# Bioinformatic Summer School - Comparative Genomics, Nanopore Metagenomics and Metataxonomic Analysis

The overall teaching aim of this module is to create competence in the students allowing them to independently perform bioinformatic analysis using the command line. This includes comparative analysis on bacterial (meta)-genomes and metataxonomic analysis.

**Specifically, we will:**

For comparative genomics:

* Download all genomes ofa bacterial species.
* Annotate the genetic content of these genomes
* Profile the prophage content of these genomes
* Profile the secondary metabolism of these genomes
* Build a whole-genome phylogeny of the genomes
* Infer the linkage between data-driven analysis and fundamental biology

For metagenomics

* Download metagenomic data
* Filter long read nanopore sequences
* Assemble nanopore sequences into metagenomes
* Analyze metagenomic sequences using methods from comparative genomics

For metataxonomy:

* Download 16S data
* Demultiplex the data
* Clean, filter and denoise the data
* Taxonomically classify the data
* Use statistics to analyze alpha-diversity
* Use statistics to analyze beta-diversity
* Discuss the data in the context of relevant biology

The successful student will

1. Have a working knowledge of the Linux command line
2. Be able to use suggested command line tools
3. Investigate the bioinformatic literature to find further tools relevant for the biological questions
4. Infer the biological relevance of the genetic content in bacteria
5. Infer the composition of microbiomes
6. Discuss the biological implications of phylogeny, genetic content, microbial composition and the relatedness of all the above

# Course content:

## Day 1:

Introduction to the command line, the conda package manager and the bash scripting language. Downloading and annotating.

9:00 – 9:15 Introduction and course overview

9:15 – 10:00: Interactive lecture [**Le1**]: setting up the command line (prior Linux install is expected)

10:00 – 12:00: Interactivate Linux exercises – navigating, copying and loops

12:00 – 13:00: Lunch

13:00 – 14:00: Lecture [**Le2**]: Basic genome assembly and what we use it for.

14:00 – 15:00: Genome download & overview – how & what software?

15:00 – 17:00: Genome annotation – new environments and our first loop

## Day 2:

Finding the functional gene content and building a phylogeny

9:00 – 9:30: Recap and follow up [**Le3**] – the hard parts of bioinformatics?

9:30 – 10:00: Lecture [**Le4**] – annotation of phages and BGCs

10:00 – 11:00: Prophage annotation – another loop, another package

11:00 – 12:00: BGC annotation – yet another loop and a different package

12:00 – 13:00: Lunch

13:00 – 14:00: Phylogeny, Trees and Genotypes [**Le5**]

14:00 – 16:00: Building a core genome phylogenetic tree and finding patterns

## Day 3:

River and presentations?

## Day 4:

Long-read metagenomics: filtering and assembling metagenomic data

9:00 – 9:30: Lecture [**Le6**] – Nanopore sequencing

9:30 – 10:00: Lecture [**Le7**] – Nanopore metagenomics

10:00 – 10:30: Downloading data

11:00 – 12:00: Filtering and assembling metagenomes

12:00 – 13:00: Lunch

13:00 – 16:00: Free play with what you learned from the single genomes!

## Day 5:

Introduction to R and Rstudio, installing R packages. The concepts of metataxonomics and the DADA2 pipeline

9:00 – 9:30 Interactive lecture [**Le8**]: setting up Rstudio projects

9:30 – 9:45: R basics [**Le9**]: Very basic R and starting the exercises

9:45 – 10:30: Lecture [**Le10**]: Metataxonomics as a concept.

10:30 – 12:00: Metataxonomic exercises

12:00 – 13:00: Lunch

13:00 – 13:15: Lecture [**Le11]** Recap for metataxonomics

13:15 – 14:00: Lecture [**Le12]** Metataxonomic Processing

14:00 – 16:00: Metataxonomic exercises

17:00 **DONE**