MMB-117

Amplicon sequence data analysis

Practicals

- Every day Mon–Fri from 10 to 16 in Bio1, 3008
- Course materials:
  https://github.com/karkman/MMB-117_EnvironmentalMicrobiology/
- All computation will be done in CSC Puhti HPC cluster and web interface can be used
  - Computing node CLI for pre-processing
  - RStudio for the rest

Learning goals

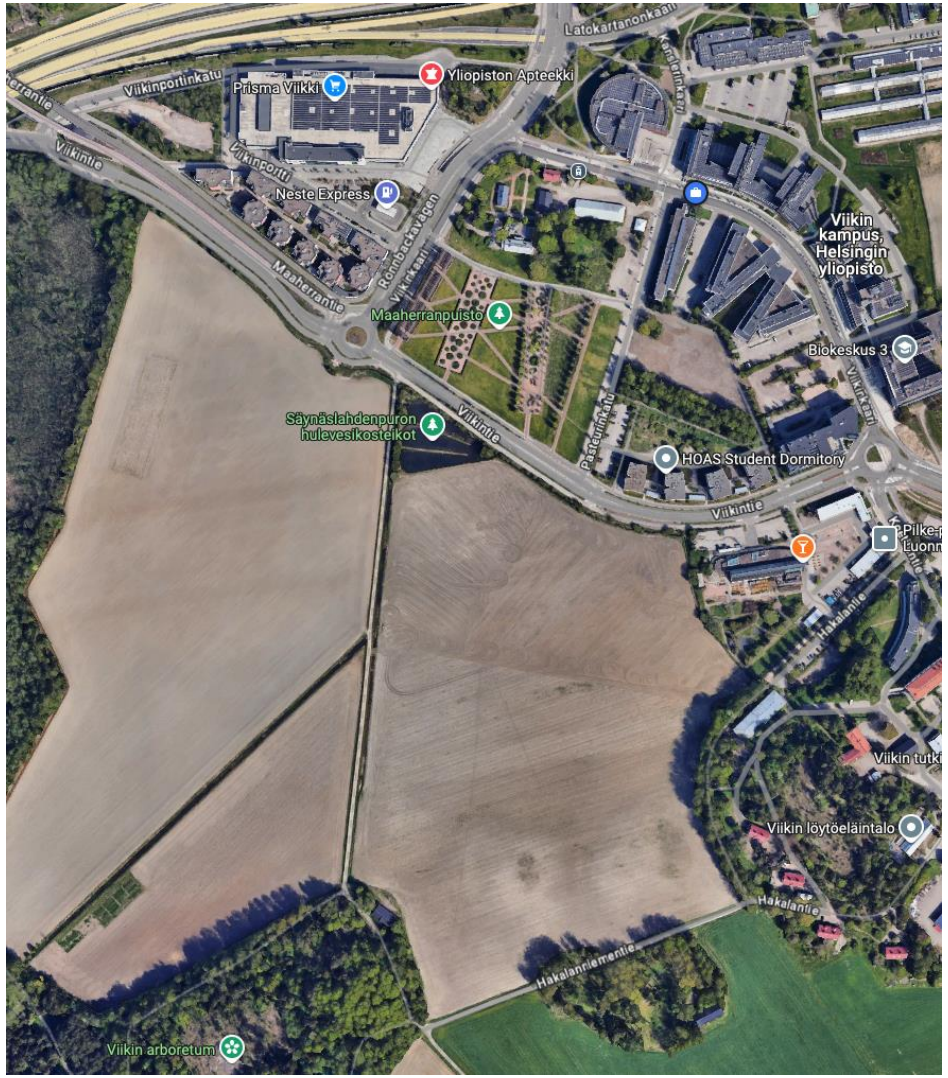
- Understand the basics of amplicon sequencing and bioinformatic approaches to analyse amplicon sequencing data
- Be able to plan an amplicon sequencing project and choose the right tools and approaches to answer your specific research question
- Have confidence to learn new methods needed to answer your research question in microbial ecology
- Empower you to ask and answer the questions you have on your own data

Workflow

Wet lab

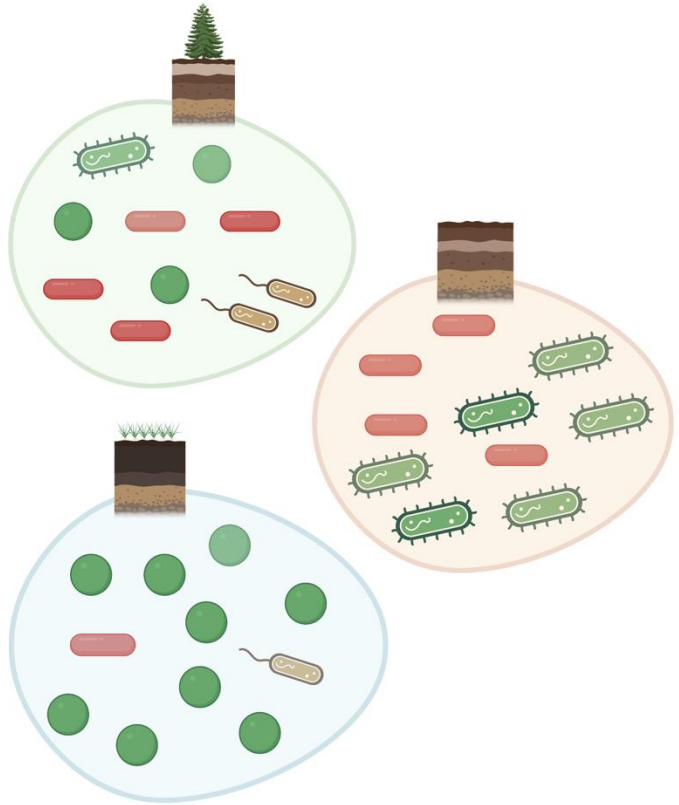
- Sampling: 4 sites x 6 replicates
  - Gas station, Park, Field, Forest

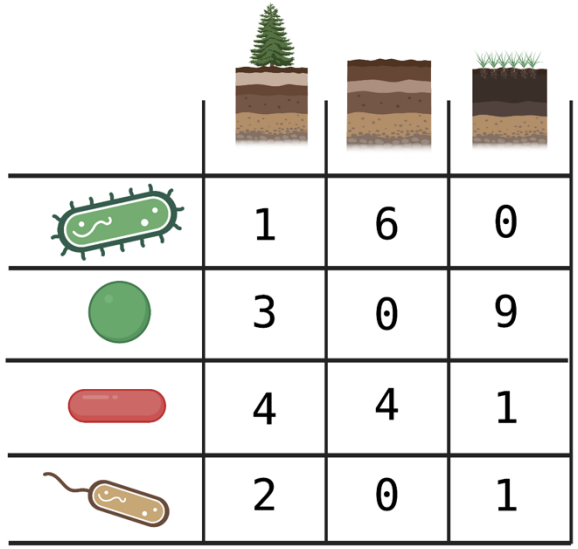
- Lab analyses:
  - Weight (wet/dry), Moisture, pH, SOM, plate counts
  - DNA extraction & 16S rRNA gene amplification



Dry lab – next steps

- Pre-processing:
  - Raw data quality control
  - Primer removal
  - Trimmed data quality control
- DADA2 pipeline:
  - Quality trimming
  - Denoising
  - Chimera removal
  - Taxonomic annotation

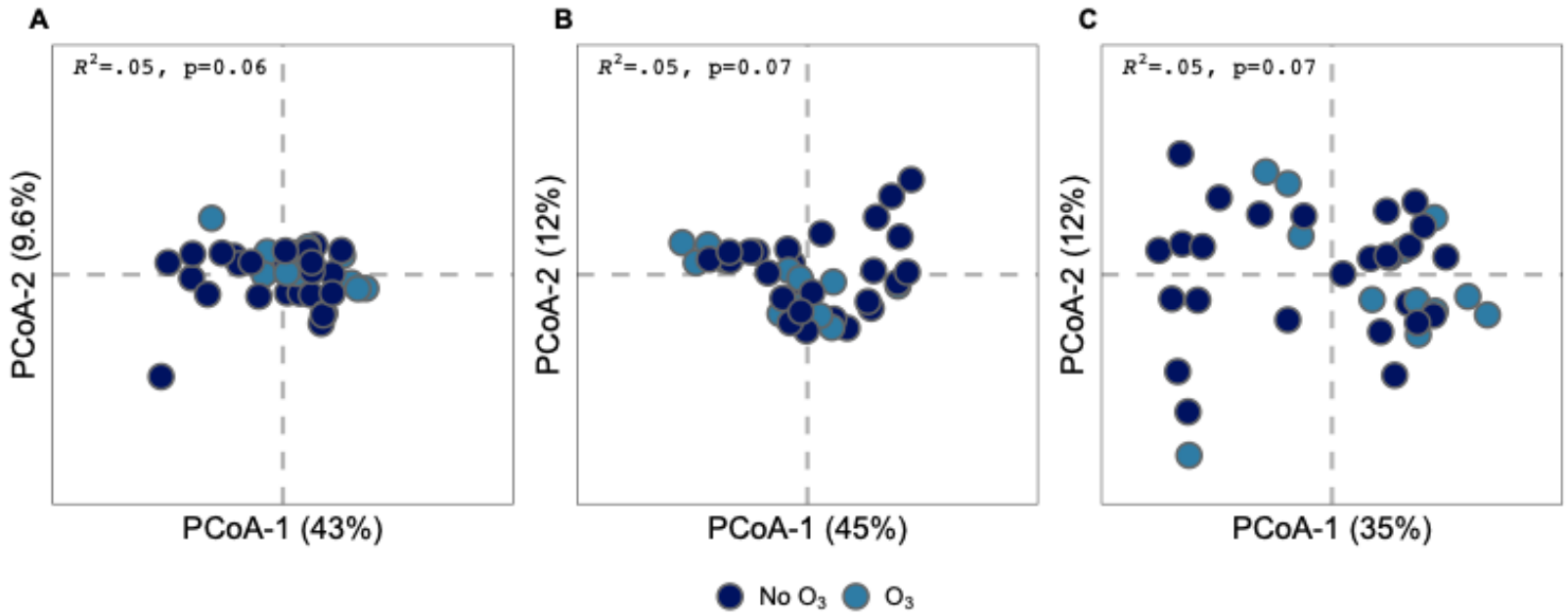




Statistics

- Data exploration

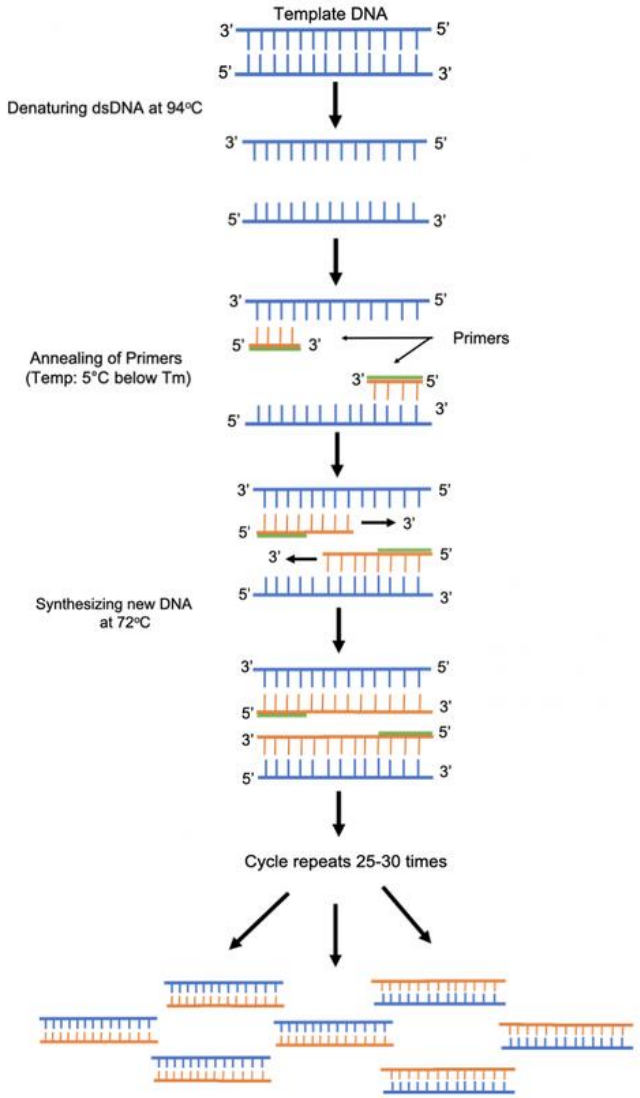
- Statistical analyses to answer the research questions

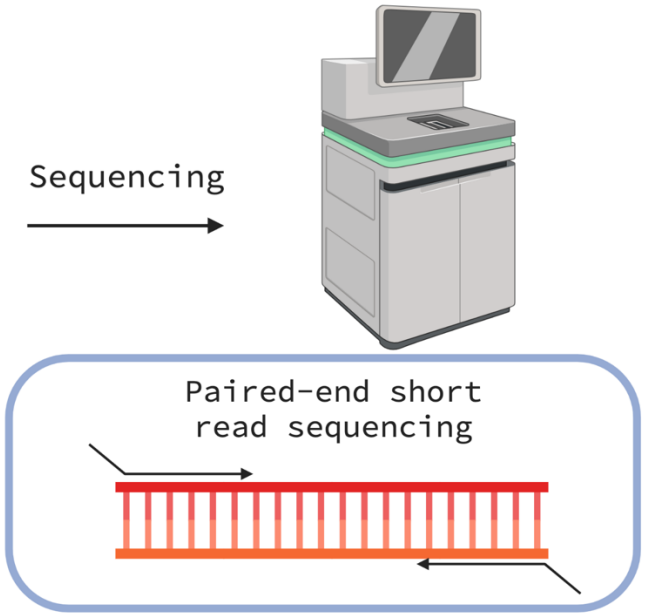


Research questions?

Amplicon sequencing

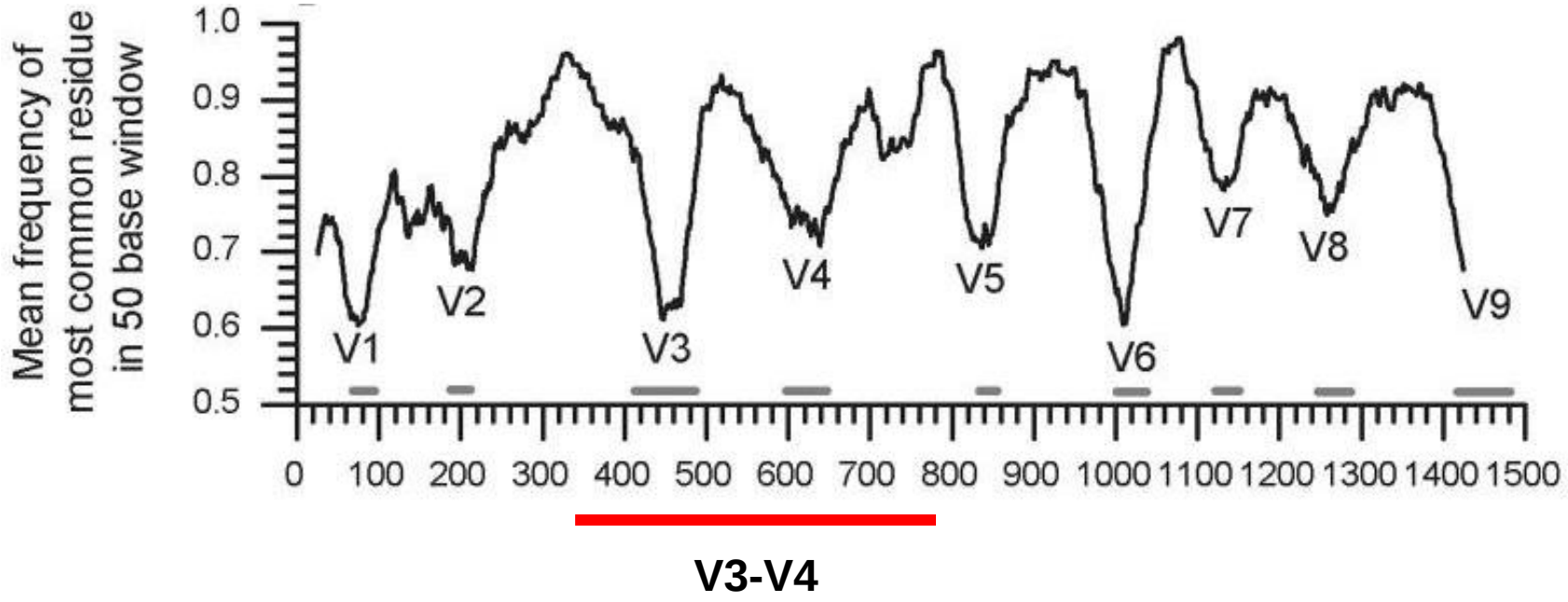
Amplicon sequencing





Amplicon sequencing

- What primers did we use? What regions have we amplified?



V3-V4