

Where to go when your bioinformatics outgrows your compute

Welcome!

The webinar will commence at 12pm AEST/ 11:30am ACST/ 10am AWST





Australian **BioCommons**

Actively supporting Australian life sciences research through
bioinformatics and bioscience data infrastructure

biocommons.org.au



[AustralianBioCommons](https://www.youtube.com/AustralianBioCommons)



[@AusBiocommons](https://twitter.com/AusBiocommons)

Acknowledgement of Country

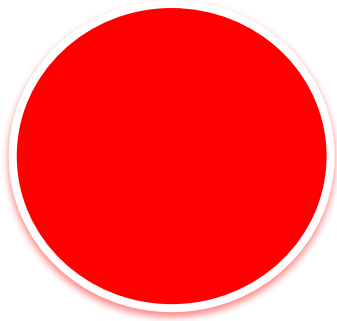
We acknowledge the Traditional Owners and their custodianship of the lands on which we meet today.

We pay our respects to their Ancestors and their descendants, who continue cultural and spiritual connections to Country.

We recognise their valuable contributions to Australian and global society.



Housekeeping



Session is recorded



Autogenerated
captions available



Questions via Q&A
function

Today's webinar

Basic needs and choosing the right infrastructure



Compute options in Australia



NCMAS for access to HPC

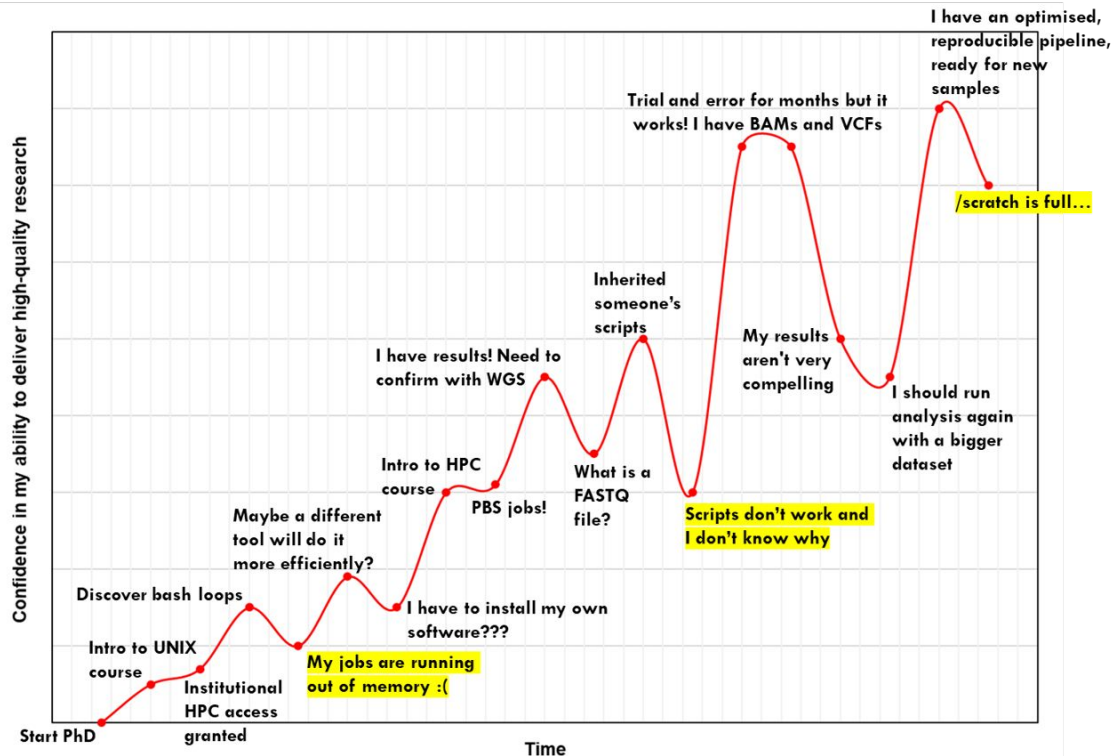


Understanding the compute requirements of your bioinformatics workflows

Dr Georgina Samaha
Sydney Informatics Hub, University of Sydney



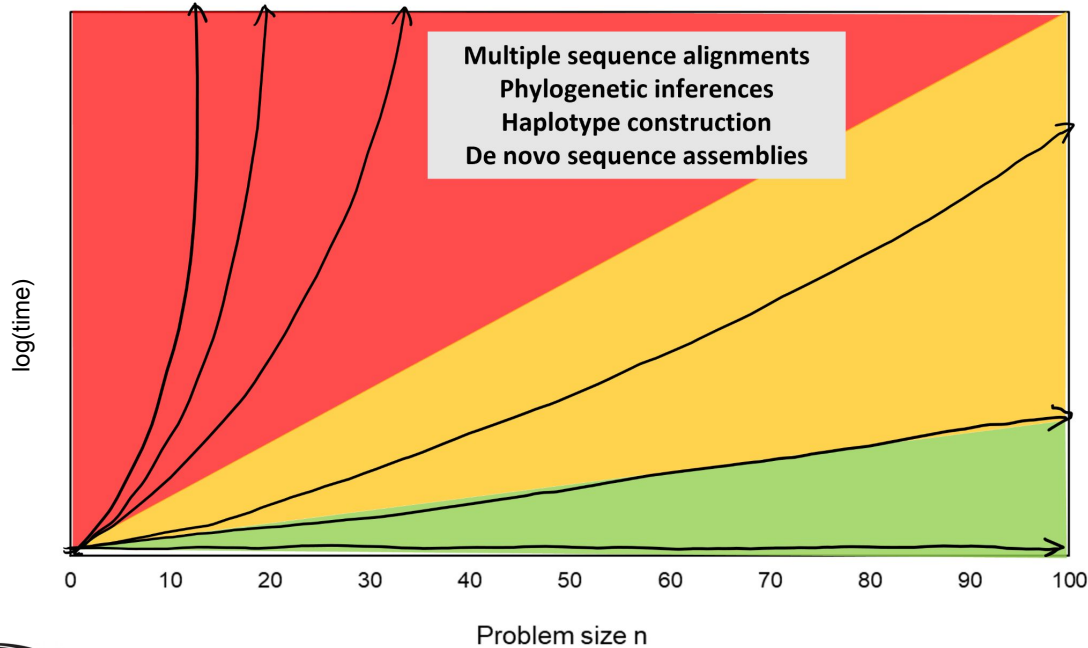
My HPC journey



Why do we struggle?

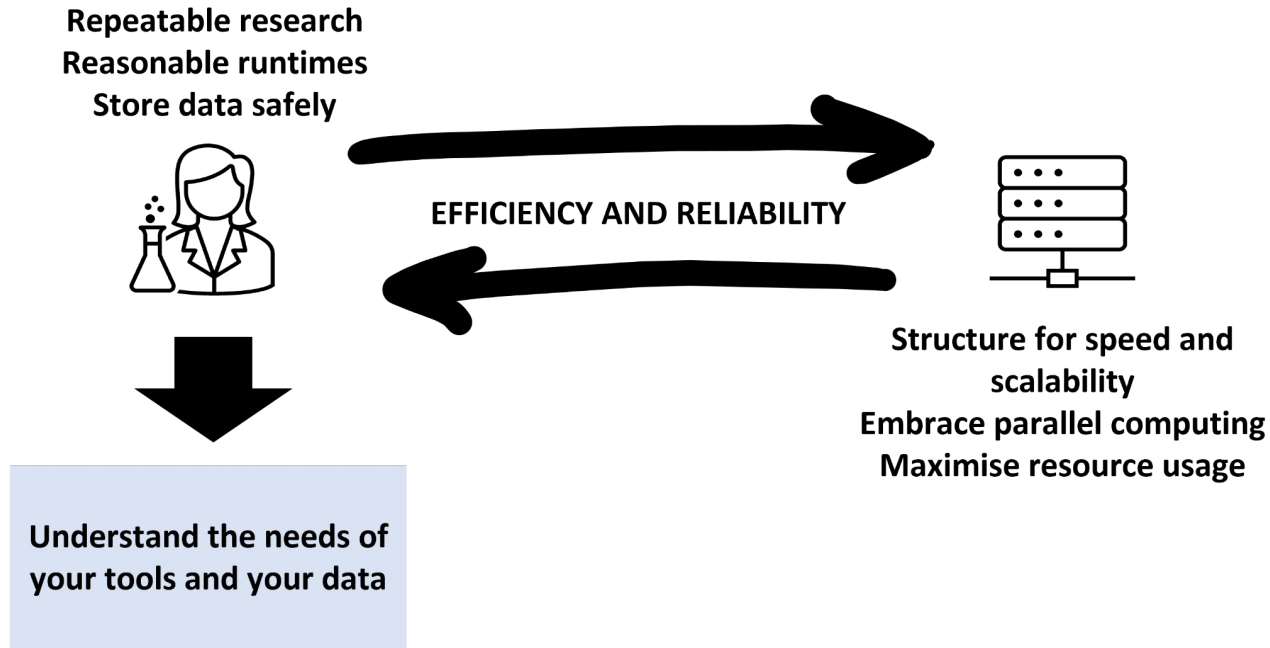
- Datasets can be massive
- Biology-fluent computational expertise is scarce
- Applications are not designed for HPCs
- Biological variation adds complexity to analyses
- Pipelines consist of multiple tools
- Applications have scale limitations

Bioinformatics is hard because biology is messy



- Bioinformatics algorithms have a complexity that limits their speed
- Biological data has a complexity that limits the algorithm's speed
- Data complexity is highly variable
- No one-size-fits-all solution, so we need to be dynamic

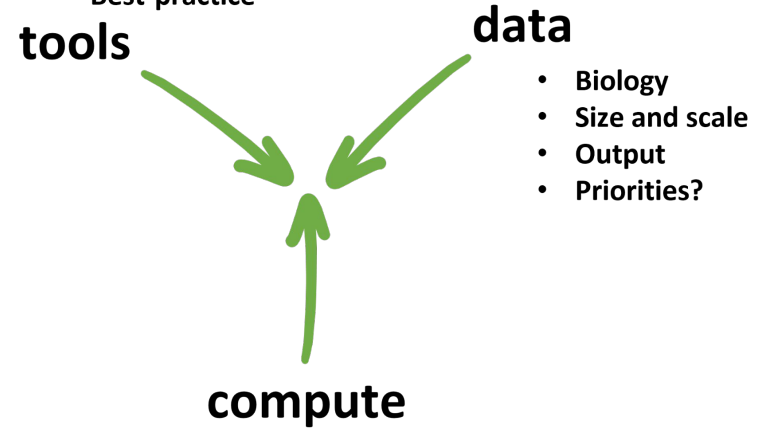
High performance computing is a two-way street



Ask questions in context

- What are my needs?
- Learn about best practice bioinformatics concepts like reproducibility and scalability
- What are the performance bottlenecks?
- What efficiencies are expected of me?
- Bigger isn't always better

- Implementation options
- Workflow logic
- Ability to scale
- Best-practice



Australian Research Data Commons

Dr Paul Coddington

ARDC - Australian Research Data Commons

People & Policy

Connecting the ARDC

Connecting the ARDC to researchers, research institutions, industry and government to enhance knowledge exchange and drive an effective national data skills ecosystem.

Communications
Engagements
Skills
Policy

Platforms & Software

Accelerating research insights and supporting collaboration

Supporting an increase in high-quality/high-impact research through eResearch platforms and better research software practices.

Research-oriented Platforms
Research Software

Data & Services

Maximising the value of Australia's data assets

Providing competitive advantage to Australian researchers by improving the discoverability, accessibility and usability of Australia's research data assets.

National Data Assets
National Information Infrastructure
National Data Capability

Storage & Compute

Providing foundation infrastructure

Supporting Australia's data and research advantage through the provision of reliable and sustainable underpinning infrastructures.

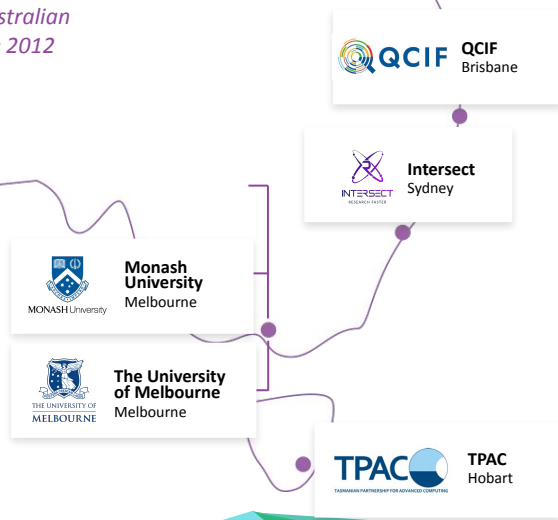
Research Compute Cloud
Data Retention

ARDC



Australia's National Research Cloud

Advancing Australian research since 2012



8,000
Virtual Machines



20,000
Physical CPU Cores



60,000+
Virtual CPU Cores



1.8PB
Object Storage



5PB
File Storage



ARDC - Nectar cloud infrastructure

- Cloud provides virtual servers on demand
 - can be used as scalable compute resources or for hosting online services
- Fills the compute gap between desktop PC and HPC
- Nectar virtual machines come in [many sizes or flavors](#)
 - From 1 virtual cpu (vcpu) and 1 GB RAM up to 32 vcpus and 128 GB RAM
 - Can request new “Huge RAM” flavors up to 48 vcpus and 360 GB RAM
 - Non-standard flavors with higher RAM are also possible
- File or object data storage, and a database service
- New services in development (available early 2022)
 - GPUs
 - Jupyter notebooks
 - Virtual desktop



THE UNIVERSITY OF
SYDNEY



Australian
BioCommons

ARDC - use of Nectar Cloud for bioinformatics

- ~20% of compute resources on the Nectar Cloud are used for projects in Biological Sciences or Medical & Health Sciences.
- 763 project allocations in Biological Sciences
- 374 project allocations in Medical & Health Sciences
- >100 NHMRC grants supported per year

Nectar Cloud has been used for:

- Galaxy Australia
- Stemformatics Virtual Laboratory
- Cancer Therapeutics CRC
- Genomics studies on threatened species (e.g. [Macquarie Perch](#))
 - “The allocation on Nectar was a huge help to our lab” - Dr Alexandra Pavlova, School of Biological Sciences, Monash University



ARDC - accessing the Nectar Cloud

- The Nectar Research Cloud can be used by any researcher with an AAF identity.
- User logs in to the [Nectar Research Cloud dashboard](#) using AAF.
- **Project Trial** (2 VCPUs for 3 months of usage) automatically allocated.
- Users can apply for a **Project Allocation** (for 3-12 months) by completing and submitting an allocation request form from the dashboard.
- Project Allocations are approved for provisioning under the [Research Cloud National Allocation Scheme](#) by the Nectar Allocation Committee
- Allocation requests usually assessed within 2 weeks.
- Users can request an update or extension of a project allocation at any time.
- No (direct) cost to the researcher – ARDC covers capex and Nodes cover opex.

ARDC - accessing the Nectar Cloud

- The Nectar Research Cloud has two categories of cloud infrastructure:
 - **National** - ARDC funded, accessible through a national allocation scheme
 - **Local** - Node funded, for Node-prioritised (local) allocations
- Projects are eligible for a **national allocation** if they meet specific criteria:
 - National competitive research grant (e.g. ARC, NHMRC)
 - Supports (or is funded by) an NCRIS capability, including ARDC projects
 - Approved by the Nectar Allocation Committee as meeting other merit criteria
- Local allocations require some arrangement with a Node

ARDC - Nectar training and user support

Nectar national helpdesk operates 7am to 6pm AEST/AEDT, accessed via:

- Email support@ehelp.edu.au
- Web interface from the [Support Portal](#) or the [Helpdesk](#)
- Live chat help from the Support Portal

The Support Portal has self-paced [online tutorials](#)

Monthly online [Getting Started with Nectar training](#) for new users.



ARDC - more information

More information is available here:

<https://ardc.edu.au/>

<https://ardc.edu.au/services/nectar-research-cloud/>

Or email support@ehelp.edu.au



Galaxy Australia

Simon Gladman

Melbourne Bioinformatics, University of Melbourne



Galaxy Australia

Galaxy Australia

Analyze Data Workflow Visualize Shared Data Admin Help User

Using 62%

Tools

search tools

Upload Data

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

Nanopore

Convert Formats

Lift-Over

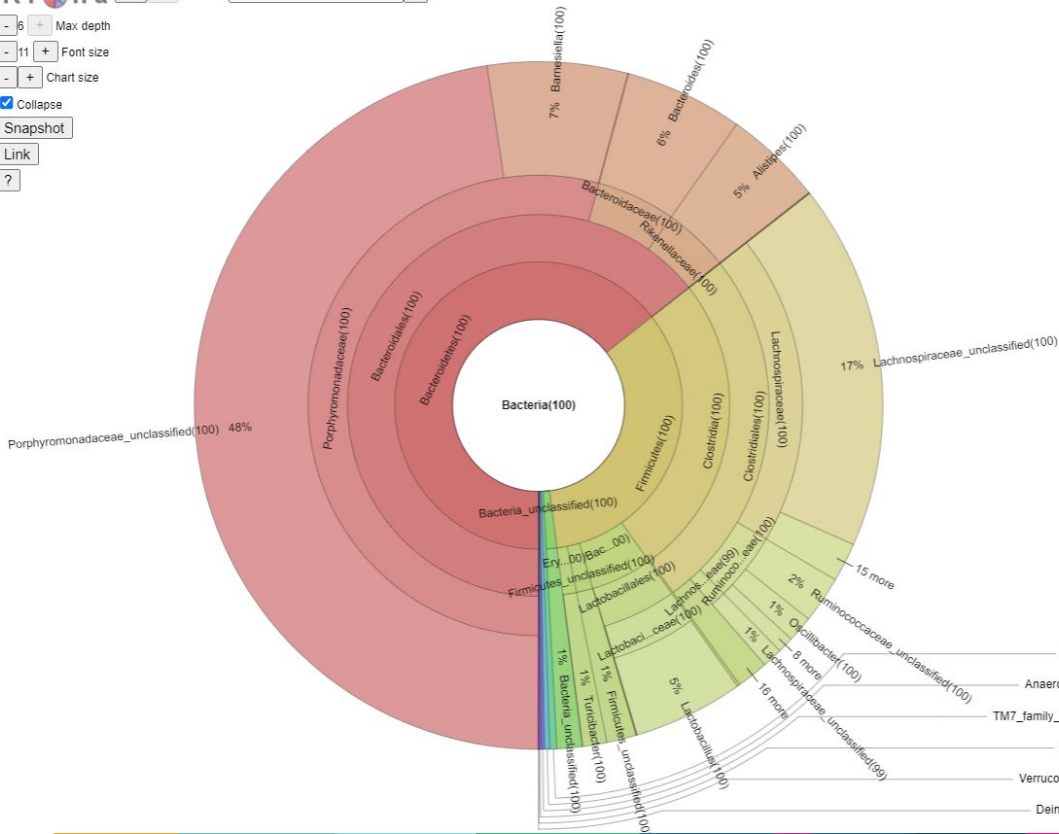
COMMON GENOMICS TOOLS

Operate on Genomic Intervals

MiModD

Krona Search: []

6 Max depth
11 Font size
Chart size
Collapse
Snapshot
Link



Bacteria(100)
Total: 113883

History

search datasets

20200811_GTN_165_Short

107 shown, 16 deleted, 59 hidden

699.94 MB

181: Convert on collection 179

a nested list with 1 item

179: Make.biom on collection 1

30 and collection 133: biom files

a nested list with 1 item

178: Krona pie chart on data 177: HTML

239.8 KB

format: **html**, database: ?

Writing /mnt/galaxy/files/002/926/dataset_292

HTML file

176: Taxonomy-to-Krona on collection 130: krona-formatted taxonomy file

a list with 1 item

169: Phylogenetic Trees

a list with 6 items

155: Heatmap.sim on collectio

Galaxy Australia

Galaxy Australia is a hosted web-accessible platform that lets you conduct accessible, reproducible, and transparent computational biological research.

- Galaxy Project - Global open source project
- Used all over the world
- Over 15 years of constant global operation and community-based contributions
- Cited in over 10,000 publications using Galaxy to enable research
- Backed by 8,000+ analytical tools and growing
- Training supported through Galaxy Training Network with over 190 training topics
- Has ongoing funding within Australia

Galaxy Australia - accessing tools and references

Currently has 1,400+ tools installed covering genomics, proteomics and metabolomics, statistics and data visualisations

100s of reference data sets and tool indices

100s of global, peer-reviewed workflows - simple one click use

And if we don't have it - just ask us

Request for installation of software tool or reference dataset on Galaxy Australia

This form has been developed to allow life science researchers across Australia to request software or reference data to be installed or updated on the national NCRIS-funded [Galaxy Australia](#) service

The information you provide will allow us to identify the applicability of the requested tool/data across the life science community as a whole and to help us in identifying resources that will be able to update/install the software/data for you.

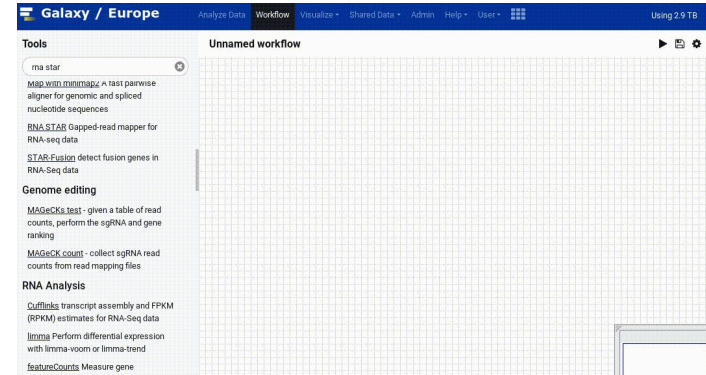
Submissions are reviewed weekly.

Please note that due to resourcing constraints, submitting a request for software or reference dataset installation using this form doesn't necessarily guarantee that assistance can be provided.

Name	Annotation	Owner	Community Rating	Community Tags	Last Updated
RNA Seq Counts To Genes	RNA-seq counts to genes	galaxy-australia	★★★★★	transcriptomics galaxy-training-network	Mar 19, 2021
CelSeq2: Multi Batch (mm10)	Pre-processing of Single-Cell RNA Data	galaxy-australia	★★★★★	transcriptomics galaxy-training-network	Mar 19, 2021
QC + Mapping + Counting - Ref Based RNA Seq - Transcriptomics - GTN	Reference-based RNA-Seq data analysis	galaxy-australia	★★★★★	transcriptomics galaxy-training-network	Mar 19, 2021
DEG Part - Ref Based RNA Seq - Transcriptomics - GTN	Reference-based RNA-Seq data analysis	galaxy-australia	★★★★★	transcriptomics galaxy-training-network	Mar 19, 2021
RNA Seq Genes To Pathways	RNA-seq genes to pathways	galaxy-australia	★★★★★	transcriptomics galaxy-training-network	Mar 19, 2021
RNA Seq Reads To Counts	RNA-Seq reads to counts	galaxy-australia	★★★★★	transcriptomics galaxy-training-network	Mar 19, 2021
De novo transcriptome reconstruction with RNA-Seq	De novo transcriptome reconstruction with RNA-Seq	galaxy-australia	★★★★★	transcriptomics galaxy-training-network	Mar 19, 2021
Visualization Of RNA-Seq Results With Volcano Plot	Visualization of RNA-Seq results with Volcano Plot	galaxy-australia	★★★★★	transcriptomics galaxy-training-network	Mar 19, 2021

Galaxy Australia - part of a global community of coders, developers, trainers and users

Using Galaxy means you can draw on the experiences of 10,000s of users worldwide. Galaxy offers - common use tool recommendations: inside Workflows and each tool execution



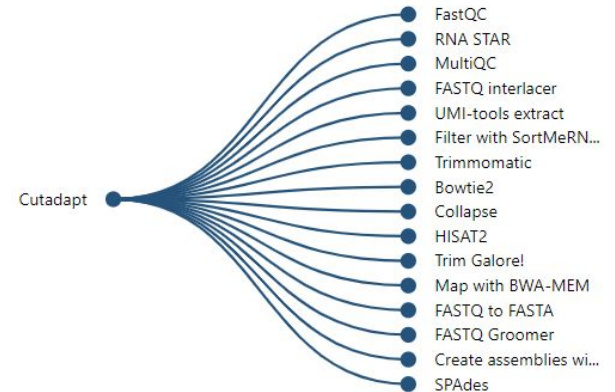
Galaxy Training Network -

<https://training.galaxyproject.org/>

Gitter - <https://gitter.im/galaxyproject/Lobby>

Galaxy Project news -

<https://galaxyproject.org/blog/>



Galaxy Australia - compute leading to publication

Characterisation of *Listeria monocytogenes* food-associated isolates to assess environmental fitness and virulence potential

Jessica A. Gray^{a,b,c,*}, P. Scott Chandry^b, Mandeep Kaur^d, Chawalit Kocharunchitt^c, John P. Bowman^c, Edward M. Fox^{b,e}

^a CSIRO Agriculture and Food, Coopers Plains
^b CSIRO Agriculture and Food, Werrisbee, V.
^c Food Safety Centre, Tasmanian Institute of Food and Nutrition
^d Biosciences and Food Technology, School of Food Science and Technology
^e Department of Applied Sciences, Northumbria University

Whole-genome reference of *Dirofilaria immitis* from Australia and identification of single nucleotide polymorphisms associated with macrocytopenia and resistance in the USA

Daisy Ching-Wai Lau, Stephanie McLeod, Sara Collaery, Selina Peou, Andy Truc Tran, Michelle Liang, Jan Šlapeta^{*}

Sydney School of Veterinary Science, Faculty of Science, The University of Sydney, NSW, 2006, Australia

NextGen molecular barcoding of larval grouper diet in an extensive green-water pond system

Lachlan Peter Dennis^a, Kelli Anderson^b, Matthew Wylie^c, Vu Van In^d, Josephine Nocillado^a, Abigail Elizur^{a,*}

^a Geneology Research Centre, University of the Sunshine Coast, Locked Bag 4, Maroochydore DC 4558, Queensland, Australia
^b Institute for Marine and Antarctic Studies, Fisheries and Aquaculture, University of Tasmania Newnham Campus, Private Bag 1370, Newnham, Tasmania 7248, Australia
^c The New Zealand Institute for Plant and Food Research Limited, Seafood Production Unit, 293-297 Port Nelson, Nelson 7010, New Zealand
^d Research Institute for Aquaculture No. 1, Northern National Broodstock Center for Mariculture (CATBANBC), Cat Ba, Viet Nam



The role of antioxidants in restoring MAPK 14 and a DNA damage marker level following autophagy suppression

Abdalla Elbially

Faculty of Veterinary Medicine, Damanhour University, Damanhour 22511, Egypt

AE, 0000-0002-6344-9425

Galaxy Australia - access

- Register using your academic email (where possible), receive the largest analysis quota



- <https://usegalaxy.org.au/>
- help@genome.edu.au
- <https://twitter.com/galaxyaustralia>



NCI

Dr Roger Edberg
Javed Shaikh



NCI
AUSTRALIA

High-performance computing, data, storage

Resource schemes

- Merit: NCMAS, ANUMAS
- Stakeholder: universities, .gov.au, institutes
- NCI Startup: development
- Industry: service agreements

<https://nci.org.au>

help@nci.org.au



NCI

NCI Systems

- Gadi: tier-1 compute system
- Lustre file systems: /scratch, /g/data
- Massdata tape archive
- Data collections
- *Cloud*
- *Open OnDemand, will supersede Virtual Desktop Interface (VDI)*

Rank	System	Cores	Rmax (TFlop/s)	Rpeak (TFlop/s)	Power (kW)
44	Gadi - PRIMERGY CX2570 M5, ThinkSystem SD650, Xeon Platinum 8274/8268, NVIDIA Tesla V100 SXM2, Mellanox HDR Infiniband, Fujitsu / Lenovo National Computational Infrastructure (NCI Australia) Australia	204,032	9,264.3	15,142.2	

Gadi - Bioinformatics

- Hugemem (1.5 TB), megamem (3 TB), gpuvolta nodes
- /g/data and /scratch file systems

NCI

Stakeholders (resource schemes)

Government: Bureau of Meteorology, CSIRO, Geoscience Australia, Antarctic Division, NSW, QLD, VIC

Universities: Adelaide, ANU, Deakin, La Trobe, Melbourne, Macquarie, Monash, RMIT, Sydney, Tasmania, UNSW, UQ, Victoria, WSU, Wollongong

Consortia: BioCommons, Intersect, QCIF

Institutes: Garvan, Victor Chang, Autism CRC, CCIA



NCI

Services

- Help Desk: help@nci.org.au
- User registration: <https://my.nci.org.au>
- Start up projects: <https://my.nci.org.au> (Gadi)
- Consultation: User services group
- Application build and install: User Services, HPC
- Workflow optimisation: User Services, HPC
- Data collections
- Cloud: coming...
- Open OnDemand: coming...
- Training: new training group, led by Dr Jingbo Wang

NCI

Any relevant activities, training, communities of practices

NCI

Resource schemes

Merit: NCMAS, ANUMAS

Stakeholder: universities, .gov.au, institutes

NCI Startup: development

Industry: service agreements



NCI

Public web site <https://nci.org.au>

User and project registration <https://my.nci.org.au>

NCMAS <https://ncmas.nci.org.au>

Pawsey Supercomputing Centre

Dr Maciej Cytowski



Pawsey Supercomputing Research Centre

Introduction

Australian Tier-1 Research Facility

- Headquartered in Perth, Western Australia on Whadjuk country
- Launched as Pawsey in 2014 with foundations back to 2000
- Critical support for SKA infrastructure on Wajarri Yamatii country, 800km north of Perth
- AU\$70m capital refresh by Australian Government

Pawsey People

- 50+ staff employed via CSIRO in UJV
- Supercomputing Applications Specialist with research background in various fields including bioinformatics, quantum chemistry, CFD, astrophysics, computational science and more...

Pawsey Researchers

- Supporting 200+ projects and 3000+ researchers across all disciplines



Pawsey Supercomputing Research Centre

Project Examples

- Trace and Environmental DNA (TrEnD) Laboratory (Curtin University)
 - Environmental DNA (eDNA) metabarcoding reveals biodiversity
 - Nextflow workflow led to joint publication of eDNA pipeline
- Computational Biology Lab (Telethon Kids Institute)
 - Leveraging public datasets to discover biomarkers in childhood cancers
 - Rare genetic diseases

Received: 27 August 2020 | Revised: 29 January 2021 | Accepted: 8 February 2021

DOI: 10.1111/1755-0998.13356

RESOURCE ARTICLE

MOLECULAR ECOLOGY
RESOURCES WILEY

eDNAFlow, an automated, reproducible and scalable workflow for analysis of environmental DNA sequences exploiting Nextflow and Singularity

Mahsa Mousavi-Derazmahalleh^{1,2,3}  | Audrey Stott⁴ | Rose Lines^{1,2,3} | Georgia Peverley³ | Georgia Nester¹ | Tiffany Simpson^{1,3} | Michal Zawierta⁵ | Marco De La Pierre⁴ | Michael Bunce^{1,6}  | Claus T. Christensen^{1,2,7}



Pawsey Supercomputing Research Centre

Supercomputing Services

- Current systems: Magnus, Zeus, Topaz
- Future system: Setonix (next slide)
- Large Scale computing
- Access: Merit Allocation Calls

Nimbus Research Cloud Services

- OpenStack based cloud environment
- Various flavours of Virtual Machines (VMs), CPU and GPU based
- Access: Pawsey Application portal (apply.pawsey.org.au)

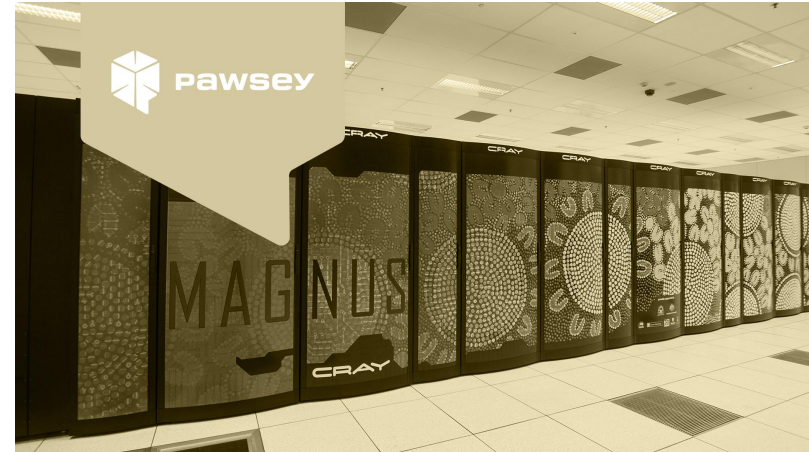
Managed Data Services

- Storage services for data collections
- Data Portal
- Access: Pawsey Application portal (apply.pawsey.org.au)

Visualisation Services

- Remote visualisation services
- High-end Linux-based and Windows based servers
- Access: Available for Pawsey researchers

Expert advice, Solution Design and Consulting
Support for Large Scale Computing
Support in Scaling Up Research
Training and Education
Help Desk
Documentation



Pawsey Supercomputing Research Centre

Setonix



Australia's new 50 petaflop system

- 200,000+ AMD Milan CPU cores
- 750+ AMD MI-Next GPUs
- 548+ TB system memory
- 15 PB ClusterStor Lustre filesystem with 2.7PB SSD



<https://support.pawsey.org.au/documentation/display/US/Changes+in+Supercomputing+Services+for+2022>



Pawsey Supercomputing Research Centre

Dedicated support for Bioinformatics

- Investing a lot in making sure we can support bioinformatics and provide necessary tools including workflow tools and container environments
- Leveraging all Pawsey services to support complex workflows (e.g. Fgenesh, Cromwell, RStudio)
- Pawsey is part of BioCommons and provides resources to BioCommons researchers

Events

- Bioinformatics Symposium - 1 day conference and workshop (2018)
- Bioinformatics at Scale and Australia's Next Generation of Supercomputers, online event (2021)

Trainings

- Implementing Scalable Bioinformatic Workflows in Snakemake & Nextflow, hybrid event organised with Australian BioCommons and EMBL-ABR (2019) 150+ participants
- Using Containers in Bioinformatics, online webinar & workshop series (2020), hundreds of participants (national and international)
- Introduction to Containers (2020), online training (80+ participants)

Hackathons

- Nextflow Hackathon (2019), 1 week, 40+ participants

Other activities

- Ask Me Anything sessions dedicated to Bioinformatics
- Bioinformatics HPC Community of Practice



Pawsey Supercomputing Research Centre



pawsey.org.au

Pawsey Friends mailing list

Pawsey Twitter feed (@PawseyCentre)

Pawsey YouTube Channel (<https://www.youtube.com/pawseysupercomputingcentre>)

User Support Portal (support.pawsey.org.au)

Pawsey Training Portal (<https://pawseyc.github.io/training.html>)

Pawsey YouTube [Home](#), [Videos](#) & [Playlists](#)

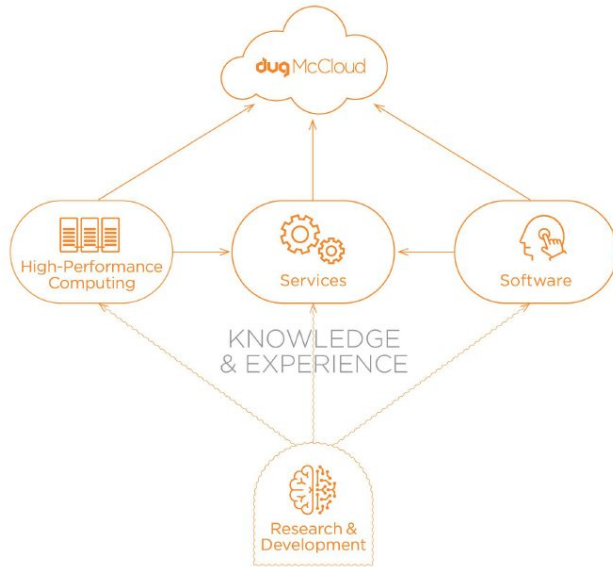


Commercial Services

Dr Rosemarie Sadsad
Sydney Informatics Hub, University of Sydney



DUG



- Australian technology company
- High Performance Computing Experts
- Big compute, big data and big storage
- **Full support** to make your research a success
 - Onboarding
 - Installation
 - Optimisation
 - Testing
- Own and operate some of the **greenest** data centres in the world
- Keep your data and compute in Australia
- Users of DUG: Harry Perkins Institute of Medical Research, CSIRO

Bioinformatics at DUG



dug



THE UNIVERSITY OF
SYDNEY



Australian
BioCommons

- **Bioinformatics pipelines** include Trinity, GATK4 FASTQ to GVCF, joint GT, 3D DNA de novo assembly, HiC
- **Software** includes Sambamba, Samtools, Bowtie2, BWA, GATK, Trinity
- **Package Managers/Repos** include Bioconda, Biocontainer, Github
- **Workflow managers include** Cromwell, Nextflow, Snakemake

Comparison: Trinity on Marsupial sample (20GB)

- Regional HPC: 76 hrs
- Institutional HPC: 55 hrs
- **DUG VAST Filesystem: 22 hrs**

For more info

www.dug.com

Aaron Lewis

Business Development Manager

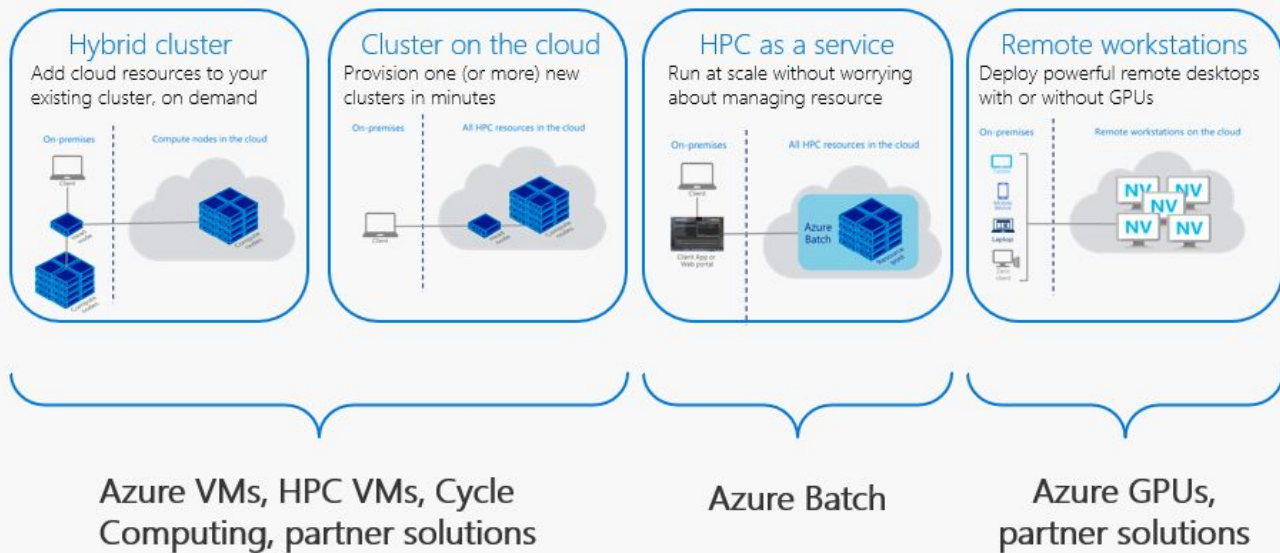
aaronle@dug.com

DUG's blog

dug.com/the-daily-dug



Azure Solutions and components

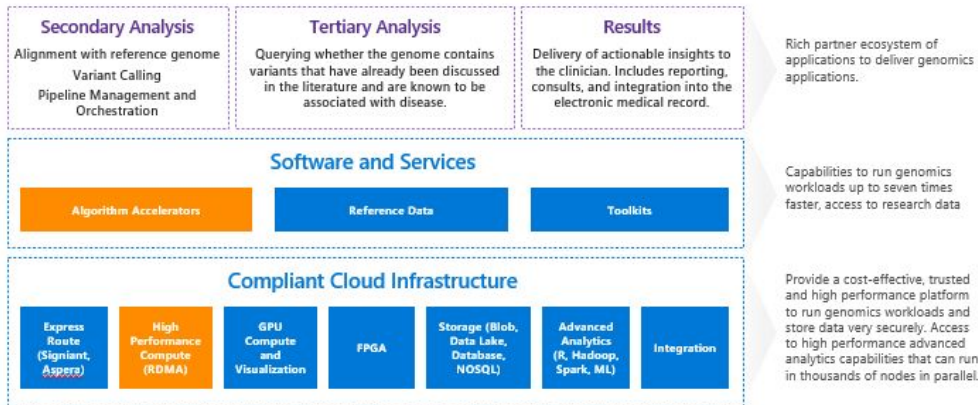


Genomics as a Service

Microsoft brings the power of software development and compute together



Provide ultra-high speed and highly secure data transfer for genomics primary analysis data, predictive maintenance, monitoring and field maintenance for equipment.



Microsoft Azure

Microsoft Genomics

SQL Server

Cortana Intelligence Suite

Dynamics365

Azure

<https://aka.ms/AzureResearch>

aus-edu@microsoft.com



Azure



THE UNIVERSITY OF
SYDNEY



Australian
BioCommons

Training and Resources for Researchers

Microsoft Azure

Get practical training to help you make the most of Microsoft Azure's intelligent, scalable, and secure environment

Introduction to Azure



[Getting Started with Azure](#)

Learn how Azure works with short video overviews, demos, and tutorials



[Overview of Azure compute services](#)

Provides a high-level overview of virtual machines and containers on Azure



[Overview of Azure machine learning](#)

Explore the full range of machine learning tools available on Azure



[Overview of Azure storage services](#)

Find the right storage options for your application on Azure



[Overview of Azure HPC](#)

Azure provides a complete set of high-performance computing solutions

Guided Trainings



[Azure Fundamentals](#)

Azure fundamentals is a six-part series that covers basic cloud concepts



[Explore Azure compute services](#)

Learn how to take advantage of several virtualization services in Azure



[Explore Azure storage services](#)

Learn about the different storage options available on Azure



[Navigating the Azure Portal](#)

Learn how to navigate and manage services using the Azure portal



[Cost Management and Optimization](#)

Plan and manage your costs in Azure with this guided series



[Introduction to high-performance computing \(HPC\) on Azure](#)

Discover the services available for your HPC research workloads



[Create machine learning models](#)

Use common frameworks to train and evaluate machine learning models



[Overview of Azure Documentation](#)

Azure documentation with steps to build and manage Azure applications

Broadest and deepest platform choice

Categories

General purpose
Burstable
Compute intensive
Memory intensive
Storage (High I/O)
Dense storage
GPU compute
Graphics intensive
HPC



Capabilities

Choice of processor
(AWS, Intel, AMD)

Fast processors
(up to 4.5 GHz)

High memory footprint
(up to 24 TiB)

Instance storage
(HDD and SSD)

Accelerated computing
(ASIC, GPUs and FPGA)

Networking
(up to 400 Gbps)

Bare Metal

Size
(Nano to 48xlarge)



Options

Elastic Block Store
Elastic Graphics
Elastic Inference



400+
instance types
for virtually
every workload
and business need





Using RONIN and AWS, we spun up training environments in minutes outside of the institutional firewall, so more than 150 bioinformaticians got access to the same virtual resources, regardless of their location."

Dr. Nathan Watson-Haigh
Research Fellow in Bioinformatics, University of Adelaide, School of Biological Sciences

With great power comes great visibility

- You control everything – your research, your way.
- All the power of AWS with none of the late nights learning Amazon EC2, networking, SysOps, DevOps, etc.
- Agility to fail fast and recover faster
- Ability to share and collaborate with a few simple clicks
- Easily manage your spend with our live budget tools for **guardrails** and **governance**

© 2020, Amazon Web Services, Inc. or its Affiliates.

RONIN PROJECTS CREATE PROJECT

PROJECT DASHBOARD

MARS PROJECT
MARS

\$5 SPENT **\$847 FORECASTED** **\$95 REMAINING** **\$100 BUDGET**

DESCRIPTION
For several decades, scientists across the globe has dedicated countless years in pursuit of finding life in Mars, even going as far as planning for humans to migrate to the red planet. And their relentless pursuit may just bear fruit soon, since NASA has already released a detailed plan of how they are going to send humans to Mars in the coming decades. The plan involves sending humans to Mars and have them permanently reside in the planet. According to NASA, "Unlike Apollo, we will be going to stay".

1 RUNNING MACHINES **3 STOPPED MACHINES** **\$1.04^{per hour} COST RUNNING MACHINES**

TIMELINE
167 DAYS REMAINING

TAGS
 Q DECADES Q SCIENTISTS Q GLOBE
 Q DEDICATED Q COUNTLESS Q PURSUIT
 Q MARS Q NASA Q MIGRATE Q RELENTLESS
 Q DETAILED Q INVOLVES

97 GB SSD STORAGE **0 GB HOT HDD STORAGE** **0 GB COLD HDD STORAGE** **0 GB MAGNETIC STORAGE**

\$11.64 MONTH \$0.00 MONTH \$0.00 MONTH \$0.00 MONTH



THE UNIVERSITY OF SYDNEY



Australian BioCommons

Genomics on AWS



Public data sets

AWS hosts a variety of public datasets that anyone can access for free. Below are just a few examples

- 1000 Genomes Project
- The Cancer Genome Atlas
- International Cancer Genome Consortium
- 3000 Rice Genome
- Genome in a Bottle (GIAB)
- The Genome Modeling System
- Medicare Drug Spending
- The Human Connectome Project
- The Human Microbiome Project
- OpenNeuro
- Physionet
- Tabula muris
- OpenStreetMaps
- and more....



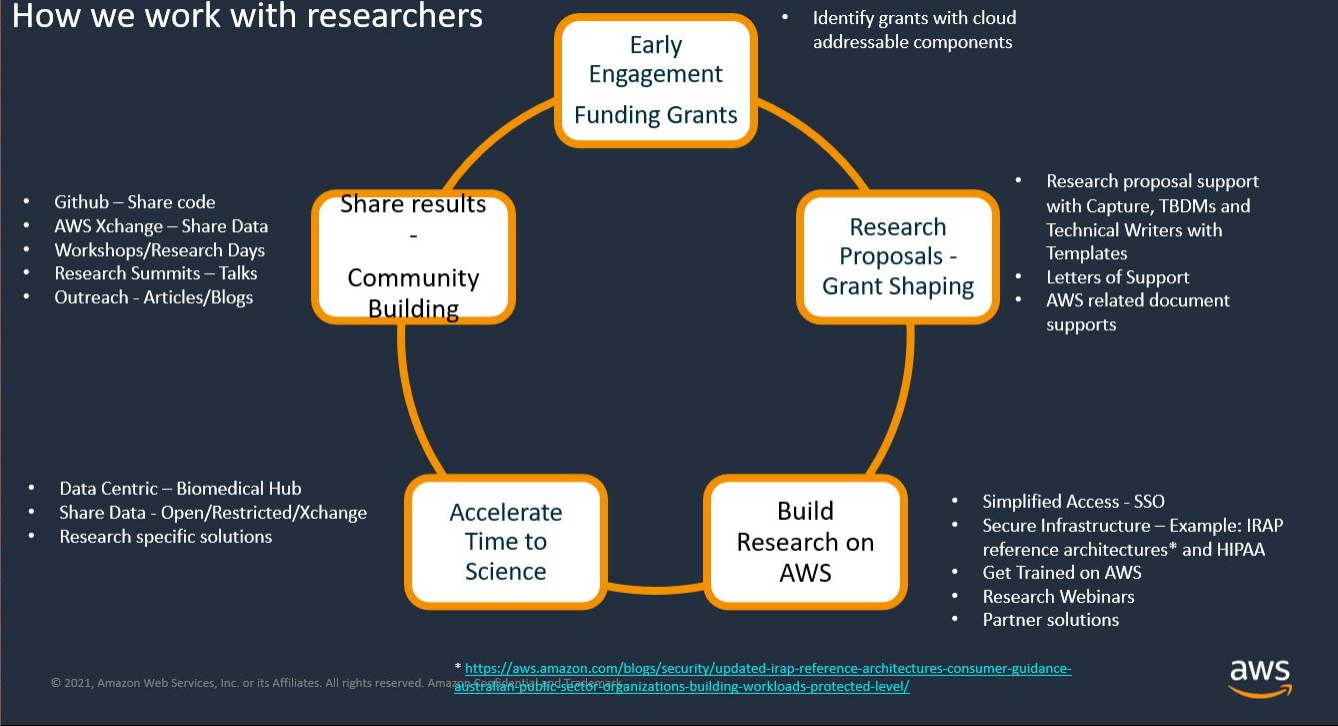
Best-in-class Genomics Workloads

- Illumina DRAGEN and NVIDIA Parabricks, Sentieon Genomics

© 2021, Amazon Web Services, Inc. or its Affiliates. All rights reserved.



How we work with researchers



AWS

More information:

<https://aws.amazon.com/government-education/research-and-technical-computing>

<https://www.amazon.science/academic-engagements>

Training:

<https://aws.amazon.com/government-education/research-and-technical-computing/research-seminar-series/>

https://pages.awscloud.com/APAC-public-OE-AWS-Research-Webinar-Series-On-Demand-2020-reg.html?did=ep_card&trk=ep_card

NCMAS

Roger Edberg

NCMAS

National Computational Merit Allocation Scheme 2022 Call for Applications

NCI: Gadi 280 MSU
Pawsey: Setonix 250 MSU (two components)
MASSIVE: M3 2.5 MSU

Allocations are made through a merit-based competitive process.
Roughly $\frac{3}{4}$ of applicants receive allocations, however many receive less than they request because demand exceeds supply.



NCMAS

2022 Call for Applications

Opens: 18 August

Closes: 5 October

<https://ncmas.nci.org.au>



THE UNIVERSITY OF
SYDNEY



Australian
BioCommons

Next steps

- Get to know the needs of your workloads
- Get access- contact your institution's ICT department
- Training resources for you to get some experience
 - BioCommons training cooperative
 - Galaxy training
- Pipelines and tools already deployed at national facilities
 - Australian BioCommons **ToolFinder**
 - Australian BioCommons **WorkflowHub**



Join us tomorrow ...

High performance bioinformatics: submitting your best NCMAS application

20 August 2021 12pm AEST/10am AWST

biocommons.org.au/events/ncmas



Tell us what you thought ...

Feedback survey



Thanks for joining us!

The Australian BioCommons is enabled by NCRIS via
Bioplatforms Australia funding

