National Biosciences Computational Infrastructure and Services

Making Australia's unique soil and marine microbiome data more accessible for all through the establishment of a cloud-based data analysis system. This makes the data easier to find, analyse and interpret for bioinformaticians and non-bioinformaticians alike.

These datasets map the microbial biology and diversity of Australia's earth and marine environments. They result from several million dollars spent collecting samples, as well as further investment by Bioplatforms Australia (BPA) and others to generate metagenomic data from these samples through the Biomes of Australian Soil Environments (BASE) consortium, and the Marine Microbes (MM) consortium.

New training materials to accompany the data analysis system have been made freely available for re-use to researchers from universities, industry or government via the Galaxy Australia and EcoEd training portal.

Start date

21 February 2018

Expected completion date

30 June 2019

Investment by ARDC

$350,000

Co-investment partners

Bioplatforms Australia

CSIRO

Atlas of Living Australia (ALA)

Queensland Cyber Infrastructure Foundation (QCIF)

Centre for Comparative Genomics (CCG)

Melbourne Bioinformatics

EMBL-Australia Bioinformatics Resource (EMBL-ABR)

Lead node

BIOPLATFORMS AUSTRALIA

1. Requirements gathering

Generate an extensive list of community requirements from Australian Microbiome researchers at meetings held in Perth, Melbourne, Sydney, Brisbane and Hobart. Group, scope and prioritise the requirements. Any requirements outside the scope of this project are being investigated for future implementation.

2. Metadata

Extend the existing BPA Data Repository API to provide end users with all metadata for each dataset accessed and/or retrieved.

3. FAIR

Identify improvements to Galaxy Australia and the BPA data repository in order to make the data more Findable, Accessible, Interoperable and Reusable (FAIR). Develop policies and register API services where appropriate. This also includes the development of service descriptions and advice on methods for citation of the software and service. Make data licensing prominent within Bioplatforms-sponsored data.

4. Improved search

Develop and implement sequence similarity (e.g. BLAST) searching of metagenomic data was.

5. Training

Develop and publish online researcher training for metagenomics. Conduct and record a nation-wide hands-on practical metagenomics training session across 11 sites and make the recording available on YouTube. Develop and publish, on EcoED, introductory material about DNA sequencing approaches for analysis of microbial diversity for an Ecosciences audience .

Core features

Meaningful presentation of metagenomic data

Metagenomics data is presented to agricultural or ecoscience researchers in a way that is meaningful to them, based on requirements gathering. The data has the diversity and abundance of species in that sample identified for them from the raw sequence data, it is denoted in a geographical location, and is accessible for further exploration or analysis.

Access through existing portals

The Bioplatforms Data Portal and Galaxy Australia are linked through formal agreements to continue embedding the developed metagenomic data solution for non-informaticians in their production systems.

Who is this project for?

- Researchers with little to no coding experience who wish to use metagenomic data.
- Agricultural researchers.
- Ecoscience researchers.

What does this project enable?

Individuals and groups of users (other than skilled bioinformaticians) can now easily and directly interact with, explore, analyse and glean information from the national “Australian Microbiome” dataset.

Handy resources

Visit the Bioplatforms Data Portal

Access Galaxy Australia

Access training materials at ecoEd

Subscribe to the Australian Microbiome project for updates including data analysis tool development.

See the newly developed ‘Environmental Microbial Diversity’ module at http://ecoed.org.au/materials/

Bioplatforms Australia

VISIT

CSIRO

VISIT

Atlas of Living Australia (ALA)

VISIT

QCIF

Queensland Cyber Infrastructure Foundation (QCIF)

VISIT

Murdoch University

Centre for Comparative Genomics (CCG)

VISIT

Melbourne Bioinformatics

BIOINFORMATICS + DATA SERVICES + INFRASTRUCTURE, FOR LIFE SCIENCES TODAY

Melbourne Bioinformatics

VISIT

EMBL Australia Bioinformatics Resource

EMBL-Australia Bioinformatics Resource (EMBL-ABR)

VISIT

The Australian Research Data Commons is enabled by NCRIS.

NCRIS National Research Infrastructure for Australia An Australian Government Initiative

Stay up to date with our latest news and events and subscribe to our newsletter.

SUBSCRIBE