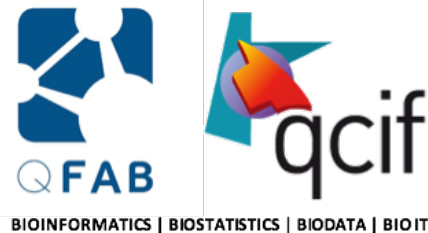




Galaxy Australia - lessons learnt from running a nationally resourced service



Biosciences: the nature of the Australian research community

30,000 health/biosciences researchers

18,000 health/biosciences HDR students

48,000 health/biosciences PG course work students

(163,000 + 40,000 =) 200,000 health/biosciences UG students

1,000 to 1,500 bioinformatician/computational biologists

=> the IT smarts need to be packaged up and delivered cleanly

Galaxy / Australia

Analyze Data Workflow Visualize Shared Data Help User

Using 66%

Tools

search tools

FILE AND META TOOLS

Get Data

Send Data

Collection Operations

GENERAL TEXT TOOLS

Text Manipulation

Filter and Sort

Join, Subtract and Group

GENOMIC FILE MANIPULATION

FASTA/FASTQ

FASTQ Quality Control

SAM/BAM

BED

VCF/BCF

Convert Formats

COMMON GENOMICS TOOLS

Operate on Genomic Intervals

Extract Features

Fetch Sequences/Alignments

GENOMICS ANALYSIS

Assembly

Annotation

Mapping

Variant Calling

ChIP-seq

RNA-seq

Multiple Alignments

Bacterial Typing

Galaxy AUSTRALIA

News

Aug 16, 2019
Galactic News August 2019

Aug 7, 2019
Galaxy Australia moves to new hardware

Jun 25, 2019
Galaxy Australia upgraded to Galaxy version 19.05

Jun 7, 2019
Galaxy Australia wins three Queensland iAwards

May 2, 2019
Text processing tools disabled

Apr 9, 2019
Galaxy Australia upgraded to Galaxy version 19.01

Events and Workshops

Jul 29, 2019 - Aug 2, 2018
Galaxy training workshops Brisbane - July - August 2019

Jul 1, 2019 - Jul 6, 2019
2019 Galaxy Community Conference (GCC2019)

Apr 1, 2019 - Apr 5, 2018
Galaxy training workshops Brisbane - April 2019

Mar 21, 2019 - Mar 26, 2018
Galaxy training workshops Melbourne - March 2019

Feb 21, 2019
GTN CoFest on Training Material

Jan 28, 2019 - Feb 1, 2019
2019 Galaxy Admin Training

Galaxy Australia Jobs (Last 12 hours)

Acknowledgements

We are pleased to maintain high connectivity and provide high quality data analysis

History

search datasets

Day 5: Metagenomics extended

99 shown, 50 deleted, 49 hidden

2.29 GB

197: FastQC on data 34: RawData

196: FastQC on data 34: Webpage

194: Krona on collection 192: H TML
a list with 1 item

192: Taxonomy-to-Krona on collection 153: krona-formatted taxonomy file
a list with 1 item

190: Tree.shared on data 176: tre

178: Heatmap.sim on collection 166: heatmap.sim.svg
a list with 6 items

169: Venn on data 167: svg
a list with 1 item

167: Collapse Collection on data 162

164: Rarefaction plot
a list with 1 item

159: Summary.single on data 151: summary

156: Sub.sample on data 151: sub.sample.shared
a list with 1 item

154: Count.groups on data 151: group.count

Galaxy / Australia
Analyze Data Workflow Visualize Shared Data Help User
Using 56%

Tools

- Inputs
- FILE AND META TOOLS
- Get Data
- Send Data
- Collection Operations
- GENERAL TEXT TOOLS
- Text Manipulation
- Filter and Sort
- Join, Subtract and Group
- GENOMIC FILE MANIPULATION
- FASTA/FASTQ
- FASTQ Quality Control
- SAM/BAM
- BED
- VCF/BCF
- Convert Formats
- COMMON GENOMICS TOOLS
- Operate on Genomic Intervals
- Extract Features
- Fetch Sequences/Alignments
- GENOMICS ANALYSIS
- Assembly
- Annotation

Unnamed workflow

Quality Checking
Search against known database
Create Reference
BIOM format conversion
Normalisation
Plots

The workflow consists of several interconnected tool modules:

- Quality Checking:** Includes tools like FASTQC, Trimmomatic, and FastQC.
- Search against known database:** Includes BLAST, BLAST+ (NCBI), and BLAST+ (local).
- Create Reference:** Includes Bowtie2, BWA-MEM, and BWA-SW.
- BIOM format conversion:** Includes BIOM-format-converter, BIOM-format-converter (old), and BIOM-format-converter (new).
- Normalisation:** Includes DESeq2, edgeR, and limma.
- Plots:** Includes Rplots, Rplots (old), and Rplots (new).

Details

Edit Workflow Attributes

Name:
F1000 workflow

Version:
Version 0, 0 steps (active)

Tags:
BIOM X

to make it easy to search for and find items with the same tag.

Annotation / Notes:
Describe or add notes to workflow;
annotations are available when a workflow is viewed.

A Brief History in Time

- Genomics Virtual Lab for public and private VLs, with Galaxy as a managed public service
 - Galaxy-Melb, Galaxy-Qld and Galaxy-Tut
 - Tool version alignment / tutorial consequences
 - Duplicate staff roles
 - Tut was re-imaged over and over, no longevity to user accounts
- Rebranded Gal-Qld to Galaxy Australia and moved to org.au
 - Staged migration of Melb to Australia
- Brought online 2 x Pulsar (repurpose Galaxy-Melb allocation and new allocation at NCI Canberra)
- Hybrid training events: 4 run nationally with local facilitators and a lead trainer, covering 11 sites and 1064 participants
- Policy, policy and policy – have a lot to thank Galaxy Europe for forging new ground

Solutions for Data Analysis

- **Freeware**

- Galaxy
- R Studio
- Command Line / HPC

Most equivocal solution

Galaxy Australia user numbers (as of Sept 2018 - 2268) as CLC-Bio users is a difference of 750k funds (+750K in kind) vs approx. \$11million annual licence fees

This does not include the cost of computers to support CLC-Bio installations

- **Commercial**

- **Office**

- Excel

- **Agilent**

- Cartagenia Bench Lab for Molecular Pathology

- **Illumina**

- BaseSpace

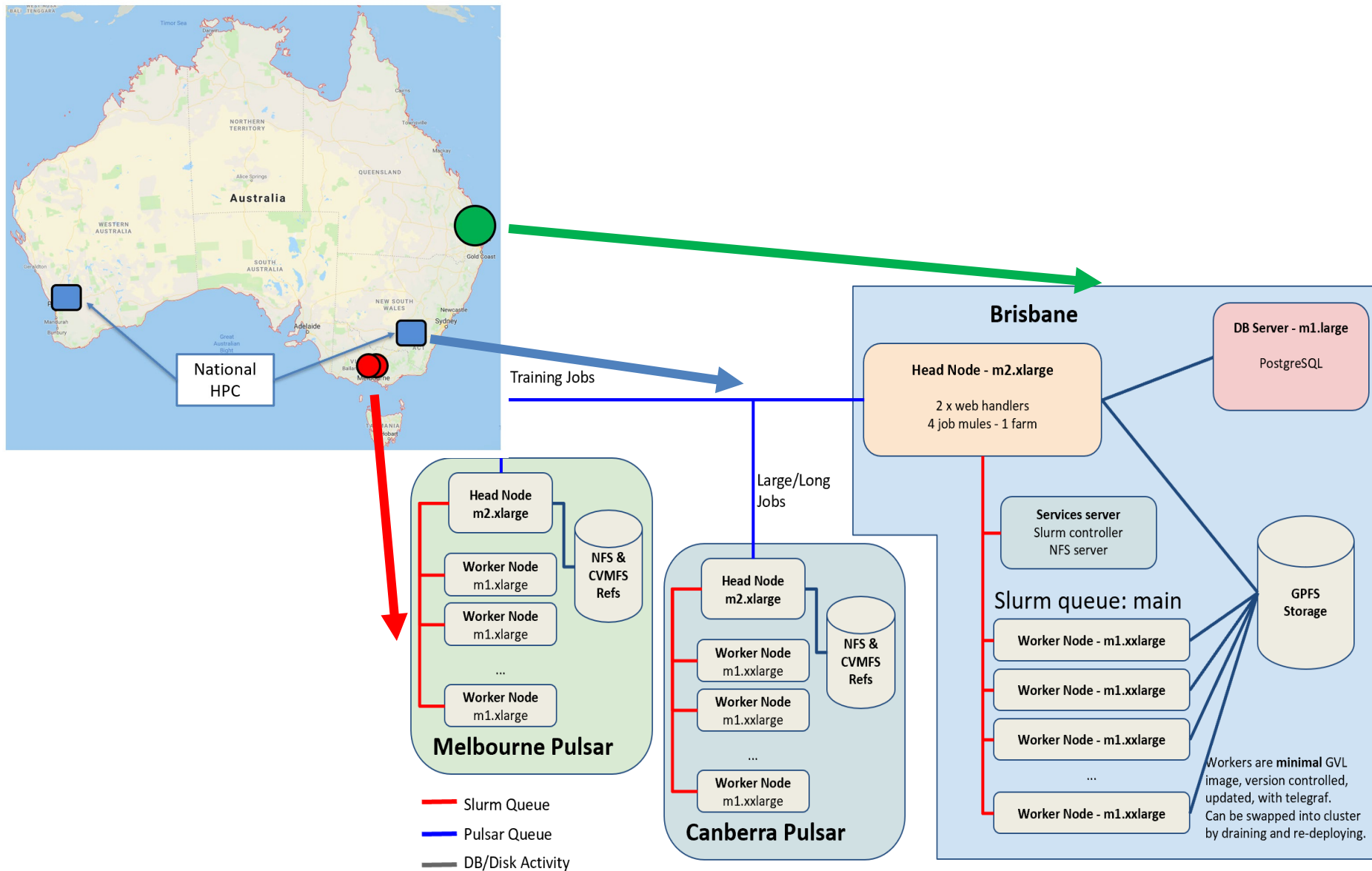
- **Qiagen**

- CLC-Bio Suite of Analysis Products

- **ThermoFisher**

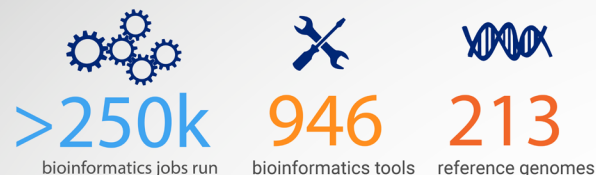
- Ion Reporter

Galaxy Australia – distributed architecture



Community Impact Summary

A complete bioinformatics analysis platform with:



... and on average 2 new tools or references added weekly on user request

Supporting published research across many areas:



57
peer reviewed papers

14
Higher degree theses

used Galaxy Australia to support their bioinformatics analysis

An active and engaged user community

5866

registered users.

1537 active users (last 90 days)

User growth 2016 - 2019



Registered users in Australia from:



30

Australian Universities



22

Medical Research Institutes or Organisations



19

Other Research Organisations

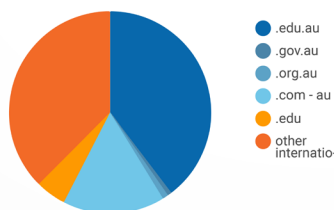
4

teaching courses utilising Galaxy Australia annually

1765

people trained in hands-on workshops

Users per domain



Users represented across

420

organisations

Users represented across

49

countries

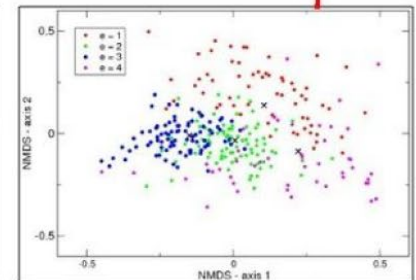
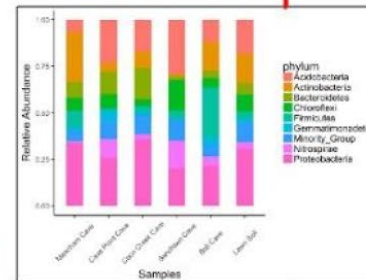
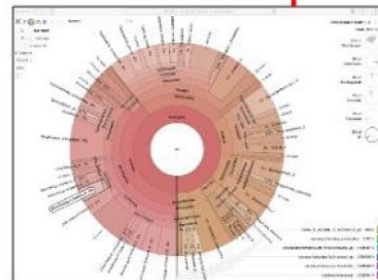


Connecting Australian Data Repositories with Galaxy Australia

Interactive plots/graphs
Presentation of input data and tools/versions used to user

Non consortium members

Australian
Microbiome
consortium



Plot on Kronos graph

Plot on graph

Plot on graph

Invoke
Galaxy Australia
API

Calculation of taxonomic diversity

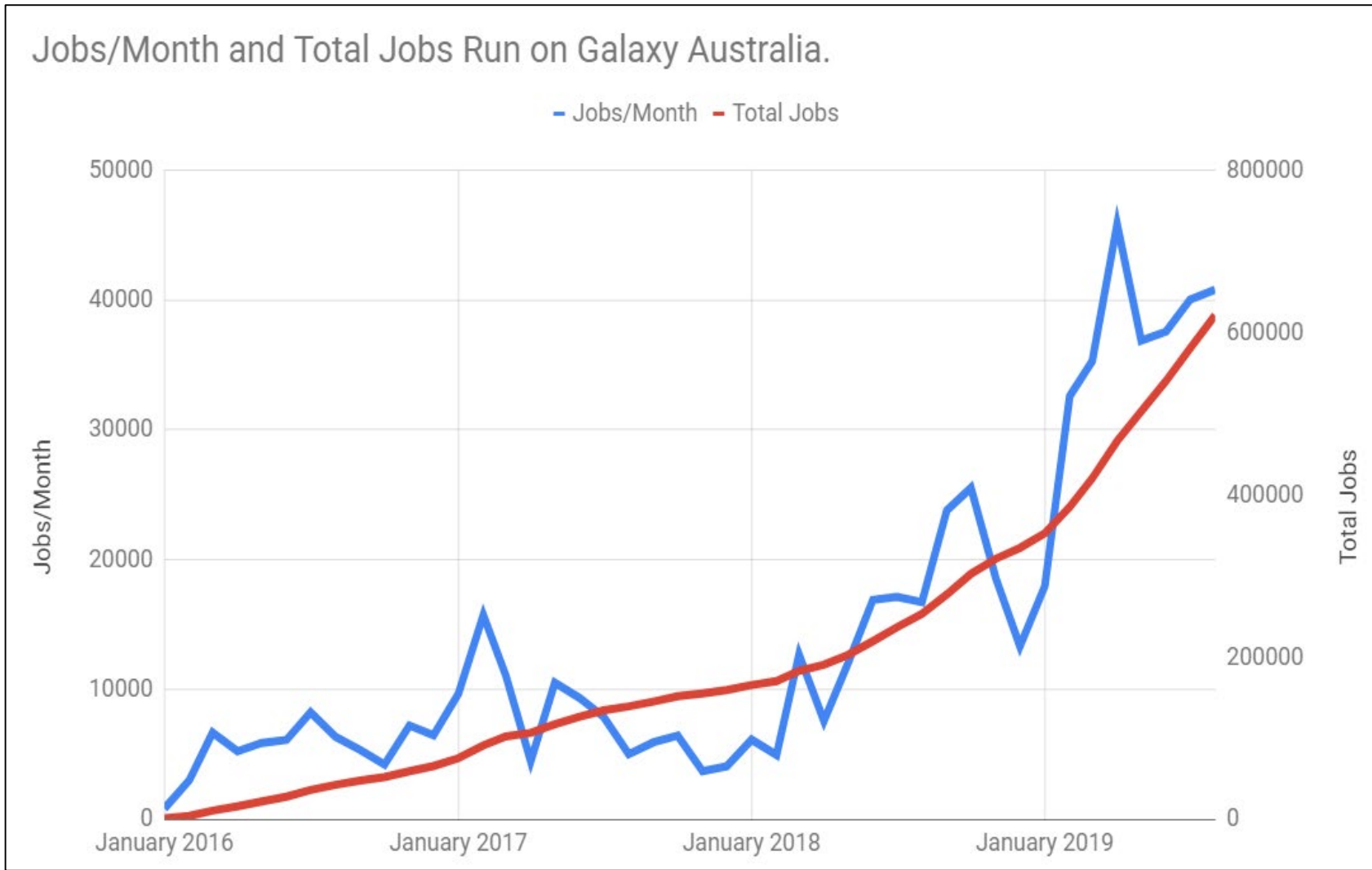
Calculation of
relative OTU
abundance at
each site

NMDS ordination

BYO and upload OTU
abundance tables



Community Impact



usegalaxy.* - a global platform and support network

Distributed reference data between .* servers

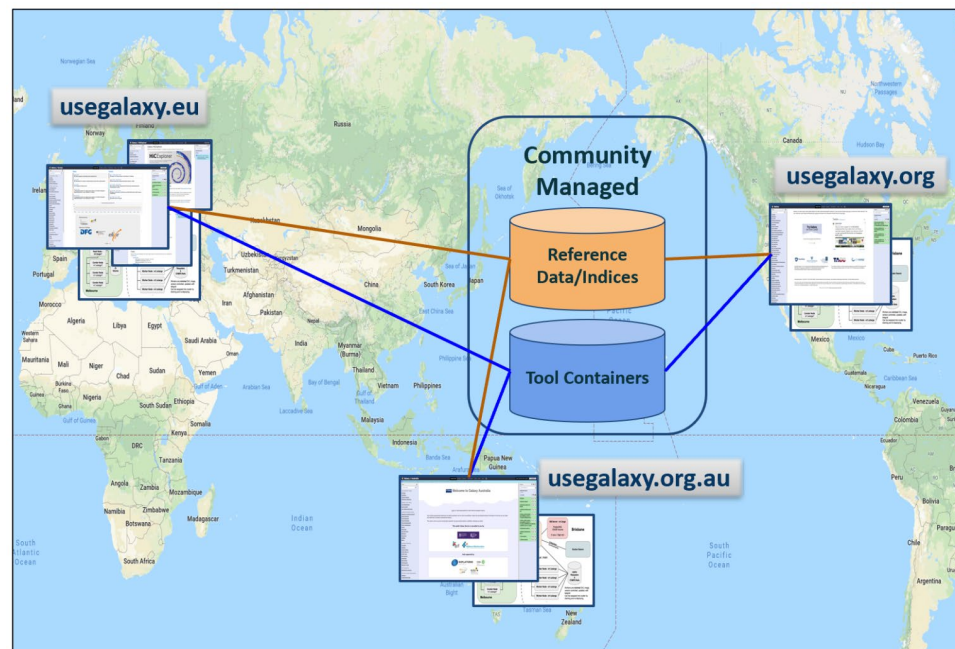
- reduced System Administration per locale
- Australian contribution to global efforts
- users are not restricted to “local” content

Intergalactic Data Commission

- formed in 2018
- regulation, automation and documentation of the CVMFS reference collections
- Australian representation on the IDC

Galaxy Project Executive Steering Committee

- formed in 2019
- Australian representation on the Committee



usegalaxy.* - a global platform and support network



Galaxy Training!

[Contributors](#) [View on GitHub](#)

Galaxy Training Network

- Australian content contributions
- Multi-language options
- Simple and comprehensive options
- Many peer reviewed best practices
- Synergised tool set

More usegalaxy.* services

- Growing number of countries and regions are forming their own usegalaxy.* service
- Increasing content development
- Distributing SysAdmin activity
- Thursday's keynote address:

The Development of ASEAN Federated Identity and Login Management, and Galaxy ASEAN Community

Welcome to Galaxy Training!

Collection of tutorials developed and maintained by the worldwide Galaxy community

Galaxy for Scientists

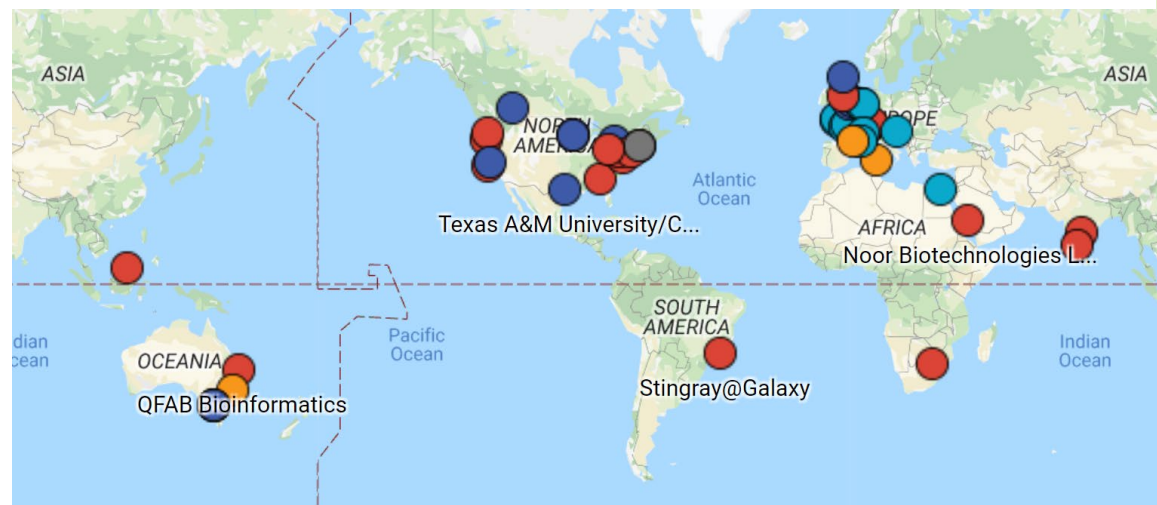
Topic	Tutorials
Introduction to Galaxy Analyses	7
Assembly	4
Computational chemistry	4
Ecology	5
Epigenetics	5
Genome Annotation	4

Galaxy Tips & Tricks

Topic	Tutorials
Data Manipulation	7
User Interface and Features	4

Galaxy for Developers and Admins

Topic	Tutorials
Galaxy Server administration	31



Growing our Australian Community – Metabolomics Australia

- Provide vendor agnostic tools and visualisations
- Leverage an increasingly mature global open source community

Galaxy / Metabolomics Analyze Data Workflow Visualize Shared Data Help Login or Register

Tools
search tools

Get Data

Collection Operations

GENERAL TEXT TOOLS

Text Manipulation

Filter and Sort

Join, Subtract and Group

Convert Formats

Metabolomics

genform Generation of molecular formulas by high-resolution MS and MS/MS data

W4m Data Subset Filter W4m data by values or metadata

NMR_Annotation Annotation of complex mixture NMR spectra and metabolite proportion estimation

xcms get a sampleMetadata file which need to be filled with extra information

Welcome to UseGalaxy.EU Metabolomics

Metabolomics.UseGalaxy.eu is a community driven metabolomics Galaxy service. The main aim is to give people a public space to discover and run metabolomics tools. If there is something that should be added, please let us know on [Gitter](#).

During the [2019 Galaxy Community Conference](#) (GCC2019) metabolomics users and developers are invited to meet during a Birds of a Feather (BoF) event to plan the aims of this service.

Content

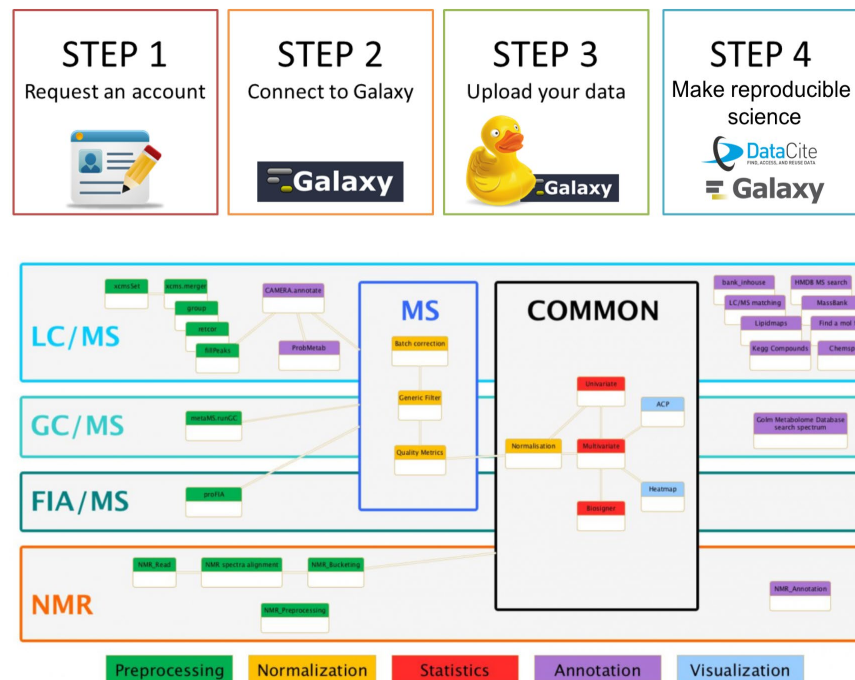
1. [Get started](#)
2. [Training](#)
3. [Available Tools](#)
 1. [Workflow4Metabolomics Tools](#)
 2. [Workflow4Metabolomics Related Tools](#)
4. [Gateways](#)
5. [News and Events](#)
6. [Contributors](#)

Get started

Are you new to Galaxy, or returning after a long time, and looking for help to get started? Take a [guided tour](#) through Galaxy's user interface.

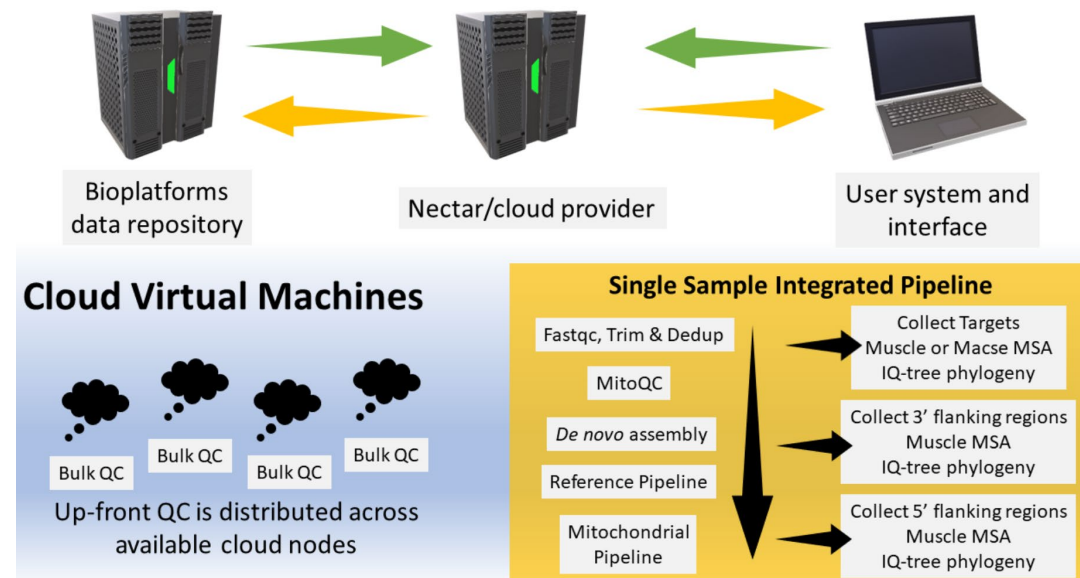
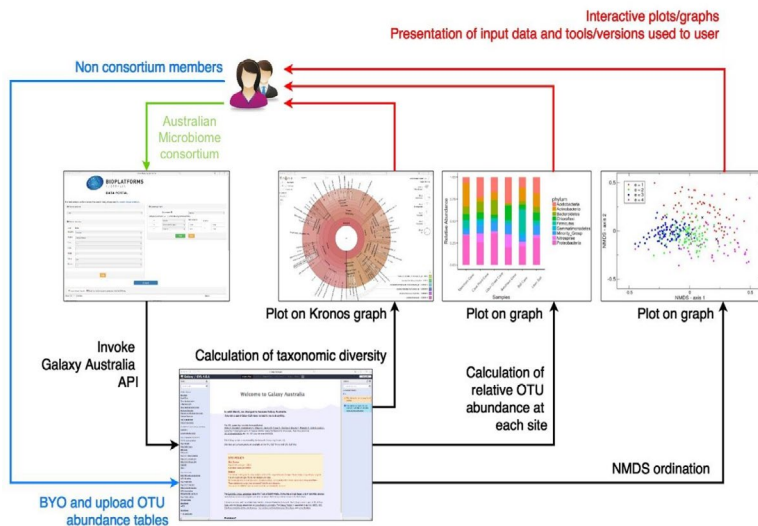
Training

Want to learn more about metabolomics? Check out the following lesson tutorials from the [Galaxy Trainings Network](#):



Growing our Australian Community – Oz Mammal Genomes and Target Capture Panel pipelines

- Build on 2018 experience with BPA Data Portal Australian Microbiome
- User interact at BPA Data Portal (and potentially never leave!) to trigger pipeline on Galaxy Australia
- Analysed data returned to Data Portal and made publicly visible



Growing our Australian Community – authentication and storage links

- Galaxy Project is an open collaborative global project
- Users on public Galaxy services rarely are required to authenticate
- But with authentication comes opportunity for tailored or seamless servicing
- Link Galaxy to AAF
 - Link to institutional HPC destinations
- Link Galaxy to AARNET Cloudstor
 - Data Input
 - Data Export
 - Service provider mediated Input



Galaxy Australia Team Members

Gareth Price

Nick Rhodes

Igor Makunin

Simon Gladman

Thom Cuddihy

Nuwan Goonasekera

Special Thanks:

Sarah Richmond, Ecoscience Research Cloud (ecocloud)

Derek Benson, CSIRO

Anna Syme, Royal Botanical Gardens, AU

Grahame Bowland, QCIF