



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.

# 2025





# 1 BSEP 2025

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 all or any part of their studies in Australia. Catering for this, we also offer BSEP remotely.

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

---

## Key dates

Date	
June	CSIRO calls for project proposals
31 <sup>st</sup> July	Program Booklet sent to the Universities
Early August to early November	Students choose proposals and get in contact with CSIRO
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 <u>faculty</u> .
Jan	CSIRO starts recruitment process (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, CSIRO will issue a contract with a visa sponsorship number in January. 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 Study Scholarships, such as Master Studies for All Academic Disciplines <https://www2.daad.de/deutschland/stipendium/datenbank/en/21148-scholarship-database/?detail=50026200> with funding between 10 and 24 months.

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](http://www.studyinaustralia.gov.au)



## 2 Experience Reports from previous students



### **2024 – Michael de Francesco**

Looking back on my time at CSIRO, I can say that I was able to build on the technical skills and knowledge that I learnt during my time at university, as well as learn how to work in a team environment on a day-to-day basis, which I something I experienced for the first time.

As I undertook this traineeship after my studies, I was keen to apply my biological background and software engineering skills that I had built over four years to this twelve-week project that I was assigned. As CSIRO work in the research space, I enjoyed not just being assigned a task, but being able to have input into the scope of the project. This was possible because I was in constant communication with my supervisor and team which allowed us to collaboratively design a plan.

Being able to pursue what interested me and the team allowed me to enjoy my work and therefore be more productive. The difference at CSIRO compared to university was that my project was more individually driven, and I was the youngest member of the team with the least amount of experience. Being able to draw ideas and skills from older team members from all over Australia who had been in research for multiple years was very beneficial and broadened my knowledge.

I was exposed to deep learning algorithms, such as AlphaFold, a protein structure prediction algorithm and Evo, a DNA sequence generation tool. Being able to learn from others who have had experience in the machine learning space was very beneficial to me and outside the scope of bioinformatic tools that I used while at university.

The difference of being able to use emerging tools on the forefront of research at CSIRO compared to university was very rewarding. Another thing that I enjoyed whilst working at CSIRO was being able to fully dedicate my time to a single project. Being able to break up my time between literature review, algorithm experimentation and presenting my findings also ensured I was not doing mundane work. Now I feel confident that I can learn new tools and algorithms that emerge in the future, better equipping me for my career at the conclusion of this traineeship.

# 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**.

<b>3</b>	<b>PROJECT .....</b>	<b>7</b>
	<i>BSEP01 Predicting disease-risk from genomic data using machine learning .....</i>	<i>8</i>
	<i>BSEP02 An agent-based framework for ETL of health records and analytics .....</i>	<i>9</i>
	<i>BSEP03 Developing Identity Management for Dynamic Consent Protocol.....</i>	<i>10</i>
	<i>BSEP04 Investigating the Impact of Population Stratification on Random Forests in GWAS using VariantSpark.....</i>	<i>11</i>
	<i>BSEP05 Building a Population Pangenome Graph for Variants Discovery in Underrepresented Populations.....</i>	<i>12</i>
	<i>BSEP06 FHIR protocol into Dynamic Consent Protocol.....</i>	<i>13</i>
	<i>BSEP07 Advancing our understanding of zoonotic evolution in pandemic influenza using AI/ML approaches .....</i>	<i>14</i>
	<i>BSEP08 Automated Classification of Influenza Strains Using NLP and Machine Learning .....</i>	<i>15</i>
	<i>BSEP09 Developing Antiviral CRISPR/Cas13 Guides Robust to Escape Mutants.....</i>	<i>16</i>
	<i>BSEP10 Develop a new approach for using random forest machine learning to identify interactions.....</i>	<i>17</i>
	<i>BSEP11 Improving Flu and Dengue Analysis with Detailed Climate Data.....</i>	<i>18</i>
	<i>BSEP12 Developing bioinformatic workflow pipeline .....</i>	<i>19</i>

<b>Project Title</b>	<b>BSEP01 Predicting disease-risk from genomic data using machine learning</b>
<b>Brief description of the project highlighting expected outcomes</b>	There is a significant genetic component to disease risks in humans. Can we predict these risks early? Find out with Machine Learning (ML) and the world's largest genomic datasets. The student will compare the predictive performance of existing genomic risk models against ML algorithms for common (polygenic) diseases (e.g., Alzheimer's, cardiovascular diseases). At the end of the project, <u>the student will determine the most suitable algorithm for generating polygenic risk scores, considering computational efficiency and its applicability across different ethnicities.</u>
<b>Duties/Tasks</b>	The student will perform <ul style="list-style-type: none"> <li>• Trait prediction analysis using VariantSpark</li> <li>• Use HAIL/AWS with python notebooks for large scale data analysis</li> <li>• Use ML algorithms to build predictive models with genetic data</li> <li>• Compare currently available PRS models (e.g., LD-pred2, LDAK-SumHer, PRS-CS, PRSice) with ML algorithms</li> </ul>
<b>Relevant field/s of study</b>	<ul style="list-style-type: none"> <li>• Machine learning</li> <li>• Bioinformatics</li> </ul>
<b>Supervisor</b>	Letitia Sng Anubhav Kaphle
<b>Contact Details</b>	<a href="mailto:Letitia.sng@csiro.au">Letitia.sng@csiro.au</a> <a href="mailto:Anubhav.kaphle@csiro.au">Anubhav.kaphle@csiro.au</a>
<b>Location</b>	Remote or in person in Sydney



<b>Project Title</b>	<b>BSEP02 An agent-based framework for ETL of health records and analytics</b>
<b>Brief description of the project highlighting expected outcomes</b>	An agent-based framework for health data ETL (Extract, Transform, Load) using natural language streamlines data processing with advanced extractors, joiners, and formatters. Users can execute these to manipulate data frames through natural language commands, facilitating seamless integration and analysis of diverse health data sources. This approach simplifies complex data operations, allowing non-technical users to interact with data conversationally. Extracted data is transformed and loaded into a structured format for analysis. Code for data analysis is generated from natural language descriptions. <u>Student will implement a pipeline with agent-based extractors, transformers and analytical script executors that will be executed using LLM based function calling.</u>
<b>Duties/Tasks</b>	The student will perform <ul style="list-style-type: none"> <li>• Design basic extractor agents (from JSON and VCF formats to tabular form)</li> <li>• Data joiners/transformer agents (to combine results of extractors using table fields)</li> <li>• Prompt engineering LLMs to generate code to perform analytics on joined data frames (executor agents)</li> <li>• Basic visualisations (frequency plots, heatmaps, etc)</li> </ul>
<b>Relevant field/s of study</b>	<ul style="list-style-type: none"> <li>• Machine learning (Scikit-learn, matplotlib)</li> <li>• Bioinformatics (Familiarity with VCF format)</li> <li>• Python programming (analytics, and plots)</li> </ul>
<b>Supervisor</b>	Anuradha Wickramarachchi
<b>Contact Details</b>	<a href="mailto:Anuradha.Wickramarachchi@csiro.au">Anuradha.Wickramarachchi@csiro.au</a>
<b>Location</b>	Remote or in person in Adelaide

<b>Project Title</b>	<b>BSEP03 Developing Identity Management for Dynamic Consent Protocol</b>
<b>Brief description of the project highlighting expected outcomes</b>	<p>In biomedical research, the concept of Dynamic Consent allows individuals to manage their consent permissions for personal data usage in real-time through online platforms. This project aims to investigate how we can enhance the security and efficiency of managing consent using self-sovereign identity (SSI) principles and blockchain technology.</p> <p>In this project, you will work with our in-house digital wallet that stores and presents credentials. The goal is to model a system where a user can interact with three different institutions, providing and updating their consent for data use. <u>You will explore how consent can be turned into a verifiable credential, shared securely across multiple organisations, and how it can lead to better data access control and management.</u></p>
<b>Duties/Tasks</b>	<p>The student will perform</p> <ul style="list-style-type: none"> <li>• Review existing identity management tools and blockchain technologies and literature</li> <li>• Model interactions between different organisations and credentials consent</li> <li>• Develop transactions models on Dynamic Consent platform</li> <li>• Conduct testing and validation of the identity management solution using the <i>Sovrin testnet</i></li> </ul>
<b>Relevant field/s of study</b>	<ul style="list-style-type: none"> <li>• Software engineering</li> <li>• Web3.0</li> <li>• Bioinformatics</li> </ul>
<b>Desired skillsets</b>	Experience with JavaScript, python and some exposure to software engineering
<b>Co- supervisors</b>	Adrien Oliva, Anubhav Kaphle
<b>Contact Details</b>	<a href="mailto:Adrien.oliva@csiro.au">Adrien.oliva@csiro.au</a> <a href="mailto:Anubhav.kaphle@csiro.au">Anubhav.kaphle@csiro.au</a>
<b>Location</b>	Remote or in person in Melbourne/Adelaide

<b>Project Title</b>	<b>BSEP04 Investigating the Impact of Population Stratification on Random Forests in GWAS using VariantSpark</b>
<b>Brief description of the project highlighting expected outcomes</b>	<p>Genome-Wide Association Studies (GWAS) are a method used to identify genetic variants associated with specific traits or diseases by scanning the entire genome. Traditional GWAS often use logistic regression to analyse the relationship between genetic variants and traits. However, machine learning (ML) approaches, such as those using Random Forests (RF), offer a powerful alternative that can handle complex interactions between variants.</p> <p>One challenge in GWAS is population stratification, where differences in ancestry among study participants can lead to spurious associations. This project aims to explore how population stratification affects the results of GWAS when using ML approaches, specifically by analysing the tree-splitting behavior in Random Forests generated by VariantSpark.</p>
<b>Duties/Tasks</b>	<p>The student will perform</p> <ul style="list-style-type: none"> <li>• Association analysis using VariantSpark</li> <li>• Use HAIL/AWS with python notebooks for large scale data analysis</li> <li>• Understanding how population stratification affects tree splitting in RF-based GWAS.</li> </ul>
<b>Relevant field/s of study</b>	<ul style="list-style-type: none"> <li>• Machine learning</li> <li>• Bioinformatics</li> </ul>
<b>Supervisor</b>	Mitchell O'Brien Letitia Sng
<b>Contact Details</b>	<a href="mailto:Mitchell.O'Brien@csiro.au">Mitchell.O'Brien@csiro.au</a> <a href="mailto:Letitia.sng@csiro.au">Letitia.sng@csiro.au</a>
<b>Location</b>	Remote or in person in Sydney

<b>Project Title</b>	<b>BSEP05 Building a Population Pangenome Graph for Variants Discovery in Underrepresented Populations</b>
<b>Brief description of the project highlighting expected outcomes</b>	This project aims to construct a pangenome graph for African / Middle Eastern population using real genomic data. Pangenome graphs are valuable as they can mitigate reference bias, a common issue when studying genetically diverse populations using a linear reference genome. <u>The expected outcome is a robust pangenome graph that can be used for downstream population genomics analyses and machine learning approaches</u> to identify novel variants associated with cardiovascular diseases (CHD) or other diseases, which disproportionately impact some communities.
<b>Duties/Tasks</b>	The student will <ul style="list-style-type: none"> <li>• Collect and preprocess genomic data from various sources</li> <li>• Construct a pangenome graph using state-of-the-art algorithms and bioinformatics tools</li> <li>• Implement bioinformatics pipelines for population genetics analyses</li> <li>• Use Machine Learning (Variant Spark) techniques to identify potential diseases-associated variants</li> <li>• Document the process, methods, and findings for reproducibility and publications</li> </ul>
<b>Relevant field/s of study</b>	<ul style="list-style-type: none"> <li>• Bioinformatics</li> <li>• Programming (Python, bash)</li> </ul>
<b>Supervisor</b>	Adrien Oliva
<b>Contact Details</b>	<a href="mailto:Adrien.oliva@csiro.au">Adrien.oliva@csiro.au</a>
<b>Location</b>	Remote or in person in Adelaide

<b>Project Title</b>	<b>BSEP06 FHIR protocol into Dynamic Consent Protocol</b>
<b>Brief description of the project highlighting expected outcomes</b>	The Dynamic Consent Protocol (DCP) enables individuals to manage consent permissions for their personal data across various platforms and applications. This project aims to incorporate the FHIR (Fast Healthcare Interoperability Resources) protocol into the DCP framework, allowing participants to share their consent preferences and personal data in a standardized and interoperable manner. <u>The expected outcome is a seamless integration of FHIR within the DCP solutions, enabling participants to fill out FHIR forms and effectively communicate their consent regarding their personal data.</u>
<b>Duties/Tasks</b>	The student will <ul style="list-style-type: none"> <li>• Conduct a comprehensive analysis of the FHIR protocol and its specifications to the</li> <li>• Create a FHIR form for participant following ensuring compatibility and adherence to standards and guidelines</li> <li>• Implement the FHIR form within the DCP framework</li> <li>• Develop user-friendly interfaces for participants to fill out FHIR forms and provide consent</li> <li>• Perform thorough testing and validation to ensure the correct functionality of the FHIR integration with the AWS backend</li> </ul>
<b>Relevant field/s of study</b>	<ul style="list-style-type: none"> <li>• Bioinformatics</li> <li>• Programming (preferably javascript and python)</li> </ul>
<b>Supervisor</b>	Adrien Oliva
<b>Contact Details</b>	<a href="mailto:Adrien.oliva@csiro.au">Adrien.oliva@csiro.au</a>
<b>Location</b>	Remote or in person in Adelaide



<b>Project Title</b>	<b>BSEP07 Advancing our understanding of zoonotic evolution in pandemic influenza using AI/ML approaches</b>
<b>Brief description of the project <u>highlighting expected outcomes</u></b>	<p>Timely and proactively detecting pandemic influenza is crucial for advancing flu antiviral strategies, reducing species mortality, and preventing financial losses. The 2009 H1N1 pandemic had a complex triple assortment where antigenic shift between human, swine and avian allowed for cross-species infection. The molecular factors involved in governing the host range and potential for host-to host transmission are mostly unknown. Identifying these factors will be crucial to detecting potential pandemic strains and mitigating the risks as a result.</p> <p>The project will expand on the current research, using AI/ML to understand how zoonotic evolution occurs, the molecular mechanisms involved and the how this influences the emergence of pandemic influenza strains.</p>
<b>Duties/Tasks</b>	<p>The student will perform</p> <ul style="list-style-type: none"> <li>· Curation of a dataset of influenza A genomes and assign multi-host groups</li> <li>· Rule-mining to determine genomic positions associated with the host range</li> <li>· Prediction of pandemic sequences from mined rules</li> </ul>
<b>Relevant field/s of study</b>	<ul style="list-style-type: none"> <li>· Machine learning (Multi-label Rule-based classification)</li> <li>· Bioinformatics (alignment, BLAST)</li> <li>· Programming (python)</li> </ul>
<b>Supervisor</b>	Nehleh Kargarfard Carol Lee
<b>Contact Details</b>	<a href="mailto:nehleh.kargarfard@csiro.au">nehleh.kargarfard@csiro.au</a> <a href="mailto:carol.lee@csiro.au">carol.lee@csiro.au</a>
<b>Location</b>	Remote or in person in Sydney

<b>Project Title</b>	<b>BSEP08 Automated Classification of Influenza Strains Using NLP and Machine Learning</b>
<b>Brief description of the project <u>highlighting expected outcomes</u></b>	<p>This project aims to develop a text mining system that automatically assigns host-specific labels—human, swine, and avian—to various influenza strains mentioned in a corpus of 150 academic papers. The system will analyze contextual information to determine the appropriate categories, improving the efficiency and accuracy of influenza strain classification.</p> <p>Influenza viruses exhibit significant diversity and cross-species transmission potential, necessitating accurate classification and tracking of strains across different hosts. Current manual methods for labelling strains are time-consuming and prone to errors. By leveraging natural language processing (NLP) and machine learning techniques, this project aims to automate the extraction and classification process, thus improving efficiency and accuracy.</p>
<b>Duties/Tasks</b>	<p>The student will perform</p> <ul style="list-style-type: none"> <li>• Data Preprocessing</li> <li>• Use NLP techniques to identify and extract contextual clues about host species (human, swine, avian).</li> <li>• Develop a Python script that integrates NLP models for contextual analysis.</li> <li>• Train a machine learning classifier to assign appropriate labels based on extracted context.</li> <li>• Validate the system using a separate dataset to ensure accuracy.</li> </ul>
<b>Relevant field/s of study</b>	<ul style="list-style-type: none"> <li>· Natural Language Processing Libraries (spaCy, NLTK, ...)</li> <li>· Machine Learning</li> <li>· Bioinformatics</li> </ul>
<b>Supervisor</b>	Nehleh Kargarfard Carol Lee
<b>Contact Details</b>	<a href="mailto:nehleh.kargarfard@csiro.au">nehleh.kargarfard@csiro.au</a> , <a href="mailto:carol.lee@csiro.au">carol.lee@csiro.au</a>
<b>Location</b>	Remote or in person in Sydney

<b>Project Title</b>	<b>BSEP09 Developing Antiviral CRISPR/Cas13 Guides Robust to Escape Mutants</b>
<b>Brief description of the project <u>highlighting expected outcomes</u></b>	<p>Objective</p> <p>This project focuses on developing predictive models to design antiviral Cas13 guides that are robust against escape mutations in RNA viruses. By leveraging artificial intelligence (AI), the aim is to analyze large genomic datasets to identify patterns in guide efficiency, particularly focusing on off-target effects. This will result in the creation of adaptable antiviral solutions to mitigate the impact of viral outbreaks in livestock, thereby advancing RNA therapies and improving the resilience of livestock to viral threats.</p>
<b>Duties/Tasks</b>	<p>The student will perform</p> <ul style="list-style-type: none"> <li>· Data Preprocessing</li> <li>· Curation of a comprehensive dataset on off-target effects of Cas13 guides from publicly available sources.</li> <li>· Feature engineering to capture critical variations affecting off-target Cas13 guide efficiency.</li> <li>· Train an off-target model by implementing and evaluating various machine learning models (Random Forest, Convolutional Neural Networks).</li> <li>· Validation and refinement of the off-target predictive model using independent datasets.</li> <li>· Integration of findings with data from related subprojects to improve the overall predictive accuracy and guide design processes.</li> </ul> <p>Validate the system using a separate dataset to ensure accuracy.</p>
<b>Relevant field/s of study</b>	<ul style="list-style-type: none"> <li>• Machine Learning</li> <li>• Deep Learning</li> <li>• Bioinformatics</li> </ul>
<b>Supervisor</b>	Emiliana Weiss
<b>Contact Details</b>	<a href="mailto:Emiliana.weiss@csiro.au">Emiliana.weiss@csiro.au</a>
<b>Location</b>	Remote or in person in Canberra

<b>Project Title</b>	<b>BSEP10 Develop a new approach for using random forest machine learning to identify interactions</b>
<b>Brief description of the project highlighting <u>expected outcomes</u></b>	<p>Searching for genetic interactions using random forests involves a very large search space for each node of each tree. Because random forests only include a small subset of features when picking splitting features for a node, it is likely that interacting features will be missed. The aim of this project is to make alterations to the random forest machine learning algorithm to improve its sensitivity to interactions between features.</p> <p><u>The expected outcome is an improved random forest algorithm that can more precisely identify interacting features in feature-rich datasets such as GWAS, including features that have a low marginal effect.</u></p>
<b>Duties/Tasks</b>	<p>The student will</p> <ul style="list-style-type: none"> <li>• Design a method for improving random forest sensitivity to interacting features (e.g. variable mTry, stochastic feature selection)</li> <li>• Tune the specific hyperparameters of the method to maximise recall.</li> <li>• Evaluate the new method against baseline RF performance with respect to interaction sensitivity and run time, using both real and synthetic GWAS data.</li> <li>• Integrate the improved algorithm into the VariantSpark random forest implementation.</li> </ul>
<b>Relevant field/s of study</b>	<ul style="list-style-type: none"> <li>• Machine Learning</li> <li>• Distributed Computing</li> <li>• Programming (Java and Python)</li> </ul>
<b>Supervisor</b>	Brendan Hosking
<b>Contact Details</b>	<a href="mailto:brendan.hosking@csiro.au">brendan.hosking@csiro.au</a>
<b>Location</b>	Remote or in person in Sydney

<b>Project Title</b>	<b>BSEP11 Improving Flu and Dengue Analysis with Detailed Climate Data</b>
<b>Brief description of the project highlighting expected outcomes</b>	<p>The relationship between climate variables and the incidence of diseases such as flu and dengue is a critical area of research. Accurate and localized climate data can significantly enhance the quality of analysis and improve our understanding of how environmental and natural factors influence these diseases. This proposal aims to refine the current methodology by utilizing precise climate data for specific locations and dates associated with flu and dengue sequences.</p> <p>Data Extraction: Collect environmental factors such as minimum, maximum, and average temperatures, as well as average precipitation for the exact locations (latitude and longitude) and dates corresponding to flu and dengue sequences.</p> <p>Data Analysis: Apply machine learning techniques to analyse the refined climate data, investigating potential links between changing climate conditions and emerging mutations in influenza and dengue viruses.</p>
<b>Duties/Tasks</b>	<p>The student will perform</p> <ul style="list-style-type: none"> <li>• Gathering and organizing accurate climate data.</li> <li>• Combining climate data with disease data to create a complete dataset.</li> <li>• Using machine learning models to study the data, improving analysis and problem-solving skills.</li> </ul>
<b>Relevant field/s of study</b>	<p>Machine learning Bioinformatics (alignment) Programming (python)</p>
<b>Supervisor</b>	<p>Nehleh Kargarfard Laurence Wilson</p>
<b>Contact Details</b>	<p><a href="mailto:nehleh.kargarfard@csiro.au">nehleh.kargarfard@csiro.au</a> <a href="mailto:laurence.wilson@csiro.au">laurence.wilson@csiro.au</a></p>
<b>Location</b>	<p>Remote or in person in Sydney</p>



<b>Project Title</b>	<b>BSEP12 Developing bioinformatic workflow pipeline</b>
<b>Brief description of the project highlighting expected outcomes</b>	<p>Workflow pipelines are widely used in the bioinformatics world to streamline and automate the execution of various tools.</p> <p>In this project, you will work with a range of bioinformatics software, including traditional machine learning models and modern large language models (LLMs). Your task will involve assessing the resource usage, as well as the input and output requirements for each tool. Subsequently, you will develop an automated pipeline to integrate these tools using a domain-specific language (DSL) like Nextflow.</p> <p>You will deploy this pipeline on our in-house HPC and Kubernetes cluster, with opportunities to practice deployment on AWS cloud.</p>
<b>Duties/Tasks</b>	<p>The student will perform:</p> <ul style="list-style-type: none"> <li>• Review of existing bioinformatics tools and literature</li> <li>• Wrap and deploy software in containers</li> <li>• Develop automated pipelines using DSL</li> <li>• Deploy pipelines on HPC, Kubernetes clusters, or AWS</li> </ul>
<b>Relevant field/s of study</b>	<ul style="list-style-type: none"> <li>• HPC, Cloud, docker, kubernetes</li> <li>• Bioinformatics</li> <li>• Programming (python, java, nextflow)</li> </ul>
<b>Supervisor</b>	Qinying Xu (Christina)
<b>Contact Details</b>	Qinying.xu@csiro.au
<b>Location</b>	Remote or in person in Brisbane



**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](mailto:csiroenquiries@csiro.au)  
[csiro.au](http://csiro.au)

**For further information**

**Health & Biosecurity**  
Denis Bauer  
+61 2 9325 3174  
[Denis.Bauer@csiro.au](mailto:Denis.Bauer@csiro.au)  
[csiro.au/H&B](http://csiro.au/H&B)

Image copyright Unsplash.Midjourney, GPT4