

CAB newsletter 30th March 2022

You are receiving this as a member of the Centre for Applied Bioinformatics (CAB)

Many thanks to all those who helped make the launch this week such a success, lots of attendees and great presentations. We plan to host presentations every two or three months, in person as well as online, if you are interested in giving a short talk, please email me at Dave.Edwards@uwa.edu.au

News:

Not yet joined CAB? Well if you are receiving this you already have joined, but if you know anyone who would like to join, they can now sign up using a simple web form here: https://docs.google.com/forms/d/e/1FAIpQLSex2vZ_24CRgh1Nb0rKLZeN5FnYVZA_-6YRHC6SxbMOuoY0g/viewform?usp=sf_link

Publication highlights:

This is a new section where members can suggest papers from their own group or elsewhere that they want to promote to centre members. They take the simple format of a title, download and short summary. Below are some examples relating to some of the launch presentations. Please forward any papers you would like to highlight to me for inclusion in future newsletters.

Title: Personalised analytics for rare disease diagnostics

Download: <https://www.nature.com/articles/s41467-019-13345-5>

Short summary: An algorithm to discover causative variants in patients suffering from rare and undiagnosed conditions.

Title: Kalign 3: multiple sequence alignment of large datasets

Download: <https://academic.oup.com/bioinformatics/article/36/6/1928/5607735?login=true>

Short summary: Fast multiple sequence alignment algorithm used in wide variety of studies including evolutionary studies and Sars-Cov-2 research.

Title: Maize Yield Prediction at an Early Developmental Stage Using Multispectral Images and Genotype Data for Preliminary Hybrid Selection

Download: <https://www.mdpi.com/2072-4292/13/19/3976>

Short summary: A deep learning model trained with images collected by UAV and field annotations for yield prediction of maize varieties two months after sowing. The code is available on GitHub as indicated in the paper.

Title: Resources for image-based high-throughput phenotyping in crops and data sharing challenges

Download: <https://academic.oup.com/plphys/article/187/2/699/6310753?login=true>

Short summary: Image datasets for plant phenotyping and crop monitoring are usually scattered across multiple platforms. This review highlights where to find the datasets (and associated data) and also talks a bit about the standards for data sharing that are needed to advance the use of machine learning for agriculture.

Title: The Mastery Rubric for Bioinformatics: A tool to support design and evaluation of career-spanning education and training

Download: <https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0225256>

Short summary: Tractenberg et al. identified 12 core Knowledge, Skills and Abilities (KSAs) vital to a bioinformatics career, and describe how each KSA is demonstrated across career stages of a bioinformatician. From these KSAs, Tractenberg et al. build a Mastery Rubric for Bioinformatics (MR-Bi) which provides a formalised framework for teaching and life-long learning.

Training:

NCI bootcamps:

A couple of exciting data science bootcamps lined up for researchers who need to deal with Tbytes or even Pbytes of research data processing pipelines. They are Python based and aims to introduce the powerful GPU for performance in data analysis and ML/DL applications. You will experience the last version of NVIDIA GPU A100 and enjoy the performance improvement with little modification on your existing workflow.

If you are looking for recipes to scale up and out multiple GPUs to accelerate your ML/DL workflows, please sign up this bootcamp (**May 4/5**): <https://gpuhackathons.org/index.php/event/nci-australia-distributed-deep-learning-gpu-bootcamp>. Application deadline is April 5.

If you are looking for an overview of Python Parallel packages and use of GPUs, please sign up this bootcamp (**May 18/19**): <https://gpuhackathons.org/index.php/event/nci-cuda-python-gpu-bootcamp>. Application deadline is April 27.

Pawsey AMD GPU Lunch n Learn & Trainings with 10-day Access to Codelab:

Join Pawsey and AMD specialists for a 90-minute / 10-day Lunch & Learn that consists of:

- **30 March, online:** A 90-minute session on portable GPU programming with ROCm
 - **Register:** <https://pawsey.org.au/event/eoi-amd-gpu-lunch-n-learn-with-10-day-access-to-codelab/>
- **4-5 April, online:** Introduction to AMD GPUs (AMD Instinct Architecture and ROCm)
 - **Register for Day 1:** <https://pawsey.org.au/event/eoi-1-day-introduction-to-amd-gpus-amd-instinct-architecture-and-rocm/>
 - **Register for Day 2:** <https://pawsey.org.au/event/2022-04-05-06-eoi-introduction-to-amd-gpus-amd-instinct-architecture-and-rocm/>

BioChats: Tuesday 5 April 12pm AEST

Regular, informal chat about life science research, bioinformatics and research infrastructure. This month we're talking about R with conversation starters Dr Ira Cooke (James Cook University) on the subject of Combining RStudio in the cloud with github workflows and Dr Sarah Williams (QCIF) on the subject of Using workflowr to share analyses.

[More information and registrations](#)

WEBINAR: Protection of genomic data and the Australian Privacy Act: when is genomic data 'personal information'?

Wednesday 6th April 12pm AEST

Professor Mark Taylor explores the relationship between the legal concept of genetic information and the concept of genomic data relevant to health and medical research, reflect on the characteristics of each, and the possibility of more clearly identifying the legal rights and responsibilities which attach to the use and disclosure of genomic data in the future.

[More information and to register](#)

WORKSHOP: R - fundamental skills for biologists

Four sessions between 1-22 June 2022, 2-5pm AEST.

Getting started with R can be a little daunting if you're new to programming. In this four-part workshop we will equip you with the foundations for efficiently using R and RStudio with biological data. Using gene expression data from a model of influenza infection, you will learn how to efficiently and reproducibly organise, read, wrangle, analyse, visualise and generate reports from your data in R.

Grants:

Australasian Leadership Computing Grants (ALCG):

NCI is seeking Expressions of Interest from researchers for the 2022 round of the [Australasian Leadership Computing Grants \(ALCG\)](#).

ALCG 2022 allocations will be **20-50 MSU** per project, valid for a 12-month period from 1 July 2022 – 30 June 2023.

ALCG aims to identify ambitious and meritorious research projects with a demonstrated ability to use HPC systems effectively at scale. ALCG provides these projects with some of the largest computing resource allocations available in Australia, to push forward the top-end of extreme scale computational science.

All fields of research are eligible and encouraged to apply.

Expressions of Interest are now open. All details on how to apply are in the ALCG 2022 [Information for Applicants](#).

Key dates:

- 31 March 2022 – [Online Information Session](#)
- 10 April 2022 – Expressions of Interest Deadline
- 15 April 2022 – Shortlisted EOIs invited to submit a full application
- 8 May 2022 – Deadline for invited applications
- 16 May 2022 – Allocations announced

Competition:

APAC HPC-AI competition:

NCI is supporting the APAC HPC-AI competition this year. We are calling participants from undergraduate to graduate students. Details can be found here. <https://opus.nci.org.au/display/Help/APAC+HPC-AI+Competition>

APAC HPC-AI competition is a well-recognised competition in the HPC community. The competition will last several months. Students have plenty of time to learn basic HPC skills, to the latest AI, GPU technology. They will be trained with experienced instructors who are experts in those domains. NCI and NSCC will provide compute resources so students can access those resources **for free** during the competition. It is a very good learning opportunity for students who don't have much chance to access HPC and run AI applications. **No prior knowledge about HPC and AI is required**, except for some basic programming experiences. We will train you up! We also welcome local mentor support to help building the

team. More importantly, it is a good learning experience just by participating. The deadline to form a team is **April 29**.

If the team gets the prize, the award ceremony will be held at the SCAsia 2023 in Singapore. You will have opportunities to present your work at the Supercomputing Asia conference. The award of the competition will be a great thing to add to your CVs, too.

Sounds interesting? Please join the information session on April 7. [Register to hear more about the competition](#).

If you have specific questions such as forming a team, local support, etc, please contact me directly Jingbo.Wang@anu.edu.au

Link with the UWA Institute of Data:

Please send an email to uwadatainstitute@uwa.edu.au if you want to be a member of the Data Institute.