

# LIC Summer Intern Project 2023/24



## **Project Title: Algorithms for contamination identification and inter-sample comparison in bovine microbiome data**

### **Briefly overview the scope of the R&D project the student will be contributing to.**

With the advances in DNA sequencing technologies over the past 10 years, the comprehensive identification of the microbes present in samples is now possible. A number of challenges exist in the data-analysis space. Samples can be contaminated by bacteria in the lab, and algorithms for identifying that this has happened and identifying and removing contamination from the data are important. The comparison of samples with each other, each containing large numbers of microbial species, is also an area that needs further research.

We have generated a rich dataset with DNA identified from over 4000 species in over 8000 samples. The student's project will be to: a) research contamination detection algorithms from the literature and examine their usefulness on our data, as well as tailoring them to apply to our data's specific needs, b) research sample comparison methodologies from the literature and examine their usefulness on our data.

Skill sets required: Statistics. Some computer programming ability needed, R & Bash preferable. Knowledge about microbes not required.

### **Briefly demonstrate how the student will contribute to the project**

The student will be working primarily with both Bash and the R software environment to analyse the data. The student will be doing extensive reading of published scientific research on the subject and will be evaluating the pros and cons of different approaches. There may be some algorithm development by the student as they tailor the literature approaches to our specific problems. There may be some use by the student of open-source bioinformatics software on the supercomputer to process the genetic data further.

### **How will LIC benefit from employing the student**

The benefits to LIC of completing this project are a) getting data analysis done on a significant set of data we have accumulated, b) contribute to gaining a more comprehensive understanding of the bacterial species involved in the NZ dairy industry and their sources in milk production, c) contribute to a better understanding of how these bacteria are interacting

with cow health, d) the ability to identify potential new employees or candidates for LIC's postgraduate scholarship programmes.

**Mentor name and job title: Dr Andrew Wallace Scientist (Bioinformatics)**  
**[andrew.wallace@lic.co.nz](mailto:andrew.wallace@lic.co.nz)**