

Galaxy Australia – the “Bring Your Own Data” platform enabling multi-omics analysis for biology researchers

Gareth Price¹

¹Gareth Price ²Simon Gladman, ¹Igor Makunin, ¹Nicholas Rhodes, ¹Thom Cuddihy, ²Helen van de Pol, ²Nuwan Goonasekera, ³Nigel Ward, ³Jeff Christiansen, ²Andrew Lonie

¹QCIF/QFAB and University of Queensland, Brisbane, Australia, {[g.price](mailto:g.price@qfab.org), [i.makunin](mailto:i.makunin@qfab.org), [n.rhodes](mailto:n.rhodes@qfab.org), [t.cuddihy](mailto:t.cuddihy@qfab.org)}@qfab.org

²University of Melbourne, Melbourne, Australia, {[simon.gladman](mailto:simon.gladman@unimelb.edu.au), [helen.vandepol](mailto:helen.vandepol@unimelb.edu.au), [nuwan.goonasekera](mailto:nuwan.goonasekera@unimelb.edu.au), [alonie](mailto:alonie@unimelb.edu.au)}@unimelb.edu.au

³QCIF and University of Queensland, Brisbane, Australia, {[nigel.ward](mailto:nigel.ward@qcif.edu.au), [jeff.christiansen](mailto:jeff.christiansen@qcif.edu.au)}@qcif.edu.au

BACKGROUND

Operation of the [Galaxy](#) bioinformatics analysis platform as a service for Australian biology researchers was nationalised in 2018 with the formation of [Galaxy Australia](#). Within its first full year of operation all service metrics have grown significantly: i.e. numbers of users, analyses performed and peer reviewed publications citing Galaxy Australia as a tool that was used to underpin a wide range of bioinformatic analyses, see Figure 1.

In its role as an easy-to-use ‘Bring Your Own Data’ Analysis platform, that is well supported by an active world-wide community of developers, Galaxy Australia has now been recognised as a key foundational component of a national bioinformatics infrastructure (the “Australian BioCommons”). Additional 2019 funding from Bioplatforms Australia, the ARDC and AARNet, through the [Australian Bioinformatics Commons \(BioCommons\) Pathfinder](#) Project, is enabling Galaxy Australia to expand the range of communities that it supports i.e.:

- To enhance the service so that it is fit for purpose for researchers undertaking metabolomics analyses (including, but not restricted to, users of metabolomics profiling instruments managed by [Metabolomics Australia](#) (MA)). Users of MA instruments will be actively encouraged by MA to utilise international community developed best-practice metabolomics data processing pipelines installed on Galaxy Australia, or to create new pipelines for more bespoke analytical options.
 - International community best-practice metabolomics data processing pipelines (identified/developed by the operators of Metabolomics Australia infrastructure) will be installed on Galaxy Australia. Any raw data arising from the specific provider and models of metabolomics profiling instruments selected by Metabolomics Australia will be of benefit to **all users** of Galaxy Australia.
- To enhance the service so that it is fit for purpose for researchers undertaking phylogenomics and phylogenetics analyses (including, but not restricted to, members of the [Oz Mammals Genomics consortium](#), and [Genomics for Australian Plants](#) consortia) by the installation of target capture data processing pipelines (specifically Exon Target Capture) onto Galaxy Australia.

- Data processing and phylogenetics analysis pipelines will be installed on Galaxy Australia and available to **any Australian researcher** wishing to access and utilise the freely available analysis pipelines.
- A process will be defined and accompanying software enhancements deployed and documented to allow individual users of the [Bioplatforms Australia Data Portal](#) to (a) move raw data from the [Oz Mammals Genomics area of the Data Portal](#) to Galaxy Australia (for the Exon target capture data processing and phylogenetics analysis pipelines), and (b) move processed/analysed phylogenetics data from Galaxy Australia back to the [Oz Mammals Genomics area of the Data Portal](#). This will be of benefit to **all authorised users** of the OMG Genomics Consortium dataset within the Bioplatforms Data Portal and further to **all users** of Bioplatforms Data Portal if data movement functionality is applied across all hosted datasets.
- Streamline connectivity between nationally available data storage tools such as [AARNet's CloudStor](#) with Galaxy Australia, so that researchers in Australia can seamlessly transport data to/from the Galaxy Australia analysis service, from their personal long-term data storage allocation provided by [AARNet's CloudStor](#).
 - A process will be defined, and accompanying software enhancements deployed, for connecting [AARNet's CloudStor](#) to the [Galaxy Australia](#) service, and publicly available documentation produced which will allow consumers of these two services to move data to/from personal cloud data storage to the computational tools/pipelines. Thereby of benefit to **any Australian researcher** wishing to connect the two services.
- To enhance the 'BYO Data' aspect of the service, some Bioplatforms Australia sponsored data production facilities (initially this will be focussed on [Metabolomics Australia infrastructure housed at the University of Melbourne node](#)) will be enabled through enhanced connectivity to connect the data production facility to Galaxy as a Data Analysis platform using [AARNet's CloudStor](#) as the transfer medium.
 - A process will be defined, deployed and documented to allow deposition of raw data from the Metabolomics Australia University of Melbourne data production facility into [AARNet's CloudStor](#) storage for users of the facility. Initially of benefit to clients of the Metabolomics Australia (University of Melbourne) data production facility, with a view to extend across the Metabolomics Australia network.

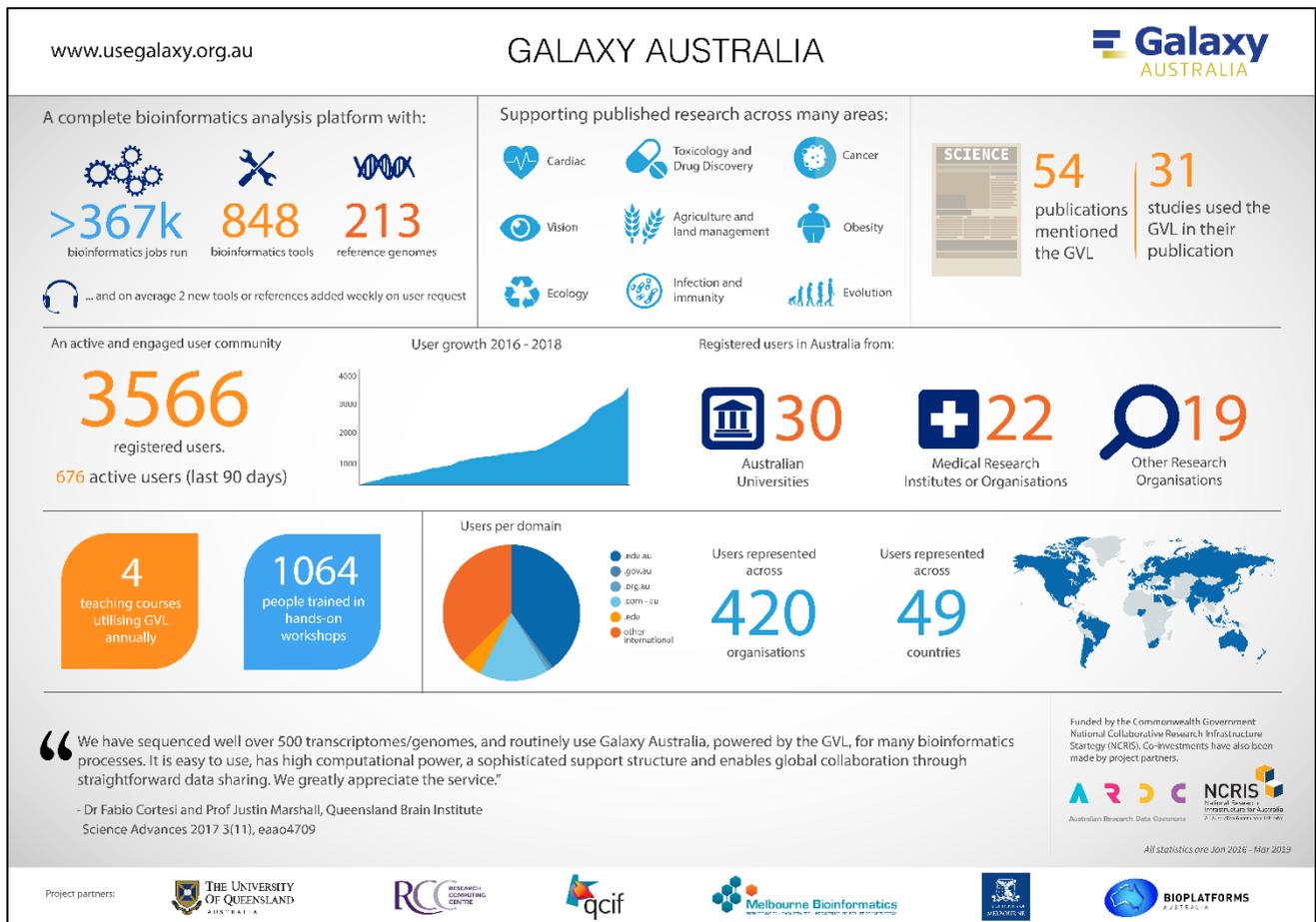


Figure 1: Galaxy Australia March 2019 Infographic