# Exploration of specific pathogens related to illness in domestic pigs and wild boars

Group 2: Anna Schaub, Maja Jager, Carolina Rocha

## Introduction

Wild boars and pigs both belong to the same species, Sus scrofa. This means that they are susceptible to similar pathogens, like for example swine fever virus, and domestic pigs can be infected from wild boars with many different pathogens and illnesses.[1] In this study, we are aiming to find out if domesticated pigs and wild boars have the same pathogens or if there is a difference in the occurrence and abundance of pathogens between these two groups. A list of pathogens typical for Sus scrofa was created and compared to OTUs that were found in the dataset. 6 Boars from Poland and 3 boars from Japan were compared to 14 pigs from France.

**Hypothesis: There is no difference in the occurrence and abundance of specific pathogens in domestic pigs and wild boars.**

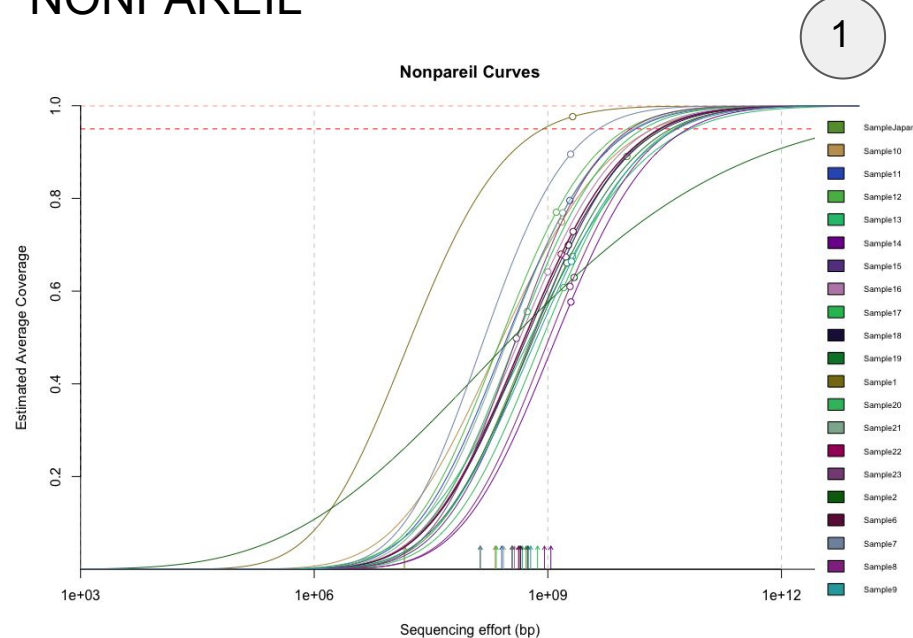## Workflow
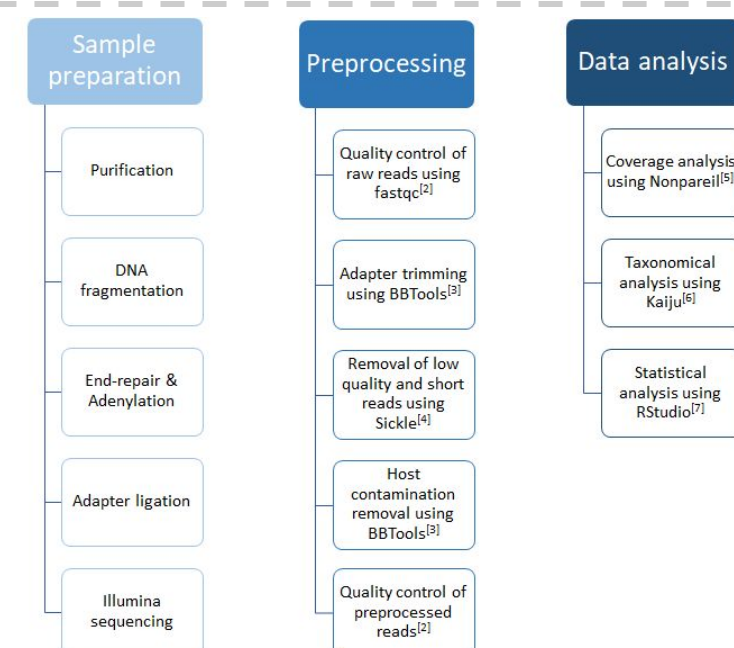


## Results

### NONPAREIL



Figure 1 shows the nonpareil curves from all the samples giving us an estimated mean coverage of 69,8%. Looking at the overall diversity the Wilcoxon test is used to determine that there is no significant difference between the domesticated pigs and wild boars with a p-value of 0.079.

### ABUNDANCE ANALYSIS

#### Proportion of Domains and Genera
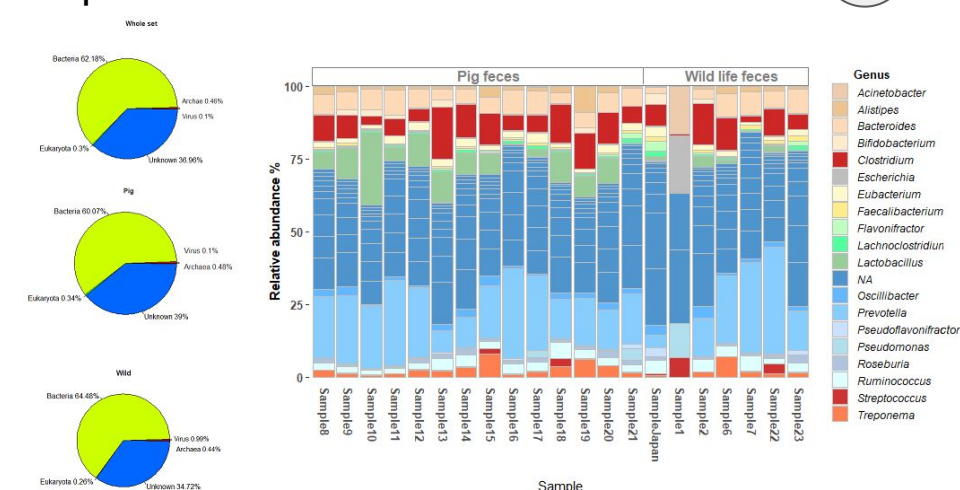


In Figure 2a, on the left the relative abundance of the different domains and on the right of the top 20 most abundant genera are visible. In general, all samples show a similar pattern in the microbial composition, except for Sample1. This sample seems to have a higher presence of Escherichia. Two of the shown genera (*Clostridium* and *Streptococcus*) contain pathogenic species from the list of pathogens.

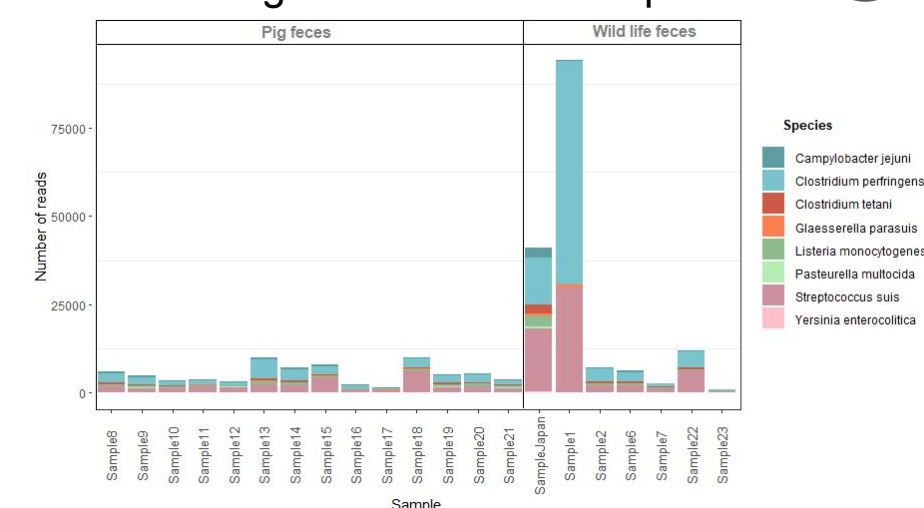#### No. of Pathogenic Reads in Samples



Figure 2b displays the abundance of reads of the pathogens found in the different samples. It is clear that Sample1 has a unusually high presence of pathogens. The reason for the high numbers in SampleJapan is because it is a pooled sample from 3 boars.

#### Pathogens Stratified on Sample Type



Figure 2c visualizes the relative abundance of the 8 pathogens of interest stratified on sample type. Only *Listeria monocytogenes* has a significant difference between the two groups with a p-value of 0.015.
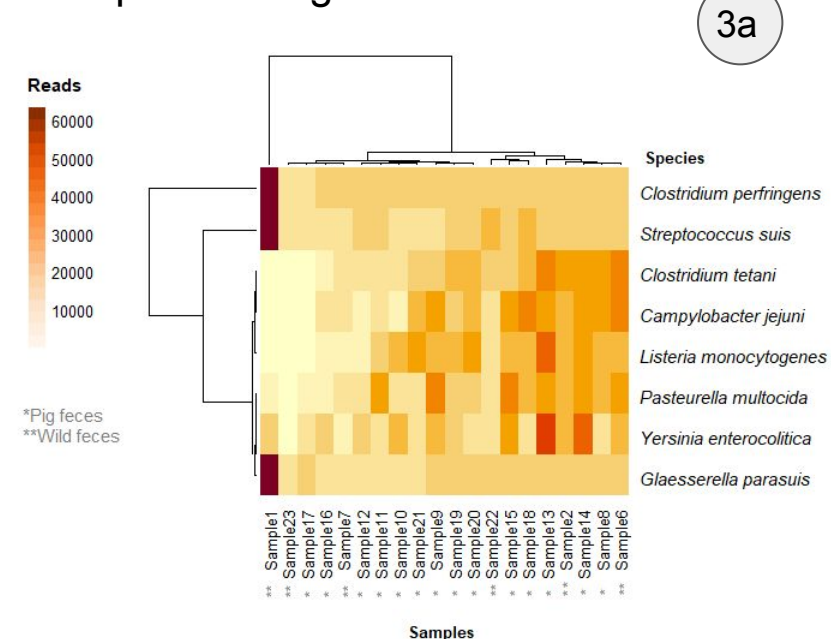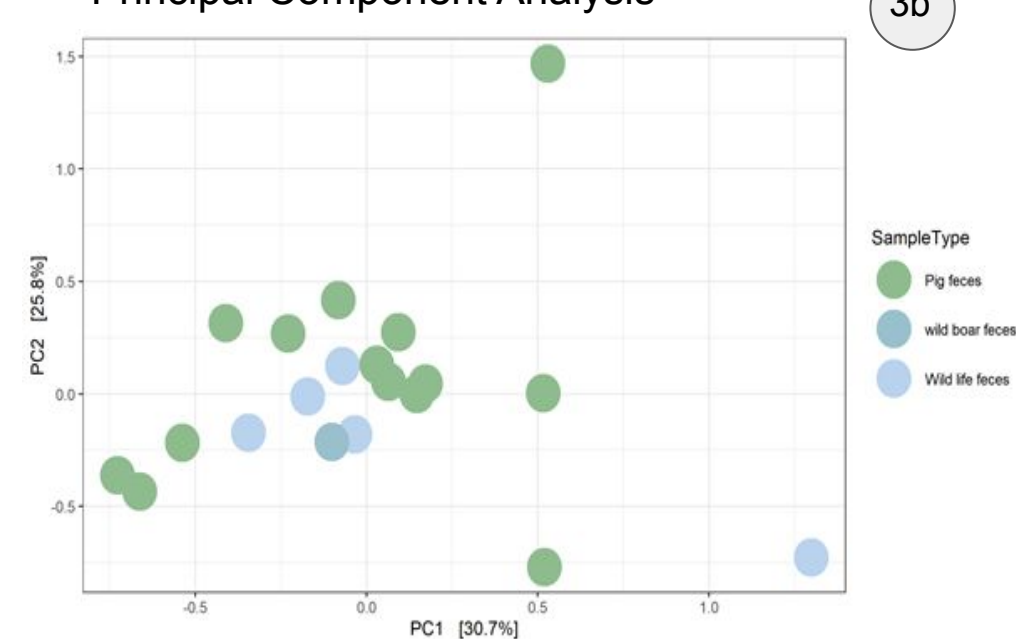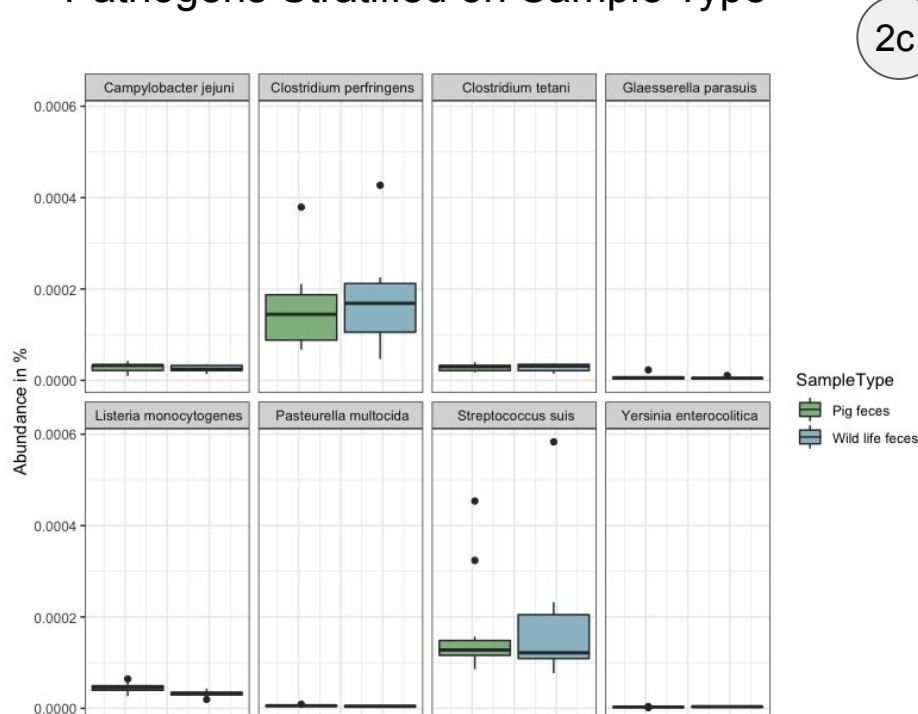
### CLUSTERING

#### Heatmap of Pathogens



Figure 3a shows the number of reads observed in the samples. The highest number of reads are observed in species such as *Clostridium perfringens*, *Streptococcus suis* and *Glaesserella parasuis*. No clustering in the two groups is detected.

#### Principal Component Analysis



PCA was done on the whole set of data and on a subset with only OTUs describing the pathogens. Figure 3b visualizes the plot of PC1 against PC2 on the subset. There is no significant clustering of the boars or pigs visible.

## Conclusion

Out of the 23 pathogens consisting of Bacteria and Viruses of interest, 8 bacterial species were found to have a high enough number of reads to consider it present, while no Viruses were found to be abundant. The bacteria with highest presence found are *Streptococcus suis* and *Clostridium perfringens*.

One boar was found to have a different abundance of the pathogens and less diversity in general than all other pigs and boars, which might suggest that this pig is ill.

Except for *Listeria monocytogenes*, no significant difference in the occurrence and abundance of pathogens between domestic pigs and wild boars was found. In addition, the entire data included only 21 samples, which is not enough to measure any significant results.

**Therefore the null hypothesis can not be rejected.**

## Outlook

- Reconstruct the genomes of the most abundant pathogens and get more context on the data.
- Collect more samples to be able to increase the significance of results.
- Look especially at Sample1: Explore its total microbial composition and determine the roles of the different species in the occurrence of diseases.

## References

[1] Wild boar as the reservoir of pathogens, pathogenic for swine, other species of animals and for humans (researchgate.net)
[2] Andrews, S. (2010). FastQC: A Quality Control Tool for High Throughput Sequence Data [Online]. Available online at: http://www.bioinformatics.babraham.ac.uk/projects/fastqc/
[3] BBMap – Bushnell B. – 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] Luis M. Rodriguez-R, Konstantinos T. Konstantinidis, Nonpareil: a redundancy-based approach to assess the level of coverage in metagenomic datasets, *Bioinformatics*, Volume 30, Issue 5, 1 March 2014, Pages 629–635, https://doi.org/10.1093/bioinformatics/btt584
[6] Menzel, P., Ng, K. & Krogh, A. Fast and sensitive taxonomic classification for metagenomics with Kaiju. *Nat Commun* 7, 11257 (2016). https://doi.org/10.1038/ncomms11257
[7] RStudio Team (2019). RStudio: Integrated Development for R. RStudio, Inc., Boston, MA URL http://www.rstudio.com/.