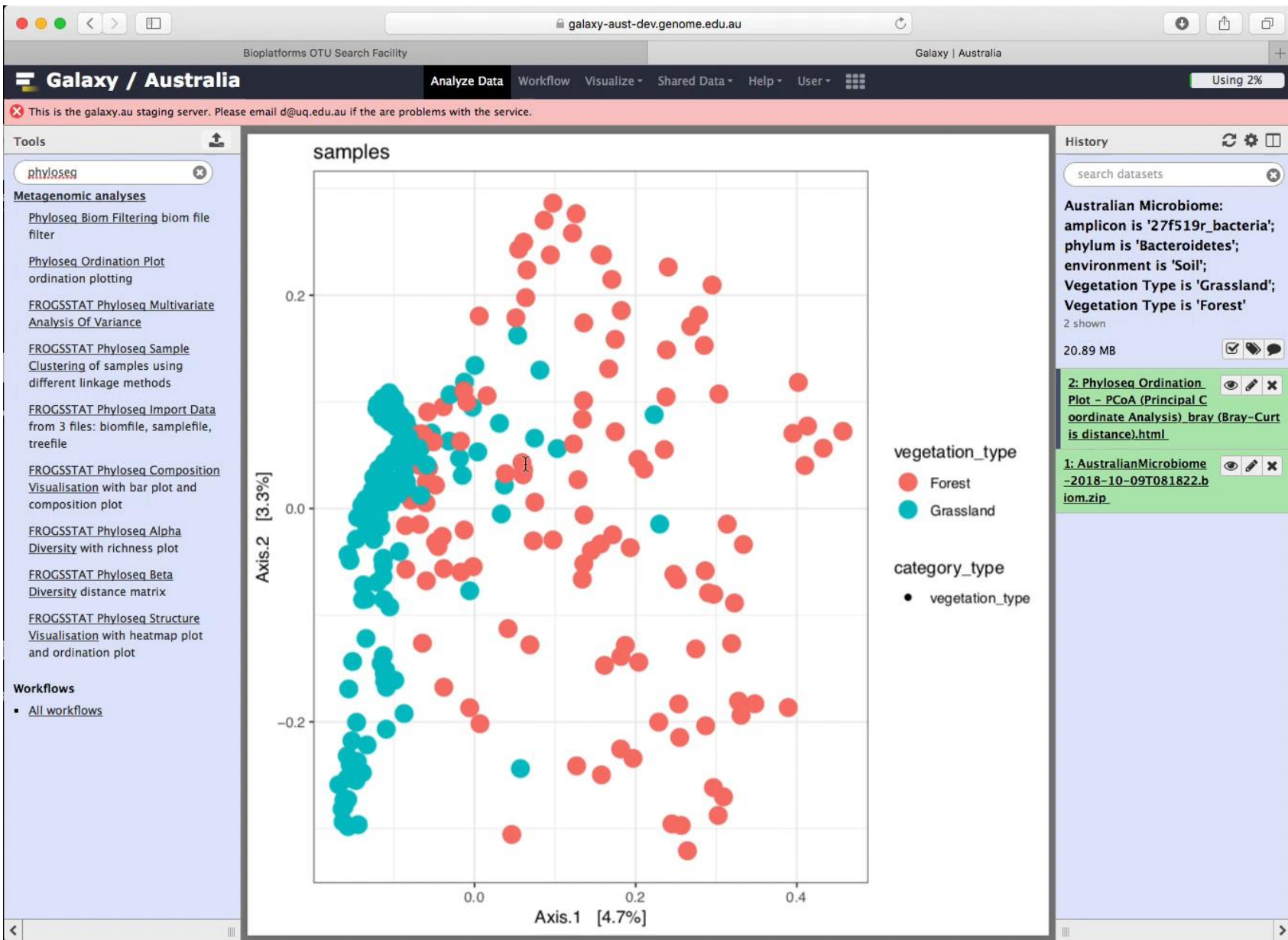
 

#93 - As a researcher, I require a 'one click biogeography' service where I can upload a sequence, and retrieve information about samples where species correspond to that sequence are found, as well as information on the taxonomic diversity at those sites, so I can quickly assess environmental metagenomics survey data that is relevant to my own species of interest

  2  2

One click


# Direct launch of R-studio in the hosted Galaxy environment



# Three user types to support


#35, 50 - As a researcher I need to export or be able to re-access the code I used in an analysis in this system, so I can re-deploy/re-use the code in future analyses - on this or other systems

1



N

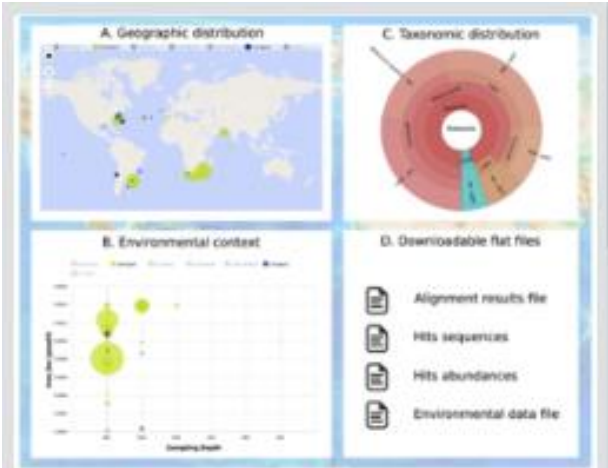
R/python code users



The system to be deployed must support users who require a Graphical User Interface to conduct complex analyses

1

GUI users



#93 - As a researcher, I require a 'one click biogeography' service where I can upload a sequence, and retrieve information about samples where species correspond to that sequence are found, as well as information on the taxonomic diversity at those sites, so I can quickly assess environmental metagenomics survey data that is relevant to my own species of interest

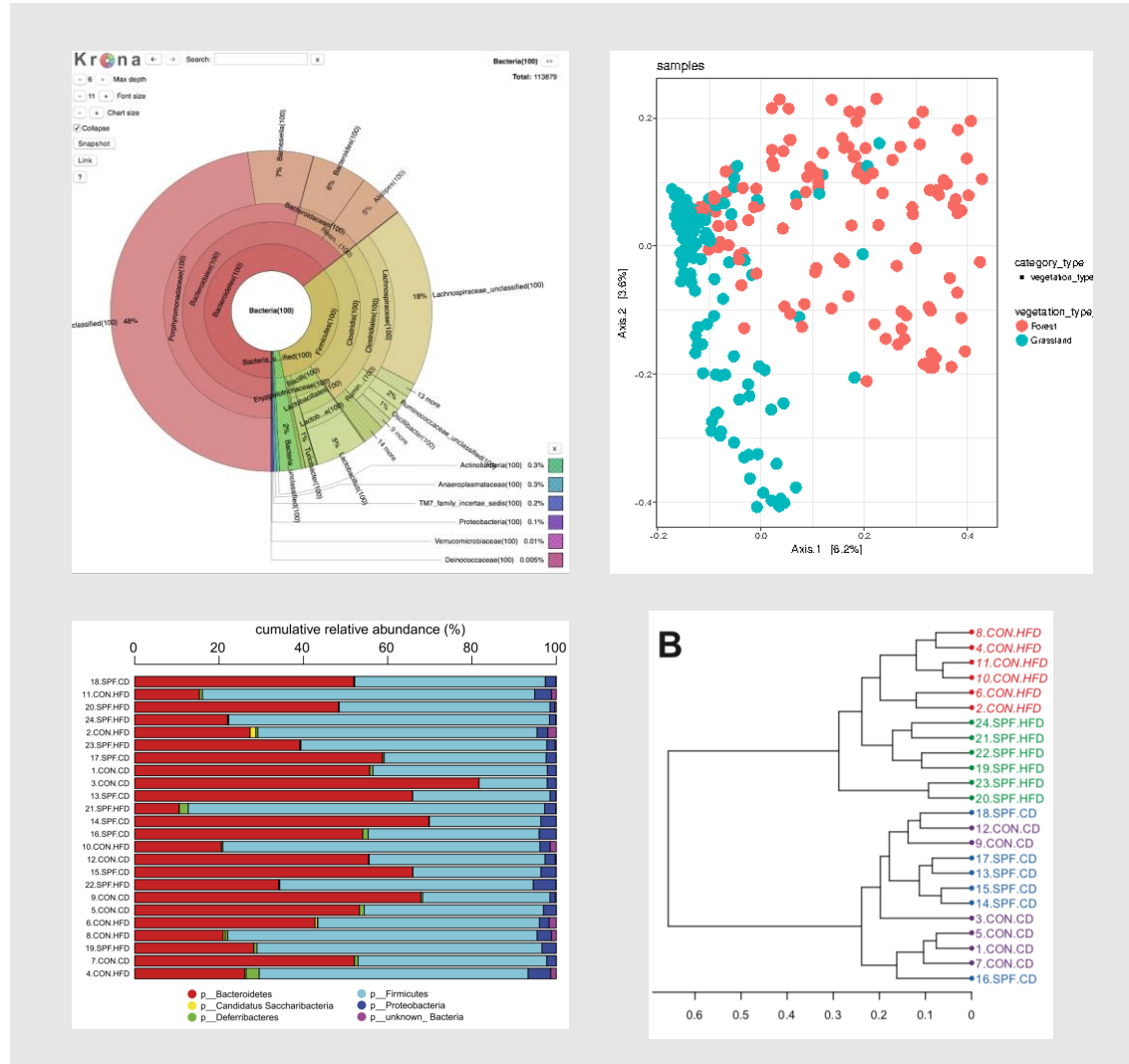
2

2

One click

# Generating Reports

- Can simply invoke multiple Galaxy Tools at once as a 'workflow'
- Then present the outputs to the user in a webpage e.g.
- Could send a link via email



# Status (I)

- Eight months into ten month project
- Demo system available for testing
- Recent/upcoming usability sessions (Brisbane, Melbourne, Hobart)
- As tools/functions pass QC, they will be deployed to production servers
- Short-list of tools/functions to add (sequence search, bar plots, bubble plots)
- Development of on-line training material  
(to be available via EcoED and Galaxy!Training portals)
- One three-hour training session planned - Nov 14<sup>th</sup> via EMBL-ABR 'hybrid' delivery model at several sites (currently: SA, Tas, NSW, Vic, Qld)

## Status (II)

- New commitment from CSIRO, BPA, Cth Dept Environment (Director of National Parks), iMOS) to develop a comprehensive national soil, marine and freshwater microbiome survey
- Ongoing and critical need to make the information within metagenomics data accessible to wide range of user communities:
  - Researchers
  - Land/Water Managers
  - National Parks and Conservation Organisations
  - Farmers/Fisheries
- The system developed represents the initial foundations for such a cloud-based analysis tool for environmental metagenomics data



# Development



Jeff Christiansen  
Derek Benson  
Gareth Price  
Mike Thang



Grahame Bowland  
Tamas Szabo  
Samuel Chang



Simon Gladman  
Anna Syme



Andrew Bissett  
Sophie Mazard



Michael Hope  
Corinna Paeper

---

<https://bioscience-rdc.blogspot.com>

---

# Funding

