An Australian open-source tool and workflow platform
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
Taking the IT out of bioinformatics

- 946 tools supporting analysis applying more than 200 reference datasets
- simple data upload with optional rule based data management
- retains “histories” of analyses
- a large number of cutting edge tools
- an app store (over 7,000 tools)
- All aspect peer reviewable and transparent
  - Workflows
  - Citations / origins
  - Histories
  - Data
Galaxy Australia – distributed architecture

Melbourne Pulsar
- Head Node m2.xlarge
- Worker Node m1.xlarge
- Worker Node m1.xlarge

Canberra Pulsar
- Head Node m2.xlarge
- Worker Node m1.xlarge
- Worker Node m1.xlarge

Brisbane
- Head Node m2.xlarge
  - 2 x web handlers
  - 4 job mules - 1 farm

Workers are minimal GVL image, version controlled, updated, with telegraf. Can be swapped into cluster by draining and re-deploying.
Community Impact Summary

A complete bioinformatics analysis platform with:

- >621k bioinformatics jobs run
- 946 bioinformatics tools
- 213 reference genomes
- ... and on average 2 new tools or references added weekly on user request

Supporting published research across many areas:

- Cardiac
- Toxicology and Drug Discovery
- Cancer
- Vision
- Agriculture and land management
- Obesity
- Ecology
- Infection and immunity
- Evolution

54 publications mentioned the GVL
31 studies used the GVL in their publication

An active and engaged user community

- 4545 registered users.
- 724 active users (last 90 days)

User growth 2016 - 2019 Q2

Registered users in Australia from:

- 30 Australian Universities
- 22 Medical Research Institutes or Organisations
- 19 Other Research Organisations

Users per domain

- 1064 people trained in hands-on workshops

Users represented across:

- 420 organisations
- 49 countries
"We have sequenced well over 500 transcriptomes and genomes, and routinely use Galaxy Australia for many bioinformatics processes.

It is easy to use, has high computational power, a sophisticated support structure and enables global collaboration through straightforward data sharing.

We greatly appreciate the service."

Dr Fabio Cortesi & Prof Justin Marshall, Queensland Brain Institute
Solutions for Data Analysis

• Freeware
  • Galaxy
  • R Studio
  • Command Line / HPC

• Commercial
  – Office
    • Excel
  – Agilent
    • Cartagenia Bench Lab for Molecular Pathology
  – Illumina
    • BaseSpace
  – Qiagen
    • CLC-Bio Suite of Analysis Products
  – ThermoFisher
    • Ion Reporter

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
Growing our Australian Community

• Ongoing funding (for 4 years) with a view to establish an enduring national research infrastructure

• adding to service functionality by adding metabolomics and phylogenetics

• greater security through institutional authentication, linking this to higher resourcing for authenticated users

• providing data sharing and data movement options through AARNET Cloudstor

• anticipating needs of other additional communities – esp. single cell and genome assembly

• adding new resources: Galaxy slave servers (aka Pulsars) around the country
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
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
Kylie McKenzie     Carmel Akhavari     Stephanie Heyneke
More Galaxy at INCOB2019

- Workshop: Wednesday 11th 1:30pm – 5pm, Workshop Room 1
  - An introduction to the Galaxy platform for computational biology – with real-world hands-on demonstration

- Breakout: Thursday 12th 11:00am – 12:00pm, Ar-rahim Auditorium
  - Building a regional Galaxy Community and Platform

Please join us
Thank you and hope to see you all in Brisbane one year soon.