Summary metagenomics Workshops

**Monday 25th of October**

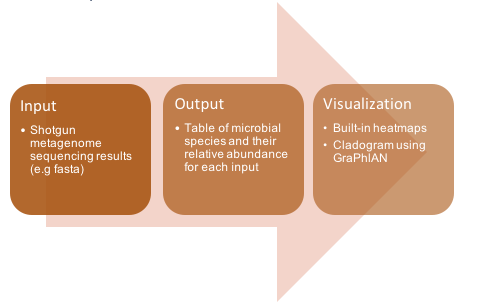
**What we did?**

* Discussed setting up Rosalind; which is High Performance Computing (HPC). HPC gives us the ability to combine computing power in a way that delivers much higher performance than one could get out of a typical desktop computer or workstation in order to solve large problems in science and in this case metagenomics analysis.
* Discussed the project report
* Installed Metahplan in Rosalind
* INPUT: shotgun metagenomics files (.fasta as extension) and as an OUTPUT we got the table of microbial species and their relative abundance which we downloaded to our local computer.
* Installation of SRA toolkit

**Key point:**

**What is Metaphlan and why we use it for microbiome analysis?**

Next-generation sequencing technologies have allowed for sequencing at a low cost and fast speed, and is used more and more to study microbial communities. Since a variety of microbes live in the microbial community at differential relative abundances, the first question researchers usually ask is who is present and at what relative abundance. A frequent used tool is MetaPhlan (http://huttenhower.sph.harvard.edu/metaphlan2), which uses clade-specific marker genes to study the microbiome taxonomic composition.



**What is SRA?**

The Sequence Read Archive (SRA) is a publicly available repository of sequence data and just one of the many databases hosted by the NIH’s National Center for Biotechnology Information (NCBI). The SRA archive consists of raw sequences from various NGS platforms and, more recently, alignment data. Submitting sequences to the database allows researchers to contribute to the reproducibility efforts of the broader community. SRA reads are typically pulled for increasing sample sizes and validating experimental results, but as a scientist you can use the database to explore new research questions.

**Tuesday 26th of October:**

**Key points:**

* **Understand what phyloseq does**
* **Understand what the script in R does**

Why phyloseq?

To explore microbiome profiles using R. The phyloseq package is a tool to import, store, analyze, and graphically display complex phylogenetic sequencing data that has already been clustered into Operational Taxonomic Units (OTUs), especially when there is associated sample data, phylogenetic tree, and/or taxonomic assignment of the OTUs.

**What we did?**

* Installation of different libraries in R which we will need to see which species are present in our samples
* We used phyloseq and made a phyloseq object

**Wednesday 27th of October**

**Key points:**

* **Understand how statistics can be used for microbiome analysis**
* **Understand what alpha and beta diversity entails**
* **Understand the code used in R to perform the statistics**

**What we did?**

* In this lecture we dive deeper into our data and introduced statistics. We look into alpha and beta diversity and with help of R we calculate this. Also we stand still by what a normal distribution is (basis of statistics) and which tests can be used to compare groups such as Wilcox test and/or student t-test.
* Alpha diversity: within sample
* Richness: what are there? 🡪 chao1 index and ACE index , shannon diversity, simpson index
* Evenness: how much is there?
* Beta diversity: measurement to characterize the diversity between two samples.