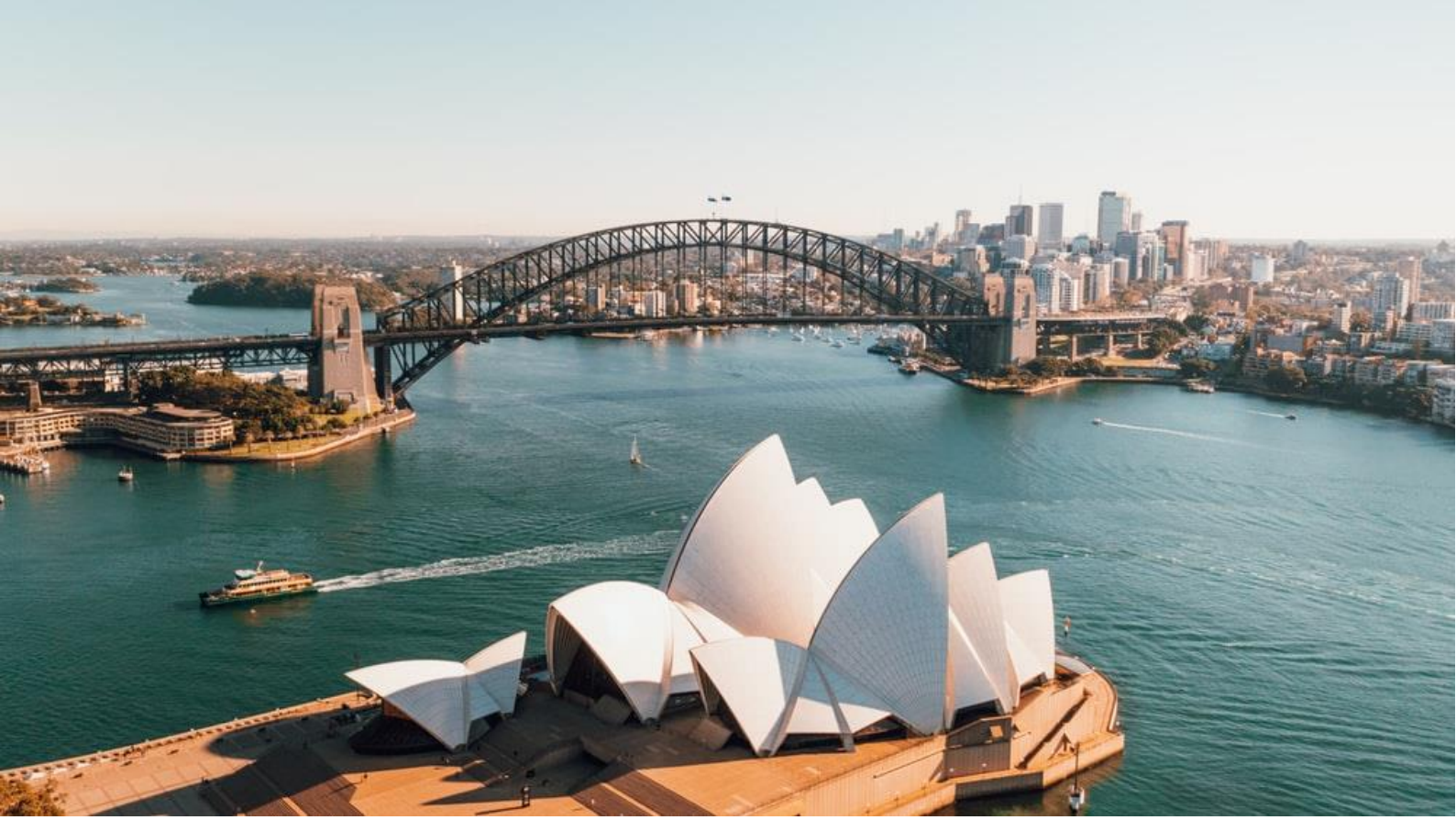
Australia's National
Science Agency

Bioinformatics Student Exchange Program

Start your career with a research project at one of the world's premier research organizations.



1 BSEP 2023

The Bioinformatics Student Exchange Program (BSEP) is aimed at giving overseas students the opportunity to contribute to world-class research and gain experience in an international research environment. Master and Honours students are invited to conduct original research as part of their University Thesis. This is an exciting opportunity to forge new collaboration and build up a strong international network.

Why Australia?

Australia is the “most productive of all G20 nations” with respect to papers published [nature Index] and a recent Nature article says that “Scientists from across the world are attracted to the country, which competes internationally by focusing on its strengths”.

The Commonwealth Scientific and Industrial Research Organisation (CSIRO) is Australia’s Government Research Agency and one of the largest and most diverse scientific organisations in the world. By igniting the creative spirit of our people, CSIRO deliver great science and innovative solutions that benefit industry, society and the environment.

COVID-19: International travel has been opened back up in Nov 2021. However, with volatility in cases future restrictions might prevent the 2023-participants to conduct all or any part of their studies in Australia. Catering for this, we also offer BSEP remotely.

University	Contact Person
Freie Universität Berlin	Ulrike Seyferth Studiengangskoordination Bioinformatik Tel.: +49-(0)30/838-75336 Email: ulrike.seyferth@fu-berlin.de
Eberhard Karls University Tübingen	Prof. Dr. Daniel Huson Algorithms in Bioinformatics Tel.: +49-7071-29-70450 Email.: daniel.huson@uni-tuebingen.de
Justus-Liebig-University Giessen	Prof. Dr. Alexander Goesmann Bioinformatik und Systembiologie Tel. +49 (0)641 99-35801 Email: Gwyneth.schulz@computational.bio.uni-giessen.de
CSIRO	A/Prof. Dr. Denis Bauer Transformational Bioinformatics, eHealth, CSIRO Phone: +61 2 9325 3174 Email: denis.bauer@csiro.au

Key dates

Date	
June	CSIRO calls for project proposals
31st July	Program Booklet sent to the Universities
Early August to early November	Deadline for PROMOS or equivalent funding application
Dec	Thesis committee assesses suitability of projects and identifies appropriate co-supervisor amongst the faculty.
Jan	Students choose proposals and CSIRO starts recruitment process (interview, visa)
May*	Students commence research in Australia
Oct*	Students return home
Nov*	Students finalise reports and write master thesis with input from CSIRO researchers

** Times for visit can be flexible*

How to apply

Please choose the project you are interested in and get in touch with your contact person listed above. Your first step will be to organize funding by applying (see below). After a successful interview in January, CSIRO will issue a contract with a visa sponsorship number. It is crucial to apply for the Australian Visa quickly as it can take up to 3 months to be approved. CSIRO will guide you through the process but please have a look at:

Funding

Students are encouraged to apply for funding. Unless stated otherwise, the projects will not provide funding.

PROMOS

German funding through PROMOS (Deadline Early October to early November), which will cover from 300 to 500 EUR per month or traveling costs up to 1950 EUR

Note, PROMOS is not explicitly paying a health insurance, this hence needs to be covered by the student.

DAAD

The DAAD offers Internationale Forschungsaufenthalte für Informatikerinnen und Informatiker (IFI). 818.93 EUR a month after deductions, travel expenses covered 1,975 EUR.

When receiving the IFI scholarship, the student is automatically insured with "Die Continentale" for the duration of the scholarship + two weeks before and after starting/end date.

There are also other funding sources available such as http://www.ranke-heinemann.de.

Other resources

Please choose the project you are interested in and get in touch with your contact person listed above. Your first step will be to organize funding by applying for PROMOS or equivalent sources (DAAD). After a successful interview in January, CSIRO will issue a contract with a visa sponsorship number. It is crucial to apply for the Australian Visa quickly as it can take up to 3 months to be approved. CSIRO will guide you through the process but please have a look at:

VISA:

The visa subclass 402 doesn't exist anymore, for us the Temporary Activity (subclass 408) visa – Research Activities applies now: https://immi.homeaffairs.gov.au/visas/getting-a-visa/visa-listing/temporary-activity-408/research-activities#Overview

Health insurance:

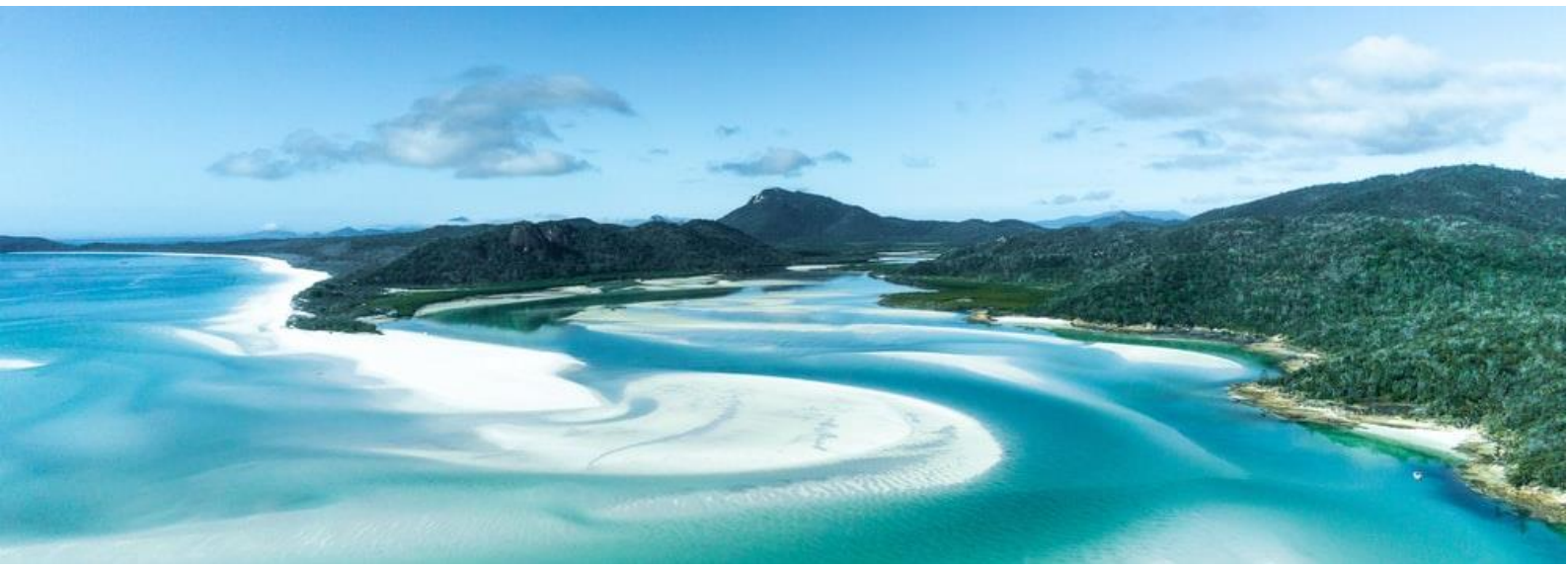
https://www.studyinaustralia.gov.au/english/live-in-australia/insurance

German information on going to Australia:

http://www.reisebine.de/

Official government website with information about studying and living in Australia

www.studyinaustralia.gov.au



2 Experience Reports from previous students

2022 – Aravind Venkateswaran

Working alongside the digital genome engineering team at CSIRO was an exciting and essential experience in my student career. My project involved pre-processing a Pac-bio dataset of AAV capsid genomes, then using machine learning methodologies and statistical analysis to classify high and low performing capsids. I had constant support from my supervisors with 2 meetings per week and also presented in interesting weekly team science meetings. I had no machine learning knowledge but after my 10 month experience, I developed a passion for it and produced fairly robust models that achieved this task.

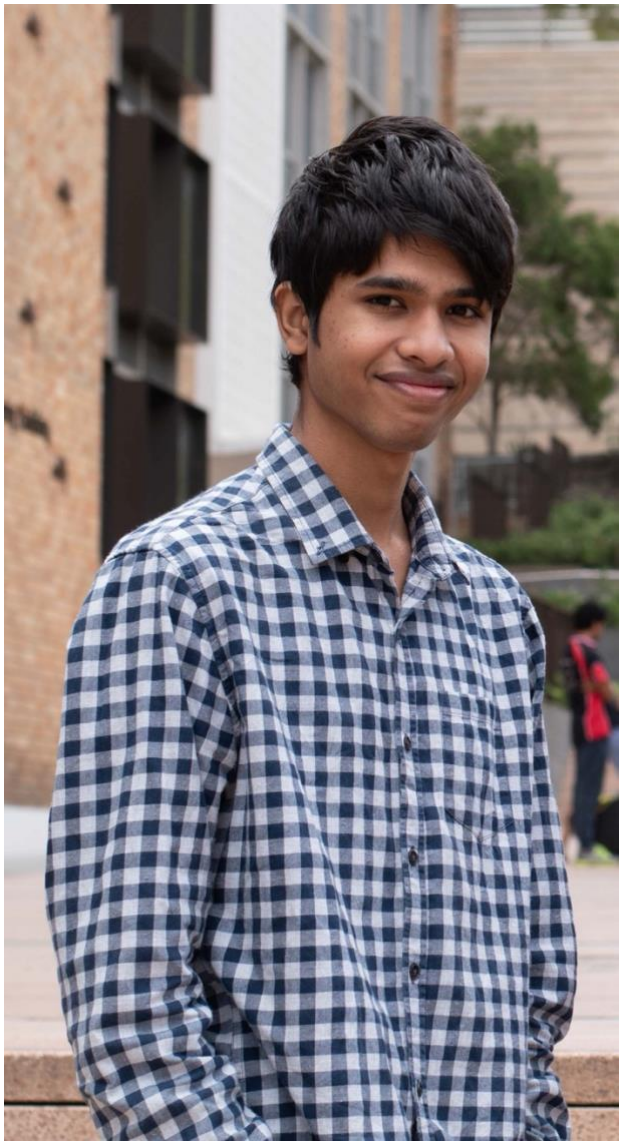


Figure 1 Aravind 2022 domestic BSEP student

Everyone in the team was friendly and encouraged to present in meetings, and submit posters at conferences. Most importantly, my supervisors were open to novel ideas and gave me freedom in literature to pursue my research topic. This helped me identify my passion in a highly diverse and multidisciplinary field, setting me up for the future.

2021 - Felix Hartkopf

As I came closer to the end of my bioinformatics master studies at the Freie Universität Berlin, I realized that I wanted to write my thesis abroad. I learned of the BSEP program through the university and thought it was the perfect fit for me. After choosing a project that I am interested in, an informal application was enough to contact my future supervisors Laurence Wilson and Suzanne Scott. My next steps were applying for additional funding and getting everything required to do so. I applied for the DAAD IFI scholarship and the PROMOS funding. However, you can only have one scholarship and must choose if you are lucky to get accepted for both. The DAAD IFI scholarship is more paperwork but should be enough to support you during your stay, including health insurance and travel costs. The language certificate, visa (Temporary Activity (subclass 408)), and your professor's

project evaluation can take some time, so plan with a buffer in mind.

Due to my planned start in March 2020, I was not able to travel to Australia. After postponing the beginning for some time, I started remotely with the option to travel to Australia if the borders opened again as requested by the DAAD. In hindsight, it would have been better to choose a second guest country to get the full scholarship and work for the CSIRO from there, but this would have prevented the option to go to Australia entirely, and it was hard to find another lab that would host me during that time. However, I spent my project in Germany and worked remotely with my supervisors. It was not what I expected, but it was great to work with the team, and I learned a lot along the way, e.g., about time zones and daylight-saving time. We had multiple meetings every week, and all questions were answered timely.

I do not regret doing the project remotely and would do it again with some minor adjustments. After all, yes, my visa was annulated, but I got credit now with the Department of Home Affairs for my next travel to Australia. So I will go there for an extended vacation as soon as possible.

3 Project

Projects can be altered to fit the students interests and skills. The Transformational Bioinformatics group at CSIRO has a very broad spectrum of activities, ranging from human health to biosecurity; from basic science to real-world applications. We highly encourage you to check our webpage (https://bioinformatics.csiro.au/) for our activities and approach us with your **own project ideas**.

PROJECT **8**

BSEP01 Predicting disease-risk from genomic data *8*
BSEP02 Science Writing/Communication for Blog posts *9*
BSEP03 Programming CRISPR scoring methods *9*
BSEP04 Machine Learning Model Selection for genetic data *10*
BSEP05 Population Genomic Simulation for Complex Phenotype PEPS *10*
BSEP06 Database solutions for COVID-19 metadata analysis *11*
BSEP07 UI/UX DevOps for Genome Editing Platform *12*
BSEP08 Cluster analysis of pathogenic COVID-19 strains *12*
BSEP09 Identification of feature interactions for continuous variables *13*
BSEP10 Exploring the background on our identified COVID-19 Variants of Concern *14*
BSEP11 Enhancing genomic searching using metadata *14*
BSEP12 Identify long-term COVID possibility for different SARS-CoV-2 variants *15*

Project Title	**BSEP01 Predicting disease-risk from genomic data**
Brief description of the project highlighting expected outcomes	How much of your disease risk is really encoded in the genome? Find out with Machine Learning (ML) and the world's largest genomic dataset. The student will learn how to run VariantSpark in an interactive notebook and optimise prediction/classification using a set of genetic drivers for phenotype of interest. The student will tackle common multi-genic diseases e.g. Alzheimer's/ALS. The resulting notebooks will help standardized the analysis of poly-genic-risk (PRS) models for a wide range of diseases annotated in the UKbiobank, i.e. supplementing the GWAS catalogue using ML methods.
Duties/Tasks	The student will perform • VariantSpark analysis • Use HAIL/AWS with python notebooks for large scale data analysis • use ML to build predictive models with genetic data

Relevant field/s of study	• Machine learning • Bioinfomatics
Supervisor	Natalie Twine
Contact Details	natalie.twine@csiro.au
Location	Remote or in person in Sydney

Project Title	**BSEP02 Science Writing/Communication for Blog posts**
Brief description of the project highlighting expected outcomes	The student will work with science experts to write up complex scientific concept into communications pieces that are understandable by the general public. For IT project, the student will document the functionality, algorithms and optimisations in a markdown format with clear diagrams. Where applicable, the student will create non-scientific diagrams to aid the understanding. The resulting blog post will be published on bioinformatics.csiro.au.

Duties/Tasks	The student will perform
	• Compile easy to understand blog articles
	• System analysis and produce software diagrams including, but not limited to;
	  ○ System Sequence Diagrams
	  ○ Architecture Diagrams
	• Implement reproducible boilerplates (terraform/ serverless/ CloudFormation)

Relevant field/s of study	• Science communication • Bioinformatics • Graphic design
Supervisor	Anuradha Wickramarachchi
Contact Details	Anuradha.Wickramarachchi@csiro.au
Location	Remote or in person in Adelaide

Project Title	**BSEP03 Programming CRISPR scoring methods**
Brief description of the project highlighting expected outcomes	The student will learn how to 1) utilise existing CRISPR scoring algorithms (MIT, Tuscan); and 2) perform research into emerging CRISPR scoring methods (e.g. Cas-13) to develop reusable high-performance computing software libraries within a cloud-computing development environment. The resulting score-function will be included into the GT-Scan web service.
Duties/Tasks	The student will perform • Programming Python/C++/AWS • Translation of biological/genome-editing research into cloud software

Relevant field/s of study	• Bioinformatics • Software Programming • Cloud computing
Supervisor	Tracey Wright
Contact Details	tracey.wright@csiro.au
Location	Remote or in person in Brisbane

Project Title	**BSEP04 Machine Learning Model Selection for genetic data**
Brief description of the project highlighting expected outcomes	You will learn to execute machine learning (ML) methods on DNA samples and analyse the results. Some of the diseases of interest are Diabetes, Alzheimer's, Rheumatoid Arthritis, and ALS. We'll provide you with some pointers on how to evaluate the model but also the freedom to pursue your own ideas. At the end of the project the aim is to have a comprehensive set of evaluation methods for quantifying the performance of the ML algorithms.
Duties/Tasks	The student will perform • VariantSpark analysis (random forest) • AWS + python (R is also possible) • Identify and pursue metrics to quantify the model's performance • Interpret and summarise results
Relevant field/s of study	• Bioinformatics • Computer science • Epidemiology
Supervisor	Roc Reguant
Contact Details	Roc.reguant@csiro.au
Location	Remote or in person in Sydney

Project Title	**BSEP05 Population Genomic Simulation for Complex Phenotype PEPS**

Brief description of the project highlighting expected outcomes	Access to genetic cohorts is limited and generating new ones is costly. Also, privacy and security are at utmost importance with this information. To overcome this problem there are several tools that generate synthetic datasets. However, current methods focus on mutation frequencies at specific loci while ignoring the epistatic interactions. We want do finalize the development of a tool that is able to generate a dataset resembling the original population and taking SNP-SNP interactions into consideration.
Duties/Tasks	The student will perform • VariantSpark analysis (random forest) • AWS + python (R is also possible) • Develop generative methods for epistatic interactions • Interpret and summarise results
Relevant field/s of study	• Bioinformatics • Computer science • Epidemiology
Supervisor	Roc Reguant Mitchell O'Brien
Contact Details	Roc.reguant@csiro.au mitchell.o'brien@csiro.au
Location	Remote or in person in Sydney

Project Title	**BSEP06 Database solutions for COVID-19 metadata analysis**
Brief description of the project highlighting expected outcomes	Millions of genomic samples of the COVID-19 virus are collected around the world and the cumulative data is reaching the limits of current infrastructure. The student will explore different database solutions i.e. relational and non-relational with the aim to better incorporate metadata (e.g. sample information) to the genomic pipeline.
Duties/Tasks	The student will • Understand COVID19 genomic analysis • Explore database solutions for metadata • Incorporate database to existing genomic pipeline
Relevant field/s of study	• Bioinformatics • Database management
Supervisor	Yatish Jain, Carol Lee
Contact Details	Yatish.Jain@csiro.au Carol.Lee@CSIRO.au
Location	Remote or in person in Sydney

Project Title	**BSEP07 UI/UX DevOps for Genome Editing Platform**
Brief description of the project highlighting expected outcomes	Genome Editing holds promise for a wide range of application areas, from human health to biosecurity. However, the key to unlocking this potential lies in enabling researchers to use the powerful but complex data analysis engine GT-Scan. The student will hence work on an intuitive user interface that empowers researchers to design genome editing experiments with ease.
Duties/Tasks	The student will perform • Conducting design interviews with users for desirable features (wireframe prototypes • Implement the Frontend on AWS cloud • Create visuals to advertise the tool
Relevant field/s of study	• DevOps • UX/UI experience • cloud experience
Supervisor	Denis Bauer
Contact Details	Denis.Bauer@CSIRO.au
Location	Remote or in person in Sydney

Project Title	**BSEP08 Cluster analysis of pathogenic COVID-19 strains**

Brief description of the project highlighting expected outcomes	This project will build on a recent paper published in the lab regarding pathogenic and protective mutations in SARS-CoV-2 viral genomes. We will aim to cluster >12,000,000 viral samples in GISAID to determine patterns and whether pathogenic and protective strains cluster together and delve into the features shared in these clusters. The outcomes of this project will include analysis which could form the basis of a follow-up paper of our publication, and we will collaborate with RONIN and Intel to optimise our solutions.
Duties/Tasks	The student will: - Help build the optimal EMR configurations for the analysis of millions of datapoints - Evaluate different cluster methods to find the optimal solution for our viral genome dataset - Explore feature extraction methods
Relevant field/s of study	· Experience with machine learning and feature extraction methods · cloud experience
Supervisor	Priya Ramarao-Milne
Contact Details	priya.ramarao-milne@csiro.au
Location	Remote or in person in Brisbane

Project Title	**BSEP09 Identification of feature interactions for continuous variables**
Brief description of the project highlighting expected outcomes	In biology, different systems interact and modulate each other. Current methods assume that one biomarker has one effect on our body. However, that is an oversimplification. We would like to improve current methods and develop a pipeline to quantify the additive effects interactions between biomarkers have on diseases. If successful, your name will be on a scientific publication.
Duties/Tasks	The student will: - Implement well known Machine learning methods and algorithms - AWS + python (R also possible) - Identify interacting model features

	- Interpret and summarise results
Relevant field/s of study	· Bioinformatics · Computer science
Supervisor	Roc Reguant Priya Ramarao-Milne
Contact Details	Roc.reguant@csiro.au priya.ramarao-milne@csiro.au
Location	Remote or in person in Sydney

Project Title	**BSEP10 Exploring the background on our identified COVID-19 Variants of Concern**
Brief description of the project highlighting expected outcomes	The SARS-CoV-2 genome comprise of numerous mutations of varying significance and many of these are understudied. The student will use known signature mutations and other relevant mutations of SARS-CoV-2 strains to explore the adaptive advantage provided by a combination of these mutations. This will focus on the Alpha, Beta, Gamma, Delta, and Omicron Variants of Concern but may involve other strains of importance or high frequency. The project will allow us to gain a better understanding of the genetic make-up of SARS-CoV-2 and characteristics enhancing virulence.
Duties/Tasks	The student will perform • Comparisons between and within major SARS-CoV-2 strains • Investigations into co-evolving mutations and their frequencies
Relevant field/s of study	• Bioinformatics • Phylogenetics
Supervisor	Carol Lee
Contact Details	Carol.lee@csiro.au
Location	Remote or in person in Sydney

Project Title	**BSEP11 Enhancing genomic searching using metadata**
Brief description of the project highlighting expected outcomes	sBeacon presents a novel and scalable method to rapidly search multiple genomic datasets for the presence and frequency of mutations of interest. Students will expand the capabilities of the service allowing, for example, the ability to search not just by genotype, but by phenotype, and to use the associated metadata to perform analysis across the phenome-genome space of millions of samples.
Duties/Tasks	The student will:

	• Add new search functionality over sample metadata • Optimise the query pipeline to handle the increased data throughput • Perform sample analysis to showcase capability
Relevant field/s of study	• Computer Science • Bioinformatics • Cloud Experience
Supervisor	Brendan Hosking
Contact Details	brendan.hosking@csiro.au
Location	Remote or in person in Sydney

Project Title	**BSEP12 Identify long-term COVID possibility for different SARS-CoV-2 variants**
Brief description of the project highlighting expected outcomes	Genomes of different variants of SARS-CoV-2 are available in GISAID, however the effects of protein fragments that remain after infection in the body is unknown. This project entails cataloging different peptides that can be generated from different SARS-CoV-2 variants and obtaining their pathogenicity score via pepbank. Overlaying this information with omics data from people with Long COVID may shed light on the molecular bases of this condition.
Duties/Tasks	The student will: • Write scripts to generate peptides of variable lengths • Compare the generated peptides against peptide databases like pepbank or peptide atlas. • Generate visualisations to present the metrics for different variants • Submit analysis to https://virological.org/
Relevant field/s of study	• Bioinformatics • Computer Science • Proteomics
Supervisor	Yatish Jain
Contact Details	Yatish.jain@csiro.au
Location	Remote or in person in Sydney

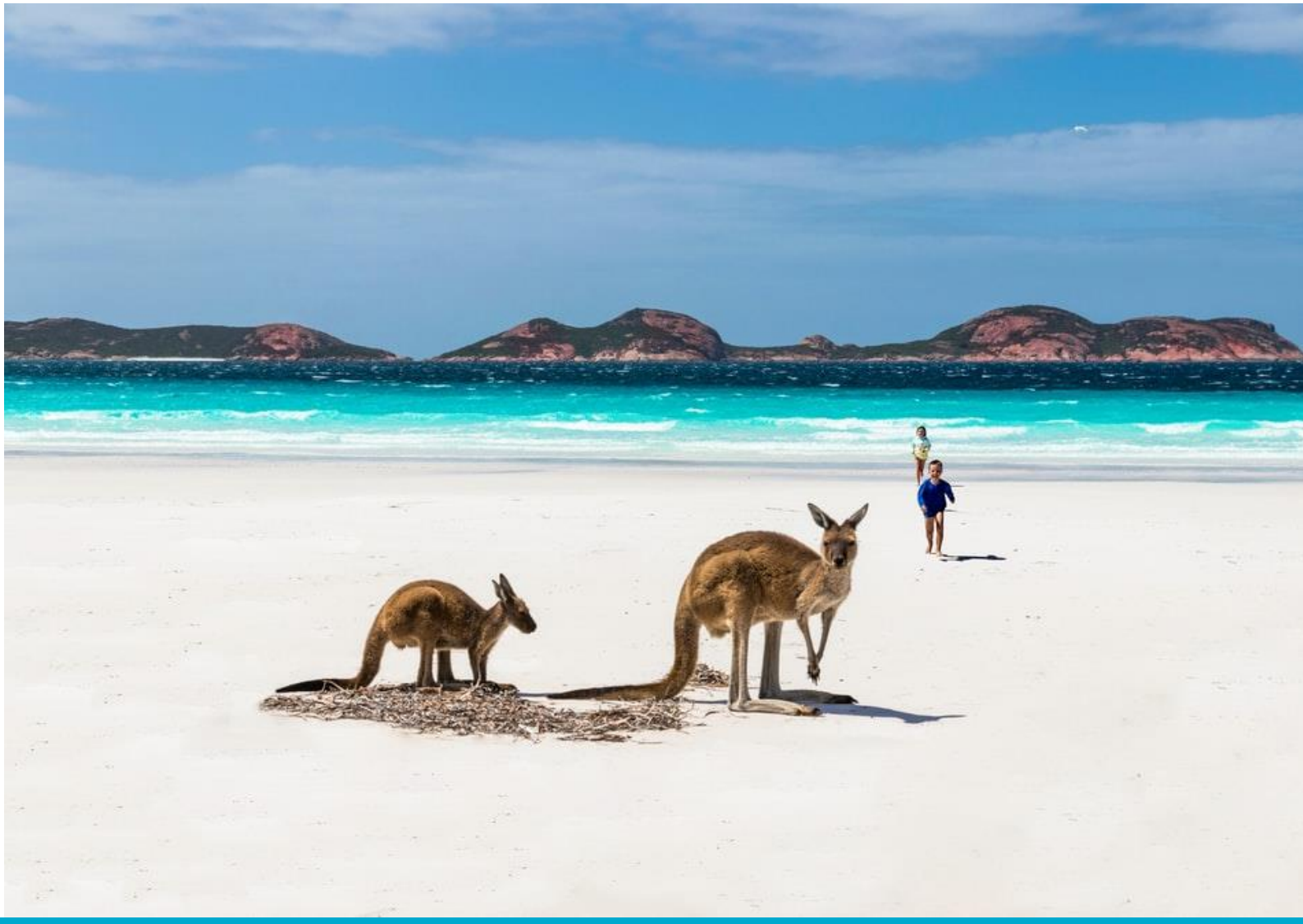


Image copyright Unsplash.

As Australia's national science agency and innovation catalyst, CSIRO is solving the greatest challenges through innovative science and technology.

CSIRO. Unlocking a better future for everyone.

Contact us

1300 363 400
+61 3 9545 2176
csiroenquiries@csiro.au
csiro.au

For further information

Health & Biosecurity
Denis Bauer
+61 2 9325 3174
Denis.Bauer@csiro.au
csiro.au/H&B