

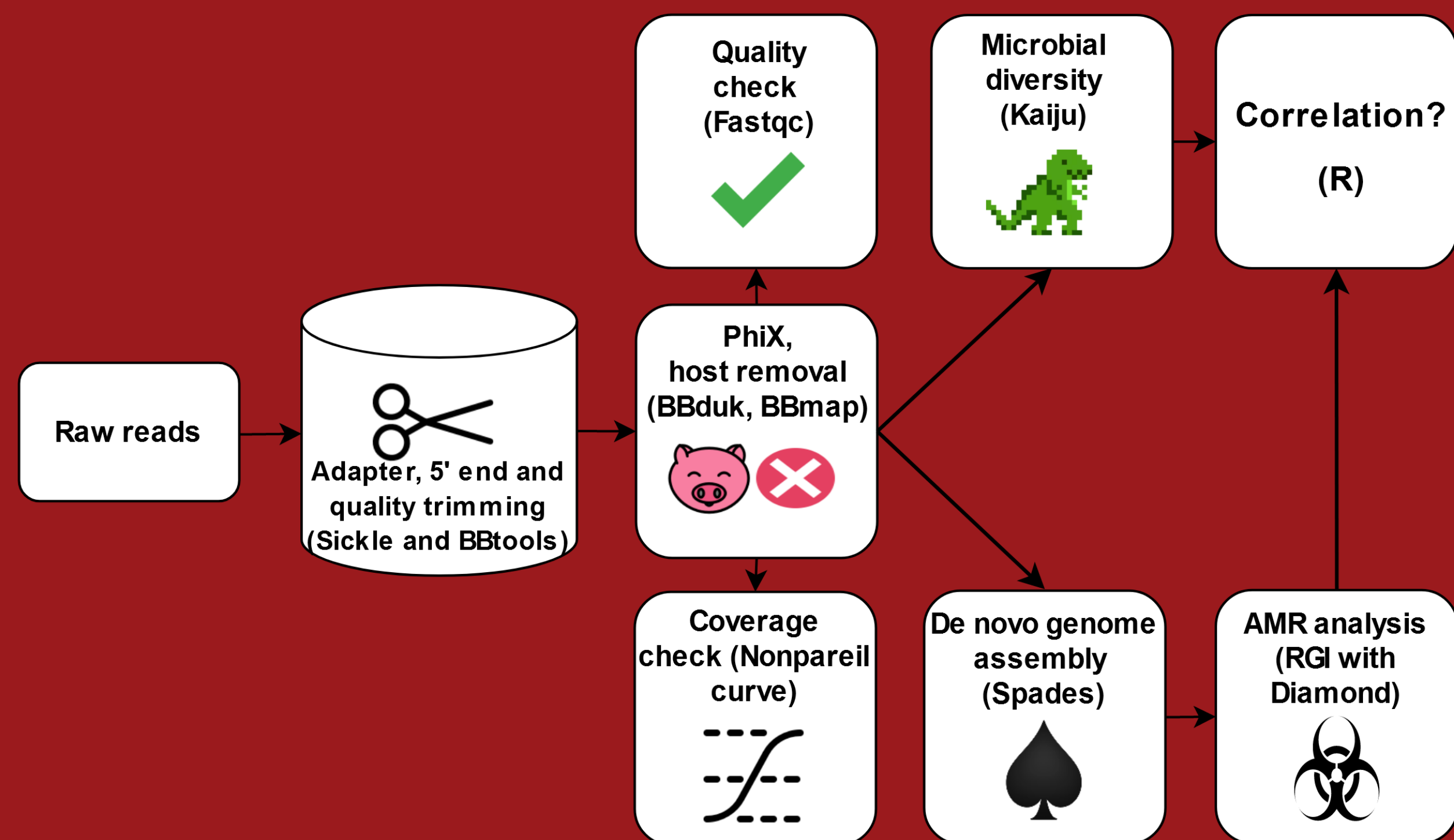
Introduction

In recent years interest in microbiomes has expanded with research into not only the human microbiome, but also other organisms such as livestock. Pigs are an important agricultural livestock that makes up a large part of the world's overall meat consumption, and the influence of the microbiome on factors such as nutrient absorption, disease and general health are of interest to the industry.

The practice of giving antibiotics to livestock to combat pathogens and promote growth is known to select for antibiotic resistance genes in the microbiome, and thus the presence of these genes can be utilized as a proxy for the antibiotic pressure on the pig [1].

In this project we wish to investigate the correlation between the abundance of antibiotic resistance genes (ARG) and the general microbiome diversity, by comparing stool samples from domestic farm pigs and wild boars.

Workflow



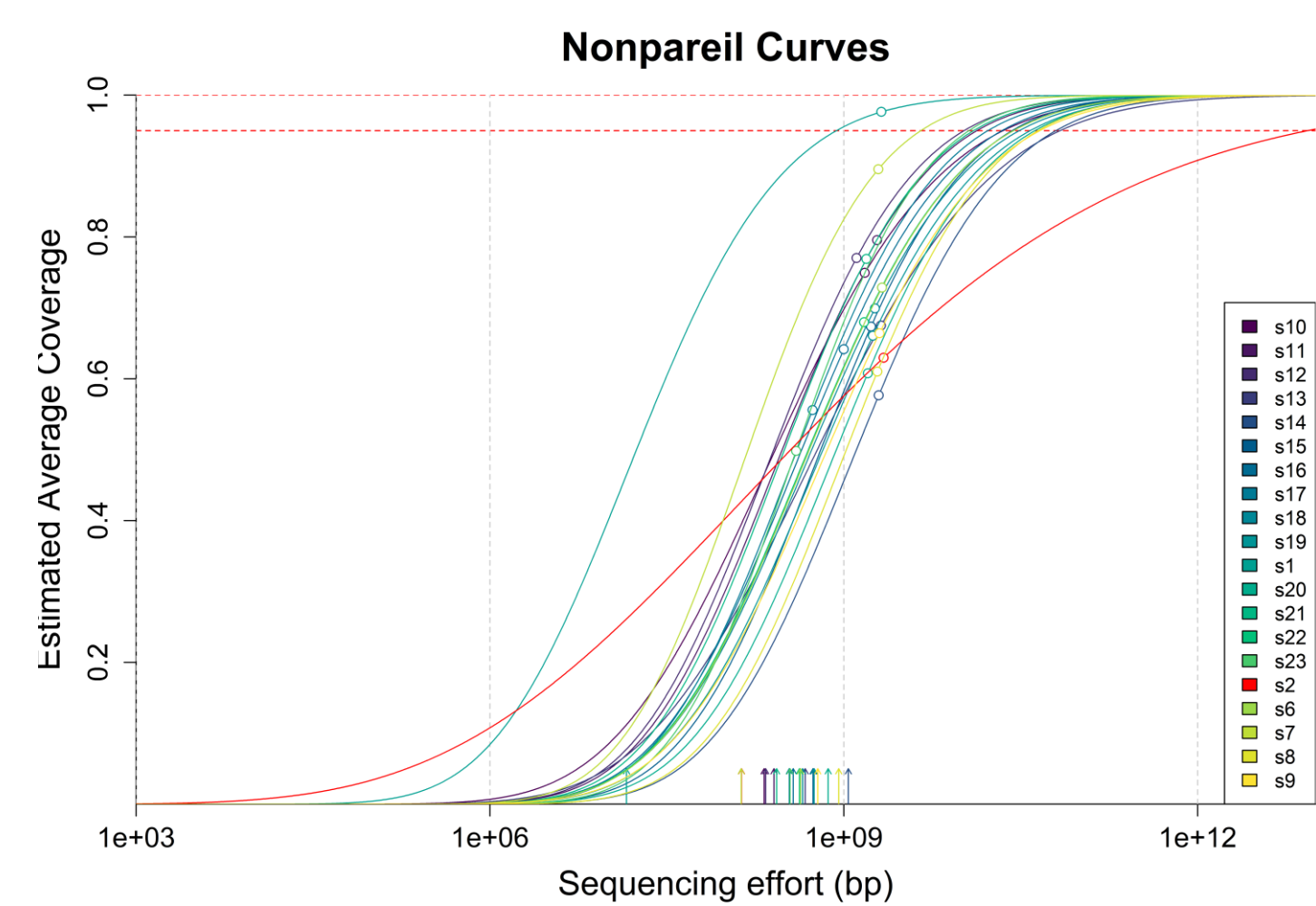
Sample processing: 20 pig stool samples were obtained from the European Union-funded EFFORT project and sequenced using an Illumina sequencer. Of these samples, 6 were obtained from Polish wild pigs and 14 from French domesticated pigs [2].

Preprocessing: Adapter and 5' end trimming was done using BBTools and quality trimming was done using Sickles [3][4]. PhiX and host DNA removal was done using BBDuk and BBDuk. Coverage and read quality were assessed from Nonpareil curves and FastQC reports respectively [5][6].

Analysis: De novo genome assembly was done for all 20 samples using Spades with K-mer sizes (21, 33, 55) [7]. For each sample, microbial diversity and abundance was determined using Kaiju and the NCBI non-redundant protein database (eukaryotic proteins included) [8]. We also identified and quantified AMR genes in the samples using RGI coupled with Prodigal, DIAMOND and Bowtie2 against the CARD AMR gene database [9] [10] [11] [12][13]. Lastly, statistical analysis was performed in R.

Quality and coverage check

Sequencing quality prior and post preprocessing was assessed using FastQC. We also checked sequencing coverage for all samples with Nonpareil curves where the estimated average coverages ranged from, 0.61 - 0.93 (see right figure)



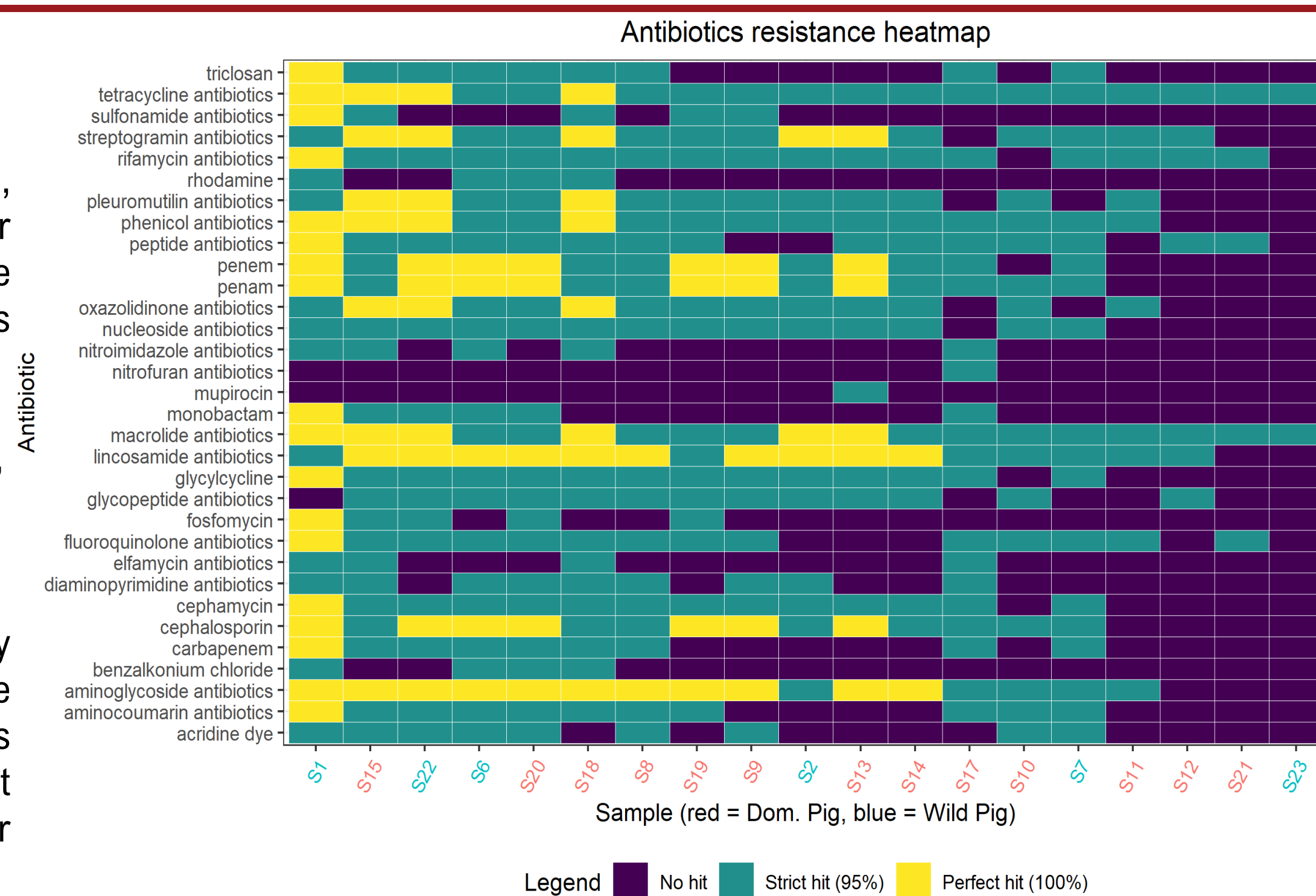
From these results, we determined that sequence quality was satisfactory for further analysis.

ARG analysis

Contigs generated by Spades were used as input for RGI, to determine a resistome for 19 of the 20 samples (genome assembly for sample 16 was not completed).

RGI first predicts ORF's and putative proteins with Prodigal, which are then subsequently mapped to CARD using DIAMOND.

A heatmap was generated by grouping antibiotic resistance genes according to drug class and reporting the highest degree of alignment match for each class.

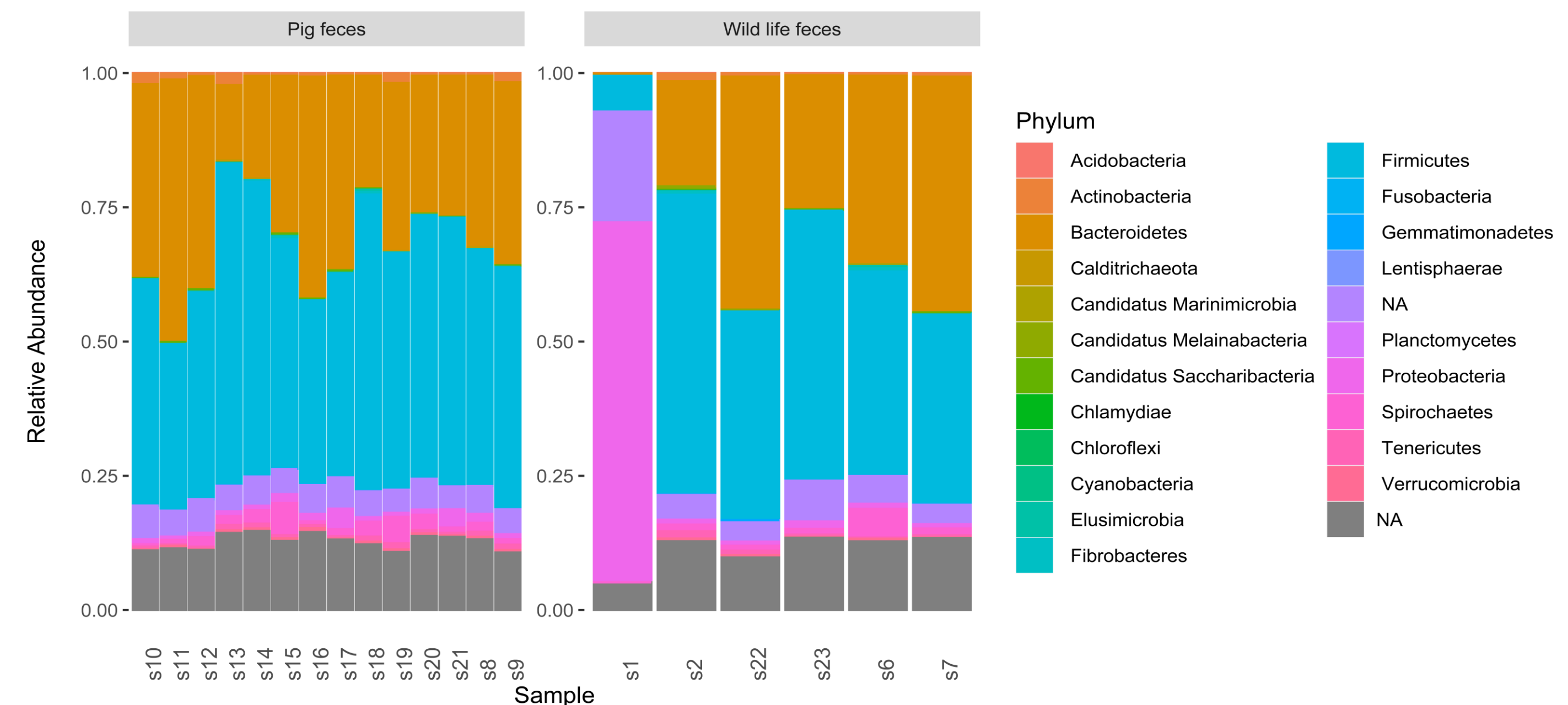


From the heatmap, two resistome clusters consisting of samples (s1, s15, s22, s6, s20, s18, s8, s19, s9, s2, s13 and s14) and (s17, s10, s7, s11, s12, s21 and s23) were detected. Note that these clusters interestingly are not distinguished by pig type but by the samples that were clustered in our OTU PCA analysis (see Microbial Diversity section). The 3 most prominent antibiotic resistance drug classes were tetracycline, macrolide and lincosamide antibiotics, all classes commonly used in the agriculture industry which are known to promote antimicrobial resistance [16][17]. Stool sample s1 had the most diverse AR drug class hits, which is interesting as our microbial diversity study suggested that the corresponding host was in a diseased state.

For ARG quantification, Bowtie2 mapping of raw reads was done against the CARD + Variant Database. The resulting mapping counts were normalized by sample read counts and used in conjunction with our microbial diversity study to determine if there is a correlation (see Discussion section).

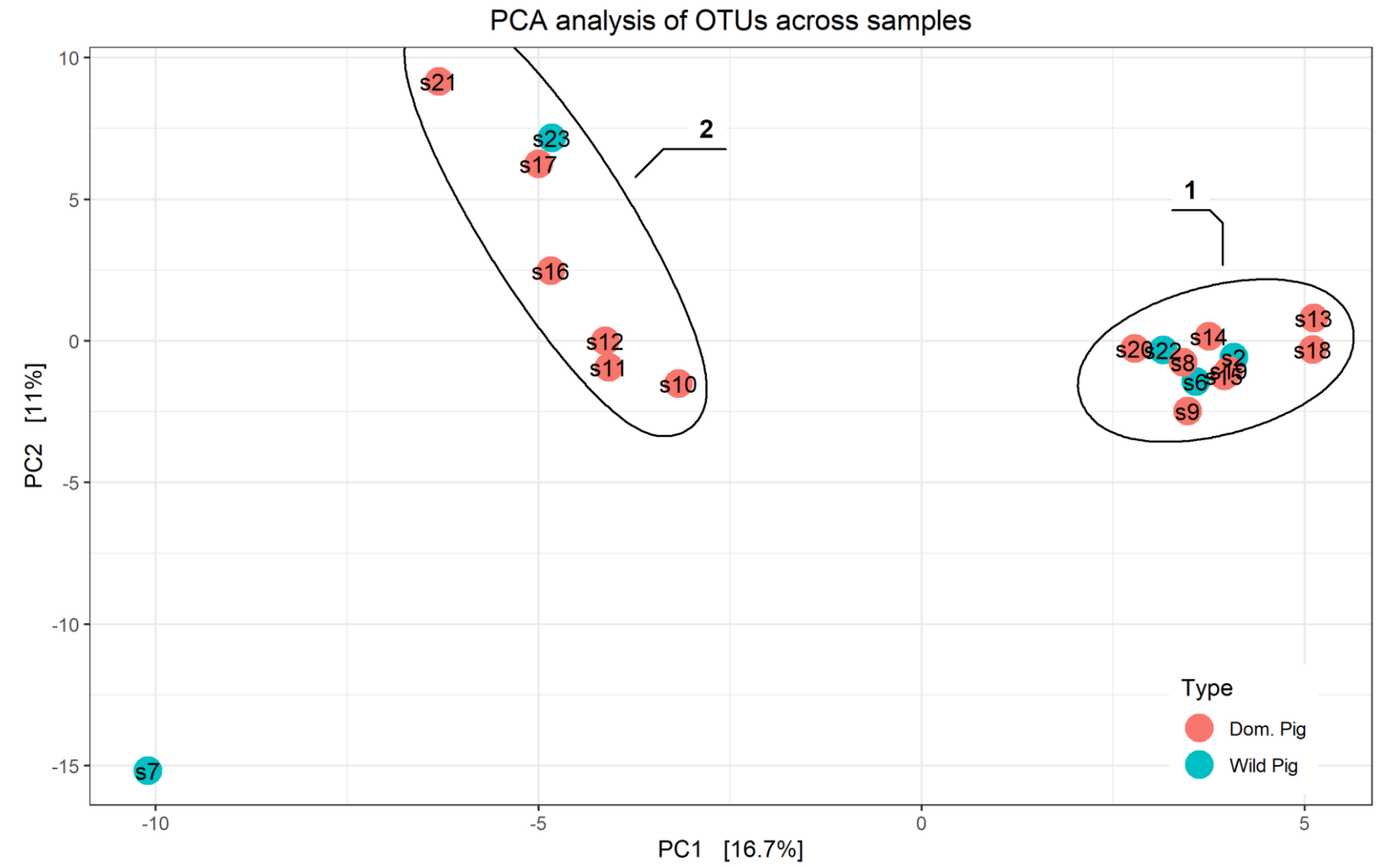
Microbial diversity

The microbial composition of the pig and boar feces samples were analyzed using Kaiju with a non-redundant database of bacterial, eukaryotic, archaeal and viral proteins. Here, 40685 operational taxonomic units (OTUs) were identified of which 32140 were bacterial. We focused the downstream analysis on this subset. In order to elucidate the microbial composition of the dataset, the relative abundance of each sample was computed. In order to simplify the visualization, OTUs which constituted less than 0.0005% of total reads were filtered out and put into one of the two NA categories (although not for the subsequent analysis).

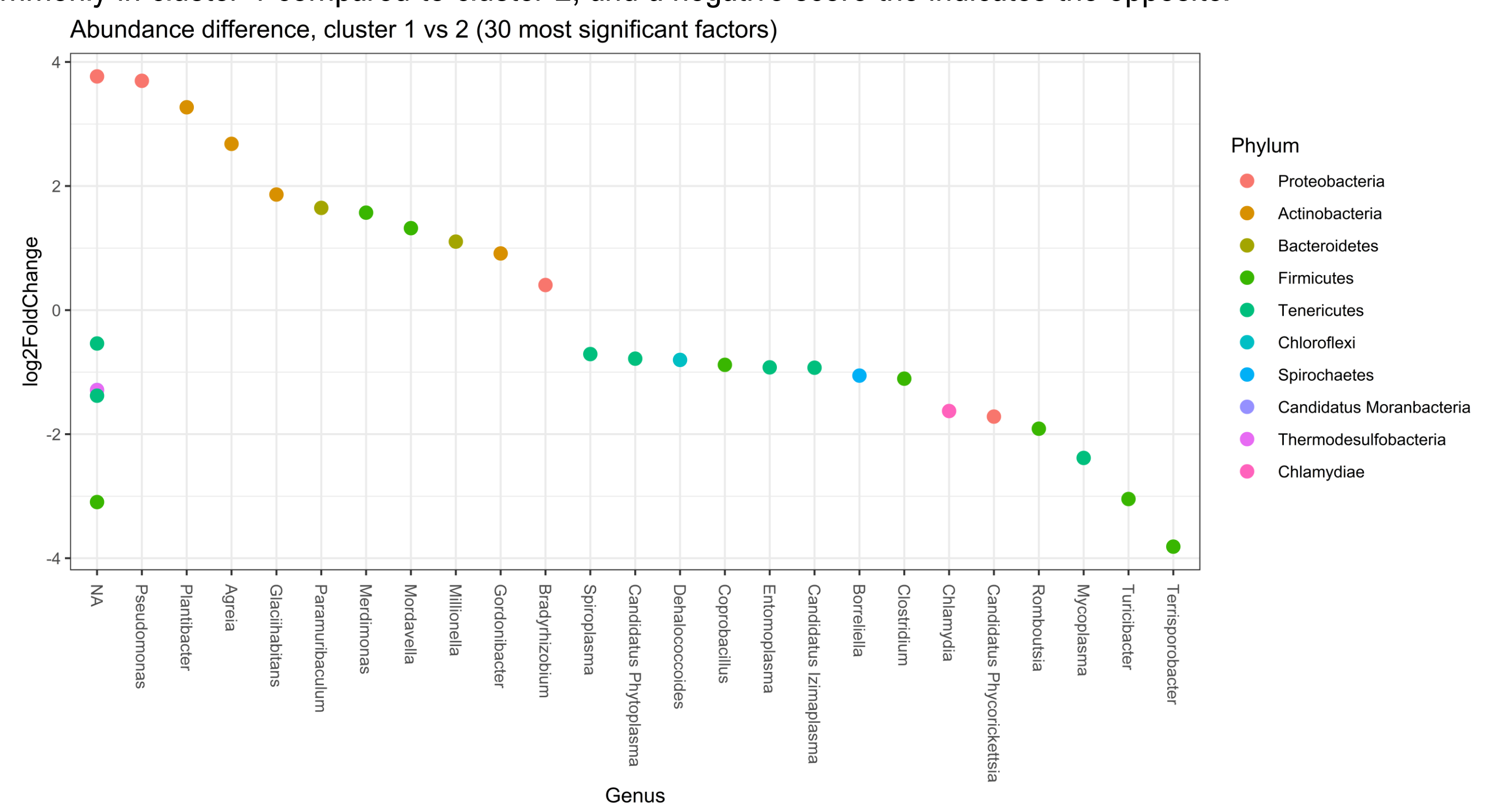


The composition of sample 1 stands out from the other samples. It is dominated by the Proteobacteria phylum which encompasses many common pathogens. An overrepresentation of this otherwise low-abundance phylum may suggest that the host is in a diseased state [15]. Using this information in conjunction with the large ARG diversity of sample 1, it can be hypothesized that the boar which provided the sample suffered from an infection at the time of death. As this sample is a strong outlier it was filtered out during further analyses. Beyond this no obvious differences can be observed between the composition of pig and boar samples.

An RDA/PCA analysis was performed on all OTUs across samples in order to further elucidate sample variance. Interestingly, 2 distinct clusters can be observed, although not based on sample type, but largely corresponding to the clusters observed in the ARG analysis. Additionally sample 7 appears as an outlier in this analysis and is thus excluded from either cluster. Lastly, a PERMANOVA was performed in order to test for statistical significance in the difference between the two clusters as well as between the two sample types. This revealed a statistical significance of 0.001 between PCA clusters 1 and 2 with significance level $\alpha = 0.001$, but no statistical significance between the pig and boar microbiomes using any common threshold. Provided more metadata was available, further analysis could reveal the factors which cause this difference. Differential abundance between PCA clusters 1 and 2 was computed at genus level in order to determine which genera drive the variance between the clusters. A positive score indicates that a genus is observed more commonly in cluster 1 compared to cluster 2, and a negative score the indicates the opposite.



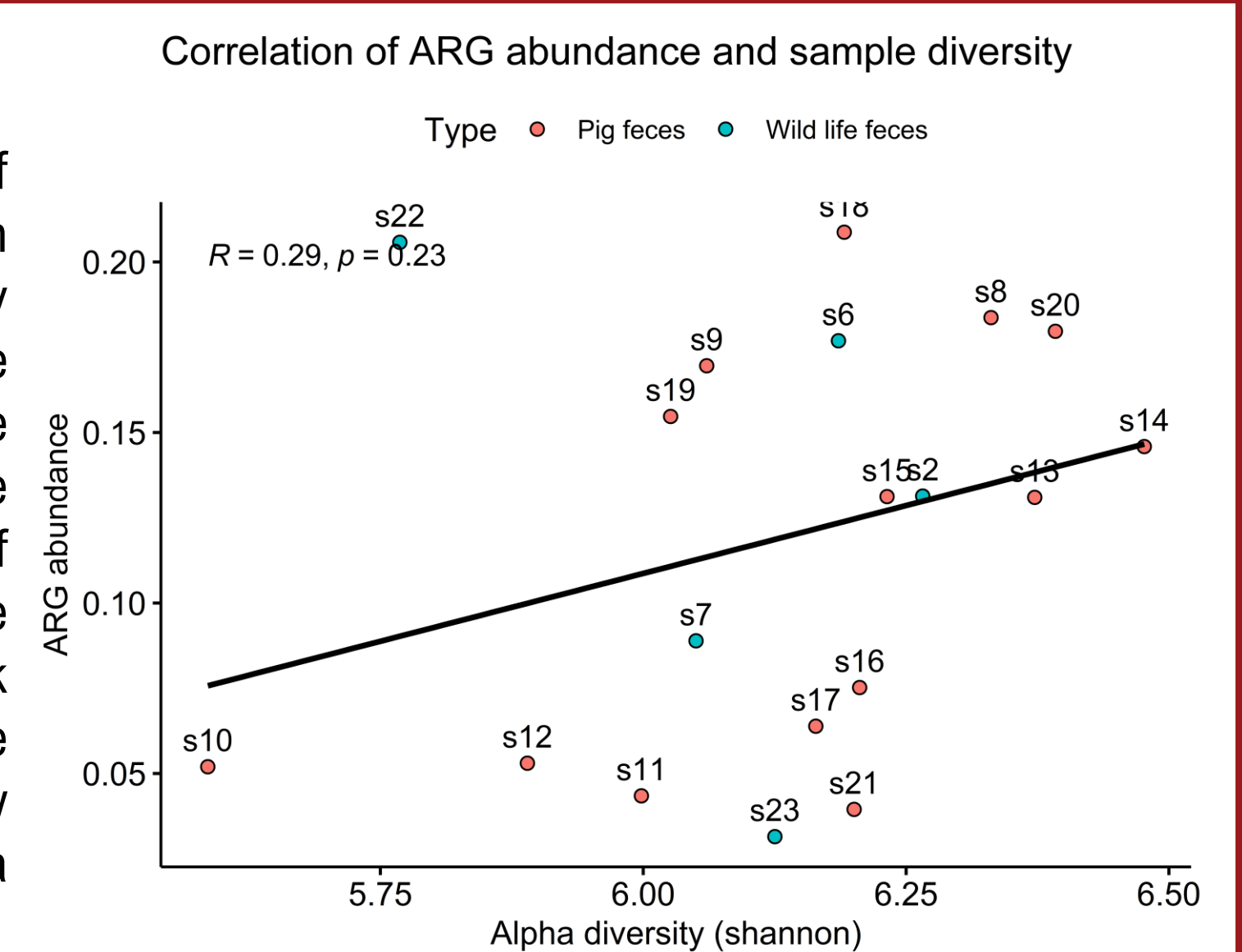
Analysis groups	R ²	p-value
Pig feces vs. Wild life feces	0.0567	0.343
Cluster 1 vs. Cluster 2	0.166	0.001



We see that genera belonging to Proteobacteria strongly drive the difference between cluster 1 and 2. As previously mentioned, this phylum encompasses many common pathogens, and thus it is interesting that it is also this cluster of samples which exhibits the greatest ARG diversity. Additionally, we see that some Bacteroidetes genera are overrepresented in cluster 1, while many Firmicutes genera are more so present in cluster 2. Interestingly, the Proteobacteria and Bacteroidetes phyla constitute a larger part of gut microbiomes of younger pigs and are diminished over time while Firmicutes becomes more dominant [18]. Thus age might be contributing factor in separating the two clusters, although this is entirely speculative and requires more rigorous analysis.

Discussion

To test our hypothesis, we estimated the alpha diversity of each sample using the Shannon diversity index. We then computed the Pearson correlation between the diversity and the ARG abundance. No significant negative correlation could be detected contrary to what might have been expected [14]. It can be speculated that the domesticated pigs simply did not receive a large degree of antibiotics and thus do not provide a baseline to which we can compare ARG abundance of wild pigs. However, lack of metadata obscures the underlying reasons for the unexpected outcomes of the analyses. Furthermore low sample count confounds the analysis and results in a diminished statistical power of the obtained results.



1. Aarestrup FM, Kruse H et al. Associations between the use of antimicrobial agents for growth promotion and the occurrence of resistance among *Enterococcus* faecium from broilers and pigs in Denmark, Finland, and Norway. *Microb Drug Resist*. 2000 Spring;6(1):63-70. doi: 10.1089/mdr.2000.6.63
2. Munk P, Knudsen B, Lückjansen O et al. Abundance and diversity of the faecal resistome in slaughterhouse pigs and broilers in nine European countries. *Nature Microbiology* 2016. 3. 896-908. doi:10.1038/51564-018-0192-9
3. sourceforge.net/projects/bbmap/
4. Joshi NA, Fass JN. (2011). Sickle: A sliding-window, adaptive, quality-based trimming tool for FastQ files (Version 1.33) [Software]. Available at: <https://github.com/najoshi/sickle>
5. Rodriguez E et al. 2018. Nonpareil 3: Fast non-parametric coverage and sequence diversity. *msystems* (3). e00393-18. doi: 10.1128/mSystems.00393-18
6. Andrews, S. (2010). FastQ. A Quality Control tool for High Throughput Sequence Data [Online]. Available at: <http://www.bioinformatics.babraham.ac.uk/projects/fastqc/>
7. Anton Bankевич, Sergey Nur, Dmitry Antonov, Alexey A. Gurevich, Mikhail Dvorov, Alexander S. Kulikov, Valery M. Lesin, Sergey I. Nikolaev, Son Pham, Andrey D. Pajestki, Alexey V. Pyshkin, Alexander V. Sirotnik, Nikolai Vyahhi, Gleen Tesler, Max A. Alekseyev, and Pavel A. Kevzhar. *Journal of Computational Biology*. May 2012:455-477. <https://doi.org/10.1089/cmb.2012.0021>
8. Moroz R, Ke G, Li X. (2016) Fast and accurate taxonomic classification for metagenomics with Kraken. *Nat. Commun.* 7:11257
9. Alcock et al. 2019. CARD 2020: antibiotic resistance surveillance with the comprehensive antibiotic resistance database. *Nucleic Acids Research*, gkz935. [Epub ahead of print] [PMID 31665441]
10. Hyatt, Doug et al. "Prodig: prokaryotic gene recognition and translation initiation site identification." *BMC bioinformatics* vol. 11 119. B. Mar. 2010. doi:10.1186/1471-2105-11-119
11. Benjmin Buchfink, Chao Xie & Daniel H. Huson. Fast and Sensitive Protein Alignment using DIAMOND. *Nature Methods*. 12, 59–60 (2015). doi:10.1038/nmeth.3176
12. Langmead B, Salzberg S. Fast gapped-read alignment with Bowtie 2. *Nature Methods*. 2012. 9:357–359
13. Alcock et al. 2020. CARD 2020: antibiotic resistance surveillance with the Comprehensive Antibiotic Resistance Database. *Nucleic Acids Research*, 48, D517–D525.
14. Xu, L., Surathu, A., Rapley, I. et al. The effect of antibiotics on the gut microbiome: a metagenomics analysis of microbial shift and gut antibiotic resistance in antibiotic treated mice. *BMC Genomics* 21, 263 (2020). <https://doi.org/10.1186/s12864-020-6665-2>
15. Sun, J., Du, L., Li, X. et al. Identification of the core bacteria in rectums of diarrheic and non-diarrheic piglets. *Sci Rep* 9, 18675 (2019). <https://doi.org/10.1038/s41598-019-56328-y>
16. Grados-Chicharro, A. et al. (2019) The effect of antibiotics in food and Feedings:stuffs: From Regulation to Analytical Methods, Bacterial Resistance, and Environmental and Health Impacts. *Anal Methods Chem*. 2017;2017:1315497. doi:10.1155/2017/1315497
17. Satu Pyörälä, Keith Edward Baptiste, Boudewijn Jacth, Engeline van Duijken, Christina Greko, Miguel A. Moreno, M. Constança Matias Ferreira Pombo, Marjatta Rantalä, Modesta Ruzauskas, Pascal Sanders, E. Thore Jhralff, Jorri Torren-Edo, Karolina Törneke, Macrolides and lincosamides in cattle and pigs: Use and development of antimicrobial resistance. *The Veterinary Record*. 2020, Issue 182. <https://doi.org/10.1016/j.vetrec.2020.04.022>
18. Zhao W, Wang Y, Liu S, et al. The dynamic distribution of porcine microbiota across different ages and gastrointestinal tract segments. *PLoS One*. 2015;10(12):e0144141. Published 2015 Feb 17. doi:10.1371/journal.pone.0117441