

## Introduction

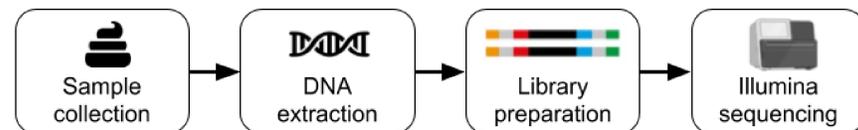
The microbiome of an animal depends of course on its specie but also differs from one individual to another, depending on its living condition. Here we investigate the metagenome of 14 pig fecal samples from conventional farms in France and 6 wild boar fecal samples from Poland - all part of the Effort program [1]. The difference between the two groups microbiome composition will be investigated (beta diversity), along with the individual microbiome diversity (alpha diversity).

Our hypothesis is that we will find a greater microbiome diversity among the wild boars compared to the pig. We base this hypothesis on the extensive use of antibiotics in pig farms [2] and that we believe wild boars have a more divers diet compare to the raised pigs.

## Processing workflow

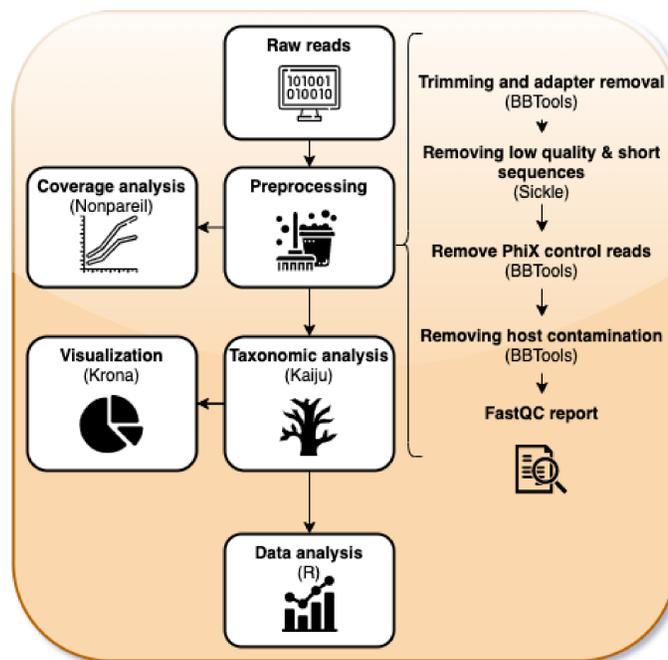
### Library preparation

Firstly, the DNA was extracted from the samples and fragmented. Next, the sequences were end-repaired and adenylated, enabling adapters containing labelling barcodes to be attached. Lastly, the paired-end fragments were sequenced using Illumina.



### Bioinformatics processing

Before starting further analysis of the sequencing output, the reads had to undergo **pre-processing**. FastQC was used to control the read quality before and after the pre-processing steps. After obtaining the processed reads, the coverage of each sample was assessed using **Nonpareil**, which estimates the coverage by comparing the redundancy of the reads in the dataset [3]. Then, **Kaiju** was used for taxonomical classification of the samples [4]. Kaiju first translates the nucleotide sequences to amino acid sequences. The data is then compressed using a Burrows-Wheeler transformation and then compared to a NCBI database.



## Quality control and data description

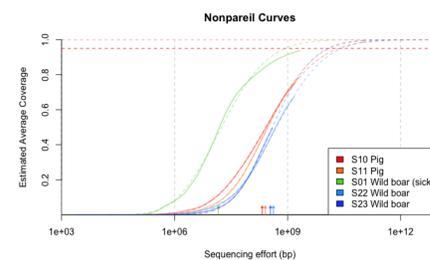


Figure 1: Nonpareil curves of selected samples.

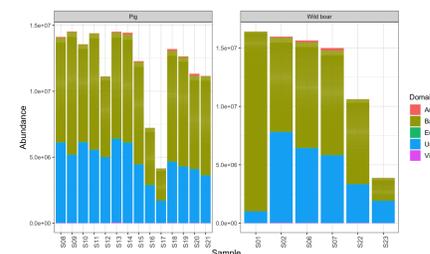


Figure 2: Read pairs classification by domain.

### Coverage result analysis

Figure 1 is showing the estimated average coverage for 5 of our samples (representing the general appearance of the samples). All samples except sample S01 have a relatively low coverage (around 75%). In order to obtain a better coverage, more reads are required. However, the average coverage of sample S01 almost reaches 95%. We will see later that this sample contains lower diversity, explaining its coverage. Note that the more reads classified as the same species, the less reads are needed to cover the whole sample.

### A first look: Kaiju output

Figure 2 shows the sample size for each sample and how the OTUs are classified between the domains, it also shows the amount of unclassified data. What is more interesting is the appearance of sample S01, it stands out compared to the other samples containing far more classified reads. This sample will require a special attention.

## Taxonomic distribution and diversity analysis

### Phyla relative abundance

To get an better overview of the microbial compositions of the samples, we plotted the relative abundance on Figure 3. The graph only shows bacterial phyla, and unclassified and unknown reads are removed. No clear difference between the two groups can be seen. However, the composition of phyla abundances in sample S01 differs significantly, as it contains a higher portion of Proteobacteria, which is hardly represented in the other samples.

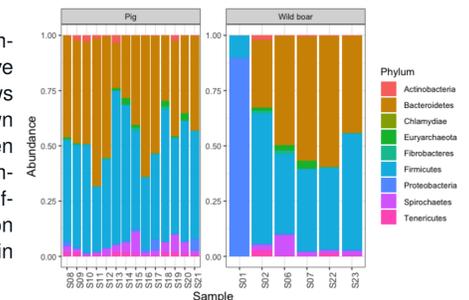


Figure 3: Distribution of taxonomical Phylum.

After a closer look at the species found in this phyla, we found that sample S01 contains far higher proportions of *E. coli* and *Shigella sonnei* than the rest of the samples (see Table 1). *Shigella sonnei* is a pathogenic bacteria, causing severe intestinal infection (shigellosis). Thus, we can infer that this sample comes from a sick wild boar. Sample S01 is not considered representative for the wild boars and is excluded in later analysis.

Bacteria	Abundance in S15	Abundance in S01	Prop. $\propto$
<i>Shigella sonnei</i>	0.0120%	2.26%	188
<i>Escherichia coli</i>	0.105%	13.1%	125

Table 1: Relative abundance of *S. sonnei* and *E.coli*. After S01 the highest abundances are seen in sample S15.

**Redundancy analysis** A constrained PCA from the taxonomical composition was performed. Figure 4 shows its first two principal components, accounting for 50% variance explained. Based on our hypothesis, we would expect one cluster containing the pig samples and the other one with the wild boar samples. Instead, we see two main clusters were the two groups are mixed, indicating that the animal species couldn't be split on this 2D orthogonal space.

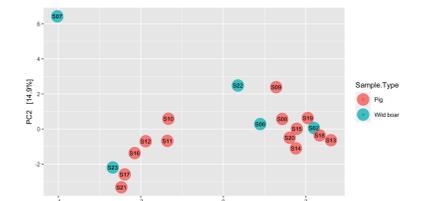


Figure 4: Redundancy Analysis based on the taxonomical composition.

### Species Richness: Alpha diversity

Alpha diversity accounts for the diversity within the metagenome sample. The more OTUs and the more evenness among them, the higher the diversity. We quantify alpha diversity in terms of the Shannon index  $H'$ .  $H' = -\sum_{i=1}^R p_i \ln(p_i)$  with  $R = |\text{observed species}|$  and  $p_i = i\text{-th specie proportion}$ .

Figure 5 shows the Shannon index for wild boars and raised pig. As expected, the sick individual has less diversity. For the other samples, Shannon index ranges from 3.8 to 4.9, indicating a similar diversity in all samples, regardless the type of pig.

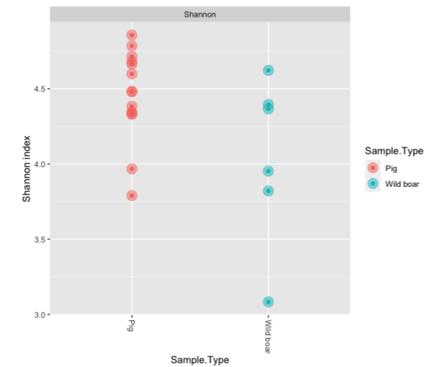


Figure 5: Shannon Index of all samples.

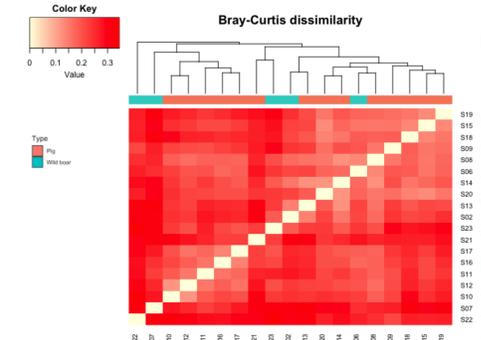


Figure 6: Heatmap of Bray-Curtis dissimilarity

### Bray-Curtis dissimilarity: Beta diversity

To study the beta diversity, we used Bray-Curtis dissimilarity. It allows to quantify the diversity similarity between two samples. The heatmap shows a similar dissimilarity between samples. Also, from the built hierarchical tree, we can see that wild boars and raised pigs are mixed together, showing a similar pattern as Figure 4.

### Pairwise test: Beta diversity

Additionally, we normalized the samples with the centered log-ratio transformation and compute their euclidean distance (Aitchison distance). We performed a post-hoc pairwise test over these distances, obtaining a p-value = 0.11, which confirms again that there isn't enough information to state for a significant difference.

### Differential abundance

Using the DESeq2 software, we generated Figure 7, which displays the differential abundance between the pig and the wild boar group, with the abundance expressed as log fold change. By applying a threshold for entries with p-values < 0.05, we removed non-significant results in the plot. The plot only shows 13 OTUs on genus level are significant. It is important to note that when sample S01 was introduced, we obtained a completely different picture, as the sick wild boar altered the average genus composition of this group.

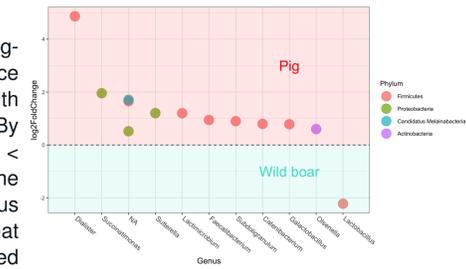


Figure 7: Estimation on Genus level of fold change between pigs and wild boars by DESeq2.

## Conclusion

From the compositional analysis, we observed that the data provided were not sufficient to differentiate the microbiome between wild boars and raised pigs. Similarly, results regarding alpha and beta diversity do not provide any additional distinction between animal species. Hence, the same study should be reproduced including more samples to ensure that these similar patterns are not due to the lack of observations.

## References

1. *EFFORT against antimicrobial resistance* <http://www.effort-against-amr.eu/>. Accessed: 2020-11-30.
2. Lekagul, A., Tangcharoensathien, V. & Yeung, S. Patterns of antibiotic use in global pig production: A systematic review. *Veterinary and Animal Science* 7, 100058 (2019).
3. Rodriguez-R, K. T. Nonpareil: a redundancy-based approach to assess the level of coverage in metagenomic datasets. *Bioinformatics* 30, 629–635 (2014).
4. *KAIJU. Fast and sensitive taxonomic classification for metagenomics* <http://kaiju.binf.ku.dk/>. Accessed: 2020-11-30.