Antimicrobial resistance classification of *Sus scrofa domesticus* fecal samples

Applied Methods in Metagenomics - 23260
Date: 13.12.2022

Group 7: Adikrishna Murali Mohan (s212940), André Rosengaard Jørgensen (s183451), Prince Ravi Leow (s220520)

INTRODUCTION

Antimicrobials are often used in livestock to keep the population healthy. Antimicrobial consumption in pigs is a major contributor to the global antimicrobial consumption in livestock [1]. The use of antimicrobial medicine puts selective pressure on the population, leading to the emergence of resistance genes in well-adapted individuals, potentially spreading Antimicrobial Resistance (AMR) to the entire population.

Sample: 24, 25 and 38

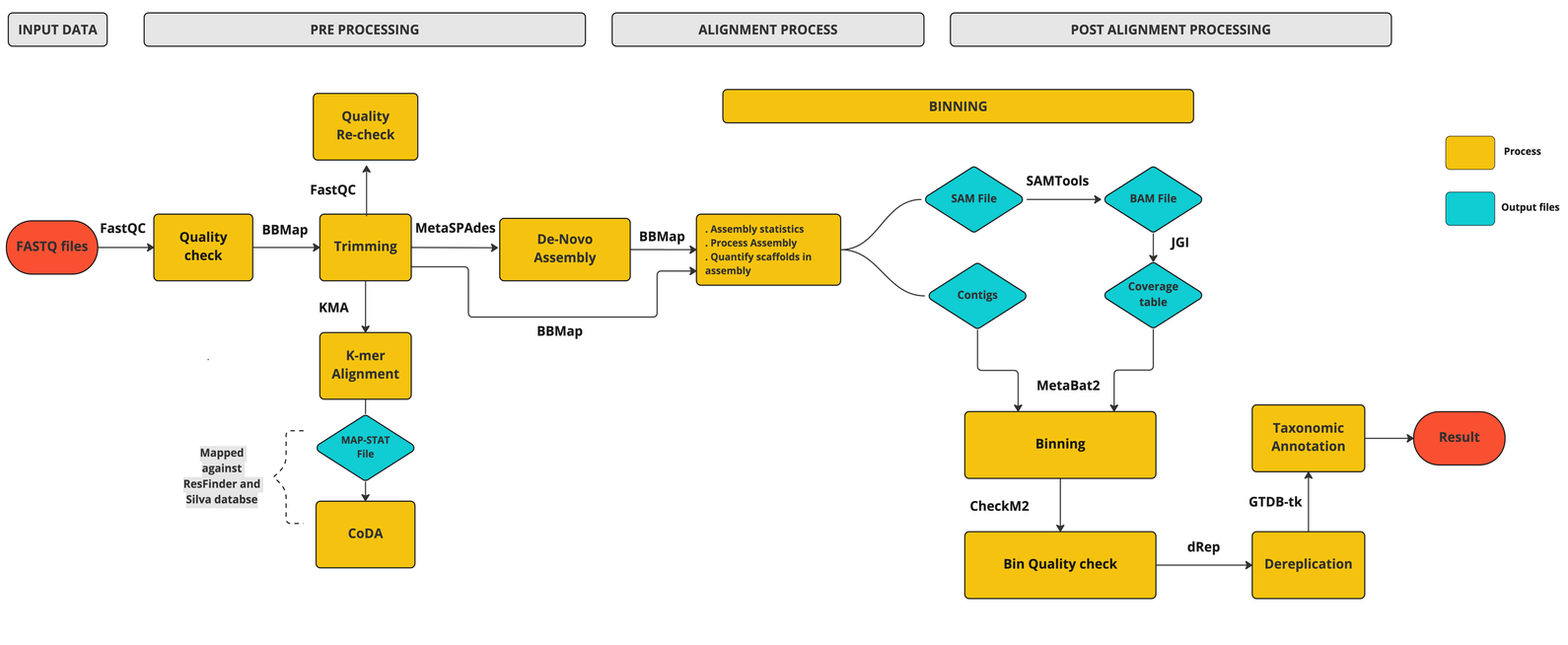
- Samples from Danish VETII project
- Collected from 83 farms from March 2014 to July 2016
- Investigated the AMR genes in the pig gut microbiome
- Sequenced by Illumina Hi-Seq
- A single fastq sample is a pool of fecal pig material from 30 pigs from the same farm located in Denmark.

OBJECTIVE

For all three samples (24, 25 and 38):

- Taxonomic identification of AMR bacterial candidates
- Identify AMR gene composition

METHOD



AMR gene composition (CoDa) workflow:

We used ALR variant, logFPKM (gene fragments per kilobase millions):

$$\log\left(\frac{\text{fragments} \times 10^3 \times 10^6}{\text{gene length in bp} \times \text{total bacterial fragments}}\right)$$

Fragment abundance and length determined from aligning KMA against ResFinder database, and total bacterial fragments from aligning against SILVA. Zeros are replaced with ones, due to ALR being a log transformation.

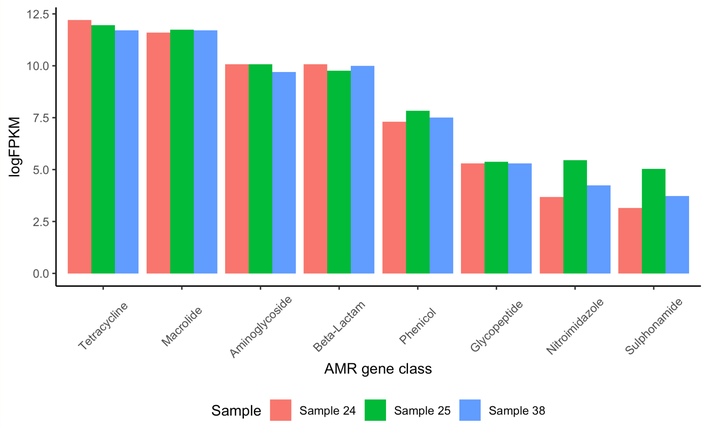
Intermediate results for metagenomic workflow:

- Read quality (FastQC): Trimming resulted in marginally better-quality scores at the ends of reads
- Assembly statistics (bbmap):
  - Sample 24: Main genome scaffold N/L50: 444,580/699
  - Sample 25: Main genome scaffold N/L50: 676,249/882
  - Sample 38: Main genome scaffold N/L50: 222,410/654
- Coverage table (Metabat2):

Sample	Largest contig size (bp)	Total average depth
24	136,384	10.8
25	291,205	12.9
38	199,456	46.6

RESULTS

COMPOSITIONAL DATA ANALYSIS



Sample	Total Fragments	Bacterial fragments	% Bacterial fragments
24	102,963	101,968	99%
25	299,937	294,141	98%
38	59,954	77,022	96%

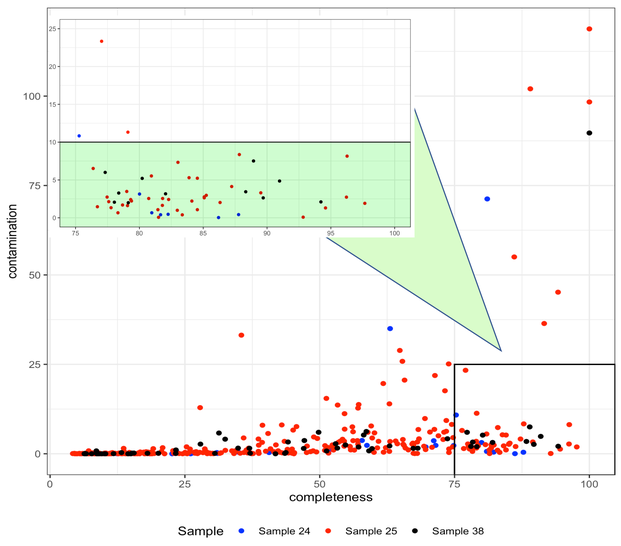
108 distinct AMR genes were determined and amalgamated into 8 gene classes. Tetracycline and macrolide are the most abundant AMR gene classes, with aminoglycoside and beta-lactam being the second most abundant AMR gene classes.

Since the ALR values are sufficiently similar, we have plotted all samples together – in spite of them technically having different closure constants (i.e., number of total bacterial fragments per sample).

Abundance values for nitroimidazole and sulphonamide in samples 24 and 38 underwent zero-replacement. These fragment lengths are relatively small, which results in respective ALR quantities being overrepresented, as a consequence of the expression for log FPKM.

CLR values were computed, but due to the similar AMR composition between samples, they have been discarded from the results.

BINNING QUALITY



As seen in the above figure, after the binning, most of the sample regions are not considered, since the cut off criteria is completeness >75% and contamination <25% – dRep's default quality criteria.

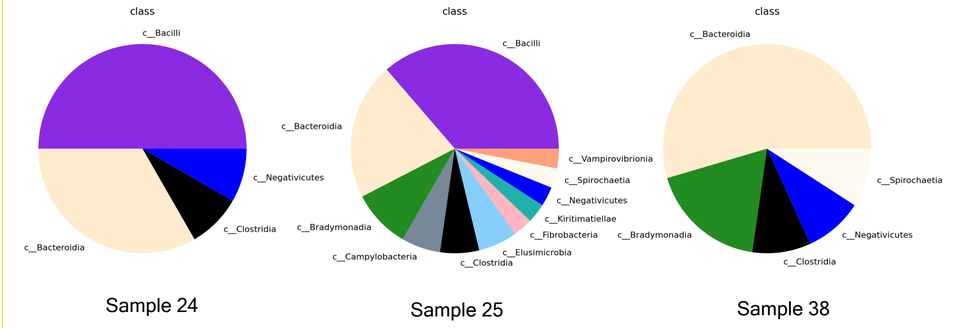
This is indicated in the above figure, where dots within the black border indicate bins that pass dRep's quality criteria. As a result, most of the bins are discarded, and therefore lost.

Due to samples 24 and 38 being significantly smaller than 25, they result in a much lower number of quality bins for the taxonomical annotation.

Sample	No. bins total	No. 'high quality' bins	No. bins discarded
24	34	7	27
25	316	39	277
38	61	11	50

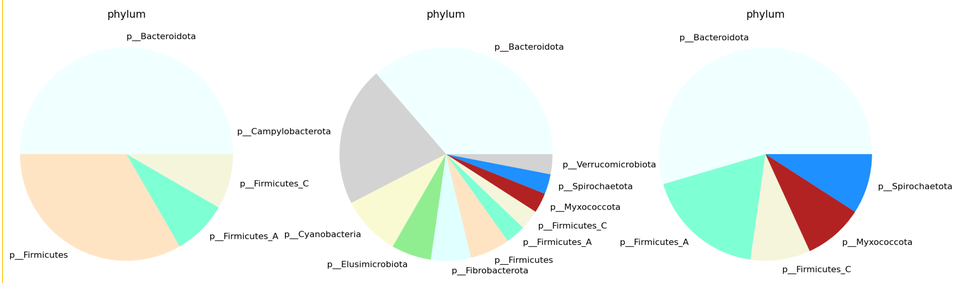
In addition, it can be seen in the magnified view, there are 3 points outside the light green border, that are clearly outliers and therefore should have been discarded.

TAXONOMIC ANNOTATION



Gtdb-tk produced annotations, split across bacterial taxonomical identifications. There were entries produced as deep as genera/species level. However, due to sample 25 having significantly more entries, and due to some entries not being able to be classified, we have chosen to compare bacterial classes between samples.

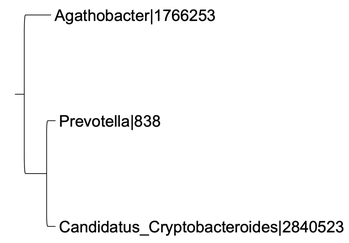
It is apparent that 38 is different from the other two samples. In 24 and 25, the majority of the bins are assigned Bacilli as class. In 38, most of the bins are instead assigned Bacteroides.



It is unsurprising that Firmicutes and Bacteroidetes are abundant in our samples as they are very common in the pig gut.

It was also noted, that all 3 samples contained the genera Agathabacter, Candidatus Cryptobacteriodes and Prevoletta.

A phylogenetic tree based on these 3 genera reveals that two of them are closely linked, but since the list is not extensive, more cannot be concluded.



CONCLUSION

The compositional data analysis of AMR gene abundance in bacteria concluded that in spite of sample size differences, the samples contain similar levels of resistance genes. Similarly, the taxonomic analysis also showed mostly similar taxa in each of the samples, although one sample (38) did contain a slightly different taxonomical binning classes.

Future studies: in order to investigate the effect of sample size on binning quality, a more extensive, and granular array of sample sizes should be investigated. In addition, bacterial diversity studies, could elucidate links between antibiotic use and AM-resistance.

REFERENCE:

1) Van Boeckel, Thomas P., et al. "Global trends in antimicrobial use in food animals." *Proceedings of the National Academy of Sciences* 112.18 (2015): 5649-5654.
2) Birkegård, A.C., Halasa, T., Græsbøll, K. *et al.* Association between selected antimicrobial resistance genes and antimicrobial exposure in Danish pig farms. *Sci Rep* **7**, 9683 (2017). https://doi.org/10.1038/s41598-017-10092-9