Detection of AMR Genes and Assembly of Genomes From Metagenomic Sequence Data

Authors:
Group 3: Lasse Johan Dyrbye Nielsen (laniel)[1], Stefanos Rodopoulos (s212895)[1], Mads Cort Nielsen (s120356)[1]
[1]Technical University of Denmark, Department of Health Technology

Introduction

Metagenomics enables the study of ideally a whole microbial community, including species that remain unculturable. With shotgun sequencing and metagenomic data analysis the taxonomic profile and genomic potential of a sample can be determined, as well as the recovery of whole genome sequences [1]. Here we apply software programs that process reads; assemble genomes; quality filters and de-replicates; and assign taxonomy to assembled genomes, all in a single pipeline using the workflow manager Snakemake. We use this pipeline to measure the relative abundances of bacterial species and antibiotic resistance genes in a metagenomic datasets.

Materials and methods: Metagenomic data from 3 faecal samples were obtained from the Danish VETI II project (PRJEB26961). Each sample was pooled from faecal pig material from 30 pigs from the same farm located in Denmark and sequenced using paired end shotgun sequencing. The workflow for processing of the raw data can be seen below. Only contigs longer than 1000bp were used for binning. Since no high quality bins resulted from the binning process, all medium quality bins (contamination < 10%, completeness > 50%) were kept. The dereplication threshold was set to 97% ANI (FastANI v. 1.33). To estimate the abundances the species resulting from the genome assembly, clean reads were mapped against the metagenome assembled genomes (MAGs). The relative abundance of antimicrobial resistance genes was estimated by mapping the clean reads against the RefSeq database. The data analysis of the mapping results is described in the section below.

Workflow

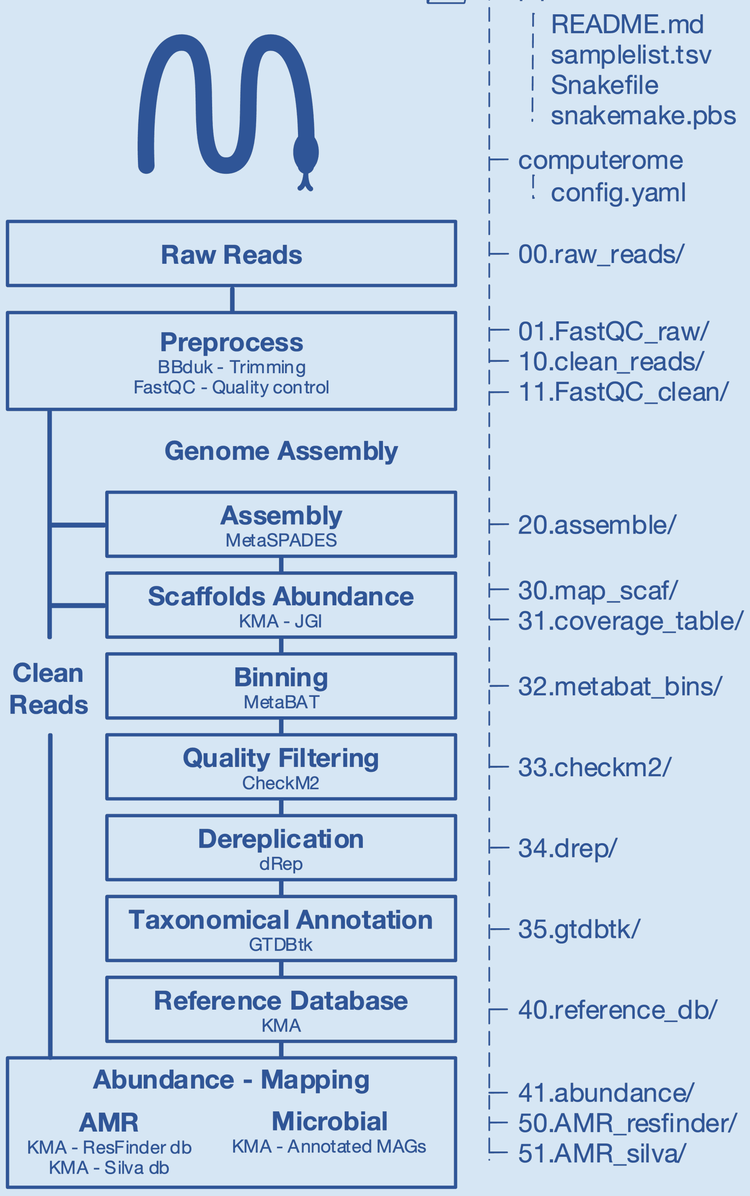

Figure 1 Workflow for the pipeline created in this project. The right side show the processes and the software used in each step. The left side show the snakemake folder structure.

Metagenome Assembled Genomes

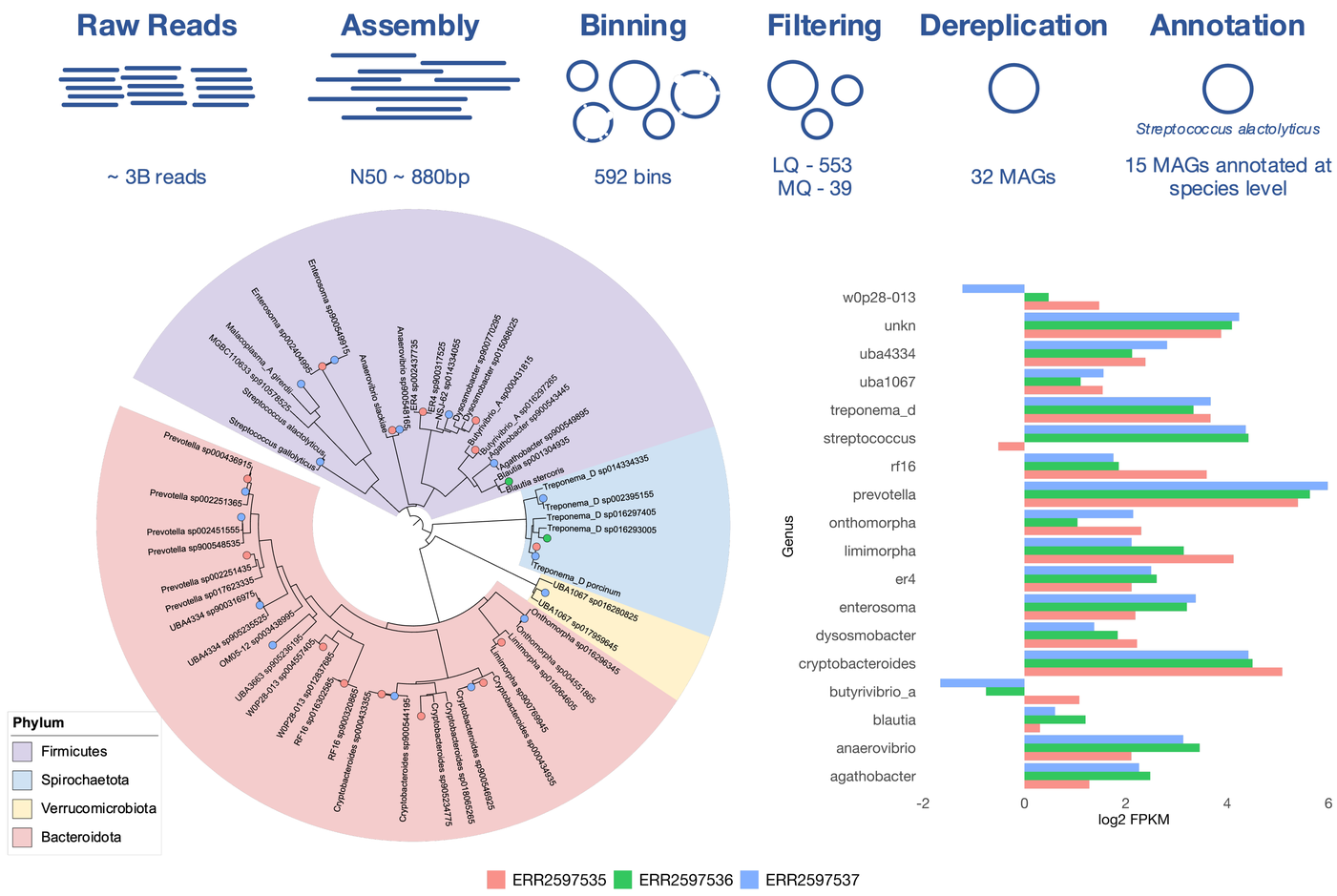

Figure 2. Phylogenetic tree resulting from taxonomical annotation of metagenome assembled genomes using GTDB-tk

Figure 3. Relative abundance of genera in three different pig faecal samples.

Phylogenetically the assembled genomes fall within the phyla Firmicutes, Bacteroidota, Spirochaetota and Verrucomicrobiota. Fifteen assembled genomes are from sample ERR2597535 or ERR2597537 respectively while only two genomes are assembled from sample ERR2597536. A total of fifteen MAGs were annotated to the species level before mapping reads from the different samples back to them.

Compositional Data Analysis

$$\log_2 \mathrm{FPKM} = \log_2 \left(\frac{\mathrm{N}_g \cdot 10^3 \cdot 10^6}{\mathrm{N}_t \cdot l_g} \right)$$

To quantify the number of reads of microbial origin, the clean reads of each sample were mapped against the Silva database. These numbers (N_t) were used as common denominator for the FPKM version of ALR transformation, see equation above. The length of each gene is used for transformation of the AMR gene relative abundance estimation and the genome size is used for the taxonomy (l_g). All data was zero-corrected using Bayesian-multiplicative treatment (zCompositions v. 1.4.0). To do between sample comparison using PCA - CLR transformation was performed (compositions v. 2.0).

Antimicrobial Resistance Genes

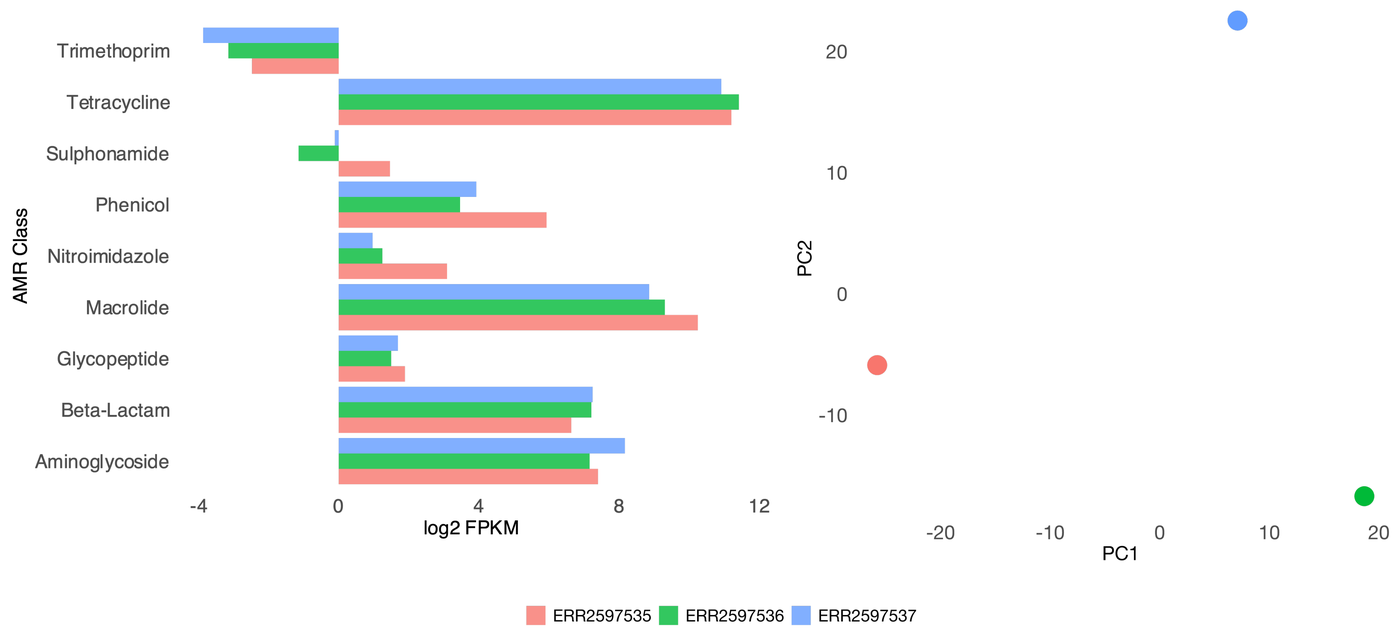

Figure 4. Relative abundance of antimicrobial resistance genes in three different pig faecal samples.

Figure 5. PCA plot of CLR transformed relative abundances of antimicrobial resistance genes.

Antibiotic resistance genes are classified according to target, resulting in ten higher AMR classes. In terms of relative abundances the rank order of categories are similar across samples. Not two of the samples cluster more with each other than with the third sample as seen by the equidistance of all samples in the PCA plot.

Discussion and Conclusion

In this metagenomic study we asses the bacterial and antimicrobial resistance diversity of three pig microbiomes, finding species from four different bacterial phyla and antibiotic resistance genes from nine different antibiotic resistance classes. 15 assembled genomes from the data are successfully taxonomically assigned and the mappings of the reads show the presence of the same species in the three samples with some variation in abundances. Similarly, representatives of all nine different antibiotic resistance genes are found within all three samples, showing some variation in abundance. We have thus successfully developed a pipeline for processing metagenomic data from raw reads to mapping, abundance estimation and taxonomy assignment.

References:
1. Quince, C., Walker, A. W., Simpson, J. T., Loman, N. J. & Segata, N. Shotgun metagenomics, from sampling to analysis. Nat Biotechnol 35, 833–844 (2017).