

# Introduction to Bioinformatics

## Part 1 and 2: Introduction to Genomic data handling

Next Generation Sequencing (NGS) data is now a mainstay of biology and bioinformatic data analysis skills are in high demand. **The first session** is designed to teach the fundamentals of computational biology to complete beginners. Attendees will learn how to navigate the command line and manipulate large files with Unix. **The second session** is designed to give hands on experience with real genomic data. Attendees will learn how to perform fundamental analyses on next generation sequencing (NGS) genomic data such as determining the genetic structure of populations (techniques such as PCAs and phylogenetic trees) and searching for adaptive genetic variation (techniques such as GWAS and PCadapt).

## Part 3: Introduction to Plant Microbiome Bioinformatic Analysis

Plant-associated microorganisms are considered key drivers for plant health, productivity, community composition, and ecosystem functioning. High-throughput sequencing of marker gene amplicons is typically used to elucidate the composition, organization, and spatial distribution of microbial communities in the environment and is increasingly used in plant microbiome studies. This course is designed to teach the basics of **targeted amplicon data processing and analysis** using DADA2 and Phyloseq packages in R (and Rstudio). Attendees will learn how to import data, perform a quality control of the sequencing data (16S or ITS), trim sequences, infer the Amplicon Sequence Variants (ASVs), assign taxonomy, and conduct basic analyses including diversity metrics, abundance and composition comparisons, and co-occurrence networks.