Genome Analysis

General notes taken from lectures, seminars and labs.

Content

[Grading 2](#_Toc35889387)

[Recap 4](#_Toc35889388)

# Grading

We will receive an individual grade for the lab project.

To get a 4 as grade I need to do at least two of the following:

* Lab manual questions: you’ve answered the questions in the lab manual, and you’ve done so correctly.
* GitHub wiki and presentation: your wiki and final presentation are well-structured and display a clear understanding of the methods used and the meanings of the results.
* Extra analysis: you’ve preformed at least one extra analysis and extensively discussed the result.

For a 5 you need to fulfill all above, and do multiple extra analysis.

**Article 4**

Basic analyses:

* Metagenome assembly.
* Binning.
* Quality check of assembly and bins.
* Functional annotation
* Basic phylogenetic placement of bins (Taxonomic ID).
* Reads preprocessing: trimming + quality check (before and after)
* Analyses of activity (expression level) of different bins.

Extra analyses:

* Abundance of different organisms/ bins.
* Refine taxonomic ID of assembled genomes.
* Metabolic pathway reconstructions for chosen bins
* Analysis of expression data of chosen gene groups (i.e: respiratory genes, genes involved in carbohydrate metabolism, etc).
* Comparisons across bins (pathways, expression certain genes groups, etc).
* Comparative genomics of bins
* Orthologue gene clustering of bins

**To keep in mind**

Don’t use raw data directly in the analysis because quality might not be optimal. First: assess the quality of your data first. Perhaps remove reads with low quality or that adapters is still present in the reads. FastQC: check quality of short reads.

You want to preprocess reads by remove the low quality base-calls and present adapters from the reads. Trimmomatic: can be used to trim and crop illumina (FASTQ) data as well as remove adapters.

Multiple softwares for genome assembly.

Assembly evaluation: QUAST and MUMmerplot.

Binning: after assembling reads coming from metagenomes we end up with a mix of contigs that come from different organisms. Binning is a process of classifying those contigs into different organisms to reconstruct their genome. Metabat software can be used for this. After binning it’s important to evaluate the recovered bins. We want statistics for each bin such as size, GC content, verifying if each bin represents a complete genome or only a part of it. CheckM provides a set of tools for assessing this.

Annotation: after assembling genome we want to see what genetic element it encodes.

Homology search: wanting to see how similar organisms are. Blastn.

Phylogenetic placements: metagenomes might contain a large microbial diversity. If we want to study organism under evolutionary context, phylogenies are useful. Phylophlan.

Mapping: aligning reads against a reference to determine reads coverage or to improve/correct assembly, account for differential expression of certain genes. BWA and Tophat.

**Deadlines**

* 30/3: Project plan
* 17/4: Assembly
* 23/4: Binning
* 29/4: Bin quality assessment + annotation
* 5/5: RNA mapping
* 8/5: Phylogeny

# Recap

**Taxonomy** = to sort organisms into groups. You classify by diving biological organisms into groups with consideration to a set of predefined criteria such as morphology and genetics. Organisms are grouped together into taxa and these groups are given a taxonomic rank. The ranks used today are domain, kingdom, phylum (plural phyla), class, order, family, genus and species. A taxon is a group of one or more populations of an organism or organisms seen to form a unit. A taxon is usually known by a particular name and given a ranking.

**Clade** = monophyletic group; a group consisting of organisms with a common ancestor and all its lineal descendants. The group consists of all the descendants!

**Metabolism** = the set of life-sustaining chemical reactions in organisms. Metabolism can be divided into catabolism and anabolism. Catabolism = degrading processes which generates building blocks, energy and reduction power. Anabolism is construction of building blocks, mainly reductive, electron absorbing reactions. The chemical reactions of metabolism are organized into metabolic pathways, in which one chemical is transformed through a series of steps into another chemical, each step being facilitated by an enzyme. Enzymes are crucial to metabolisms.

Microorganisms are facultative if they can alter between different metabolic pathways (f.ex. how they receive energy).

**Metagenomics**

Metagenomics it the science that studies genes and genomes of non-cultative organisms. It’s a cultative independent set of methods used for large scale analysis of organisms. It describes what organisms are present in different environments and thereby gives us understanding of the diversity and evolution of organisms in their natural habitat. Metagenomics will answer the following questions:

* What organisms are present in the sample? What kind of bacteria?  
  You often use ribosomal genes to get some understanding of what organisms are present. You want to make PCR and can’t therefore use species-specific primers – need primers that can attach to a gene most organisms have. It’s common to use 16S/18S and amplify with PCR for phylogenetic analysis.
* What is the purpose of the organisms? What happens as a whole in the system?   
  You want to investigate what genes are present in the environment because it will tell a great deal of what functions exist.
* Which organism does what?

**Metatranscriptomic**  
Science that studies gene expression of microbes within their natural environments. It allows to obtain whole gene expression profiling of microbial communities. Metagenomics focuses on studying the genomic content and to identify which microbes are present. Metatranscriptomics is used to study the diversity of the active genes and quantify their expression levels and monitor how levels change under different conditions.