IDENTIFICATION OF A CORE METAGENOME THROUGH FUNCTIONAL ANNOTATION IN SWINE AND WILD BOARS

Jorge Carrasco (s192528), Martí Morera (s192447) and Anna Oliver (s202436)
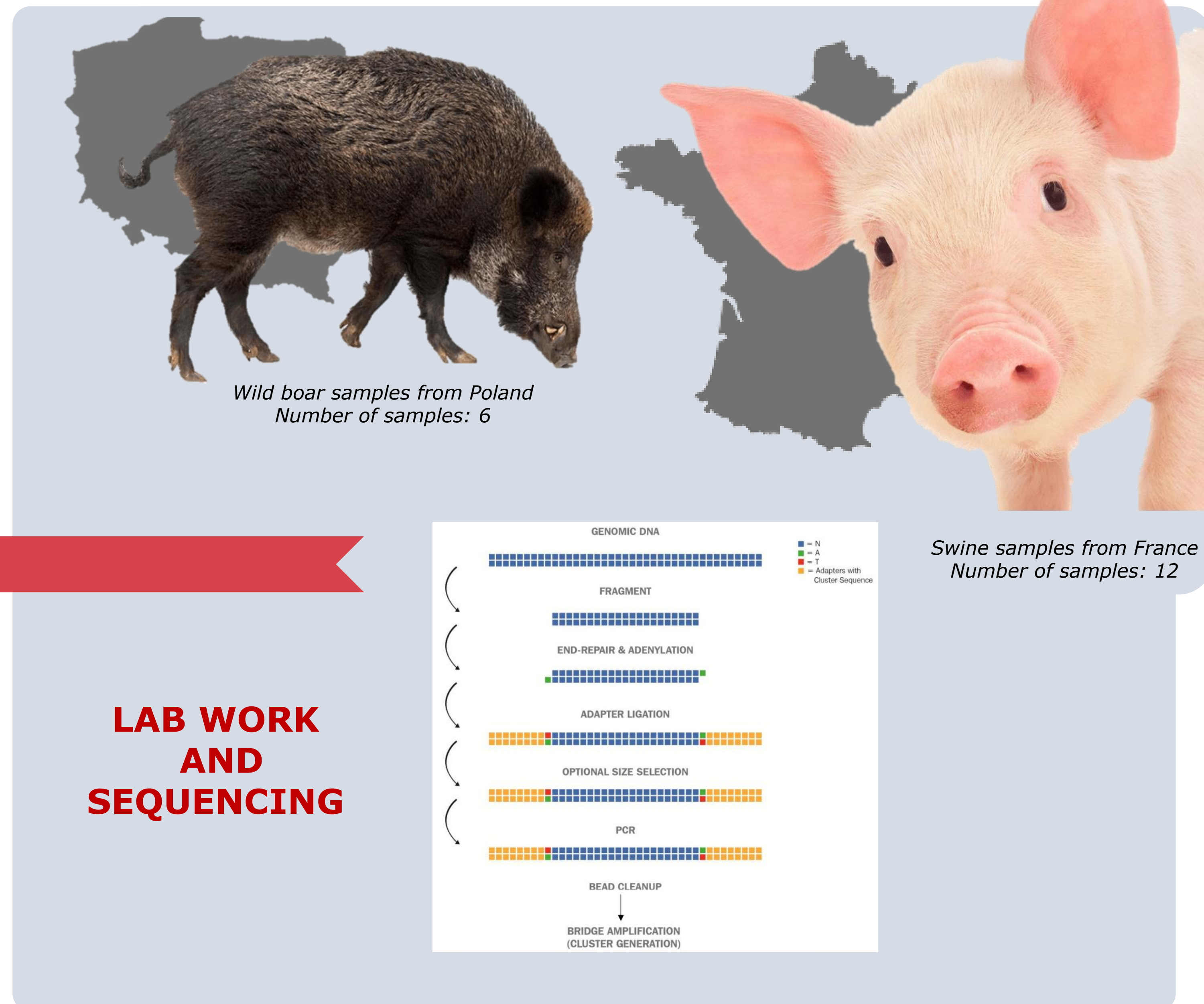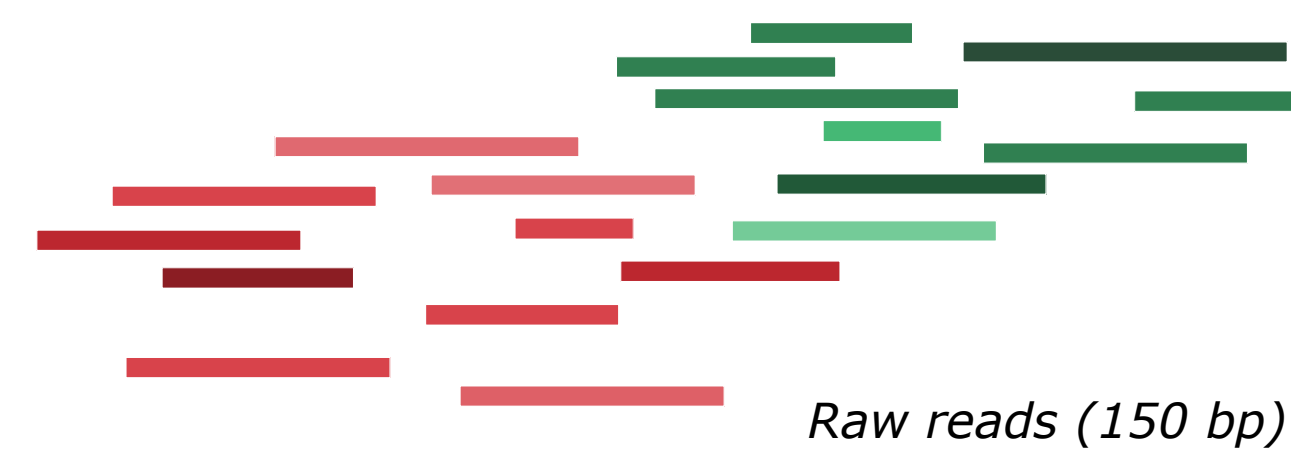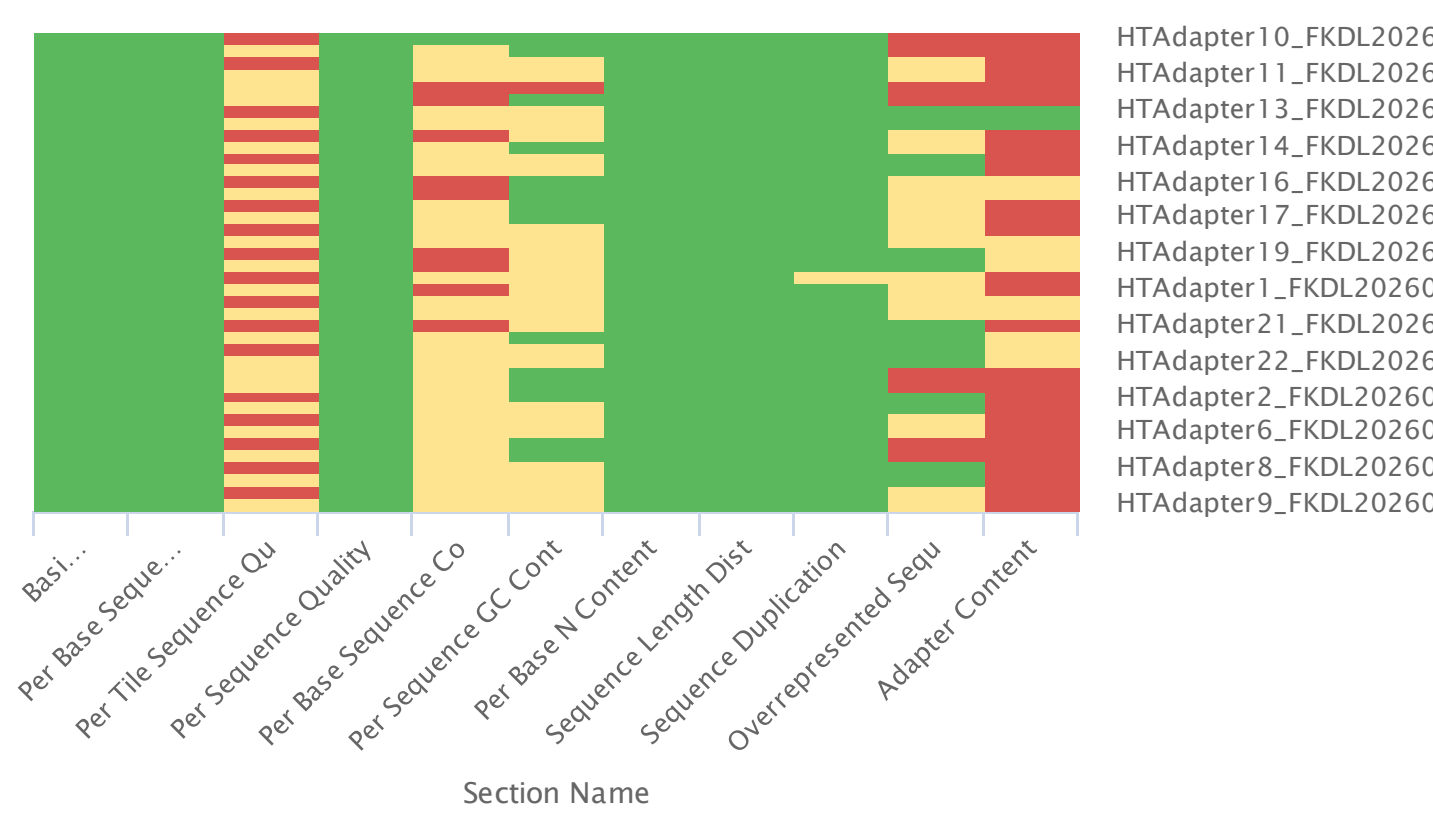
DTU

- **Hypothesis 1**

  A functional core metagenome can be defined for both species

- **Hypothesis 2**

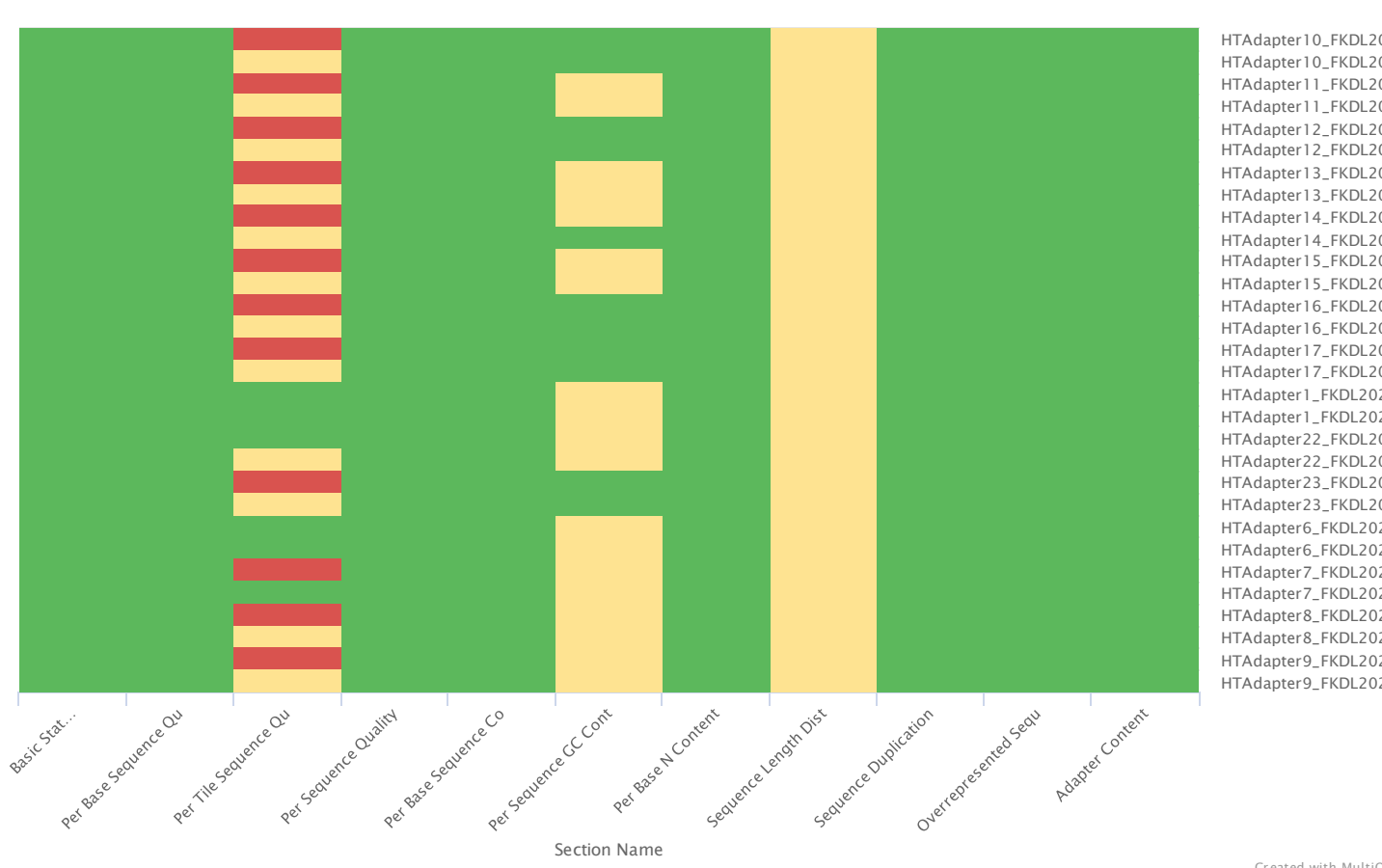 *Swine and wild boars have a different functional core metagenome*

Wild boar samples from Poland
Number of samples: 6

Swine samples from France
Number of samples: 12

Multi FastQC before preprocessing [1,2]

Raw reads (150 bp)

LAB WORK AND SEQUENCING

TRIMMING AND ADAPTER REMOVAL

BBDuk [3]
forcetrimleft=5 ktrim=r k=23
mink=11 hdist=1 hdist2=0 tpe tbo

QUALITY FILTER

Sickle [4]
-l 50

HOST AND PHIX REMOVAL

BBTools [3]
Phix: k=31 hdist=1

Host removal:
– Swine (*Sus scrofa domesticus*) → GCA_000003025.6
– Wild boar (*Sus scrofa scrofa*) → GCA_006511355.1

NCBI

FUNCTIONAL ANNOTATION

eggNOG [5,6]

GENE ONTOLOGY

topGO [7]

❶ Filter those reads with predicted gene ontology

❷ Use elim algorithm with Kolmogorov-Smirnov method to find the most specific (down in the hierarchy) gene ontologies that are significative.

❸ Count number of samples where a significative ontology appears

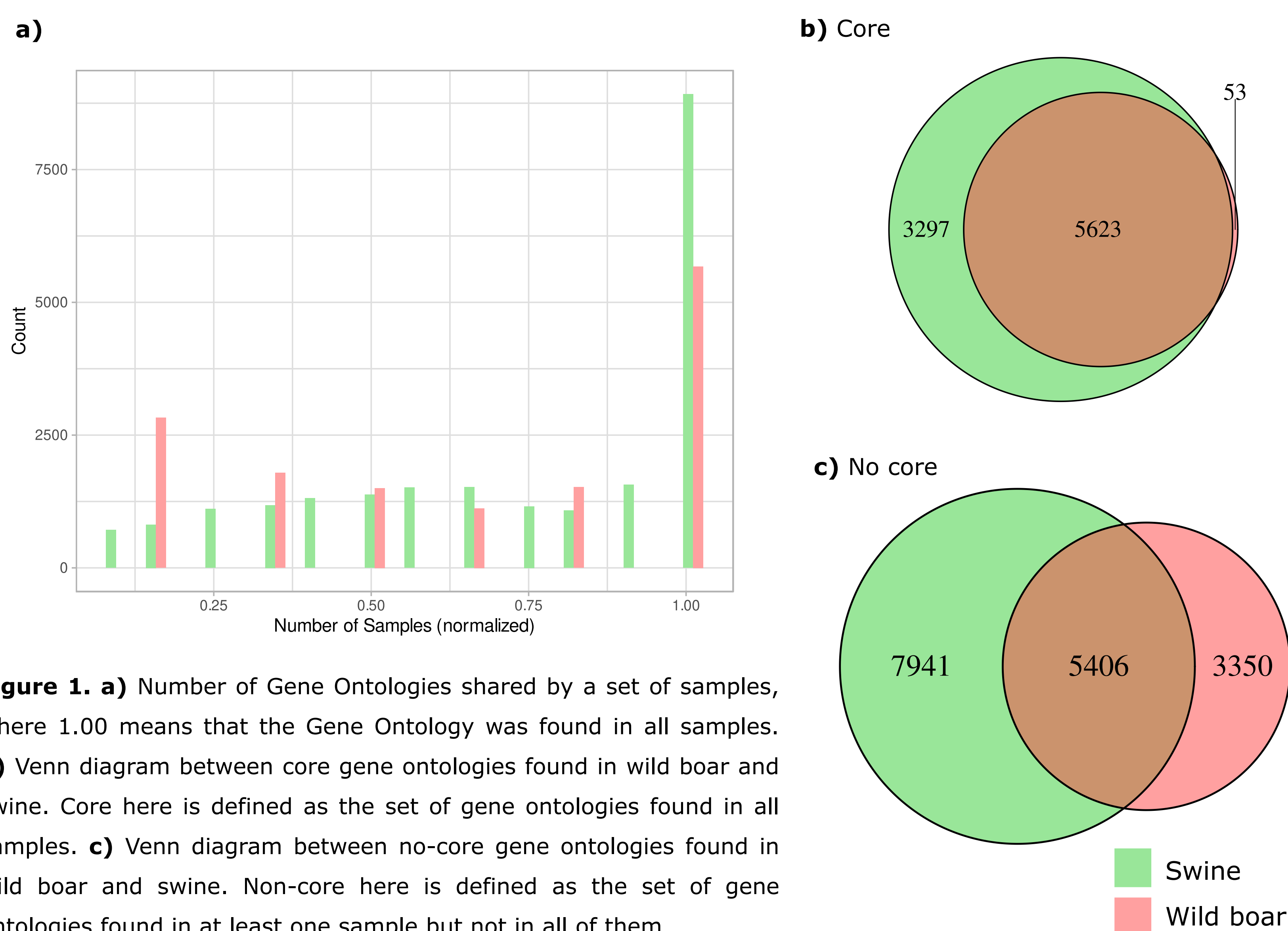## Multi FastQC after preprocessing [1,2]

RESULTS

Figure 1. a) Number of Gene Ontologies shared by a set of samples, where 1.00 means that the Gene Ontology was found in all samples. **b)** Venn diagram between core gene ontologies found in wild boar and swine. Core here is defined as the set of gene ontologies found in all samples. **c)** Venn diagram between no-core gene ontologies found in wild boar and swine. Non-core here is defined as the set of gene ontologies found in at least one sample but not in all of them.
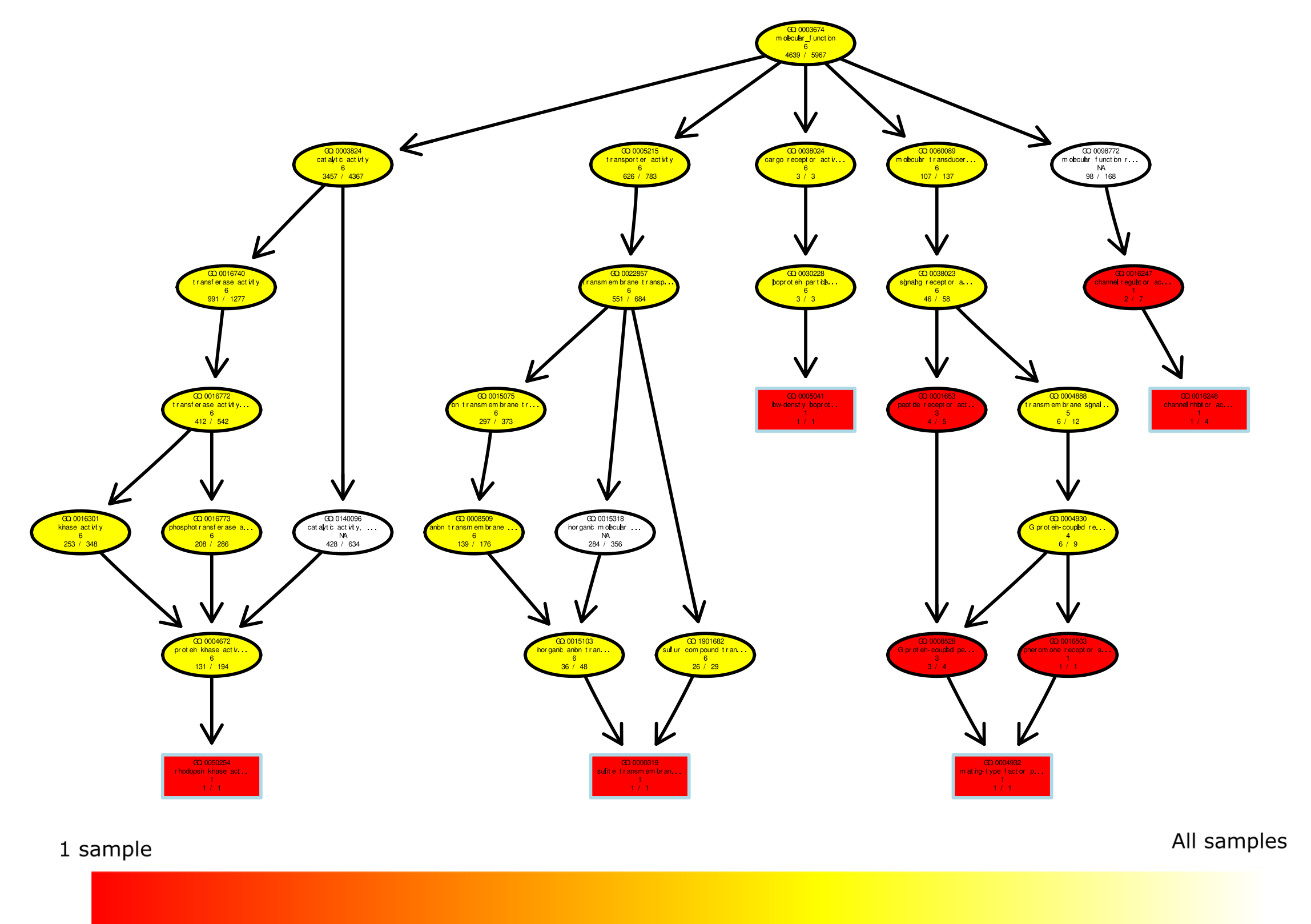
a)

b) Core

3297 5623 53

c) No core

7941 5406 3350

Swine
Wild boar

1 sample   All samples

Figure 2. Wild boar gene ontology tree of significative Gene Ontologies scored by number of samples where they appear. Here the number of tips is 5, but for it to be meaningful this number should be increased. However, it was impossible for us to keep it readable with more tips using topGO.

– **Hypothesis 1**

A functional core metagenome can be defined for both species. Most of the Gene Ontologies are shared by all samples for both species.

– **Hypothesis 2**

The core Venn diagram suggests that, if more samples were provided, swine and wild boars would not have a different functional core metagenome. A deeper analysis should be done to differentiate alpha and beta diversity within and between the two organisms gut metagenomes.

[1] Philip Ewels, Måns Magnusson, Sverker Lundin and Max Käller (2016). MultiQC: Summarize analysis results for multiple tools and samples in a single report. Bioinformatics. doi: 10.1093/bioinformatics/btw354 PMID: 27312411
[2] Wingett SW and Andrews S. (2018). FastQ Screen: A tool for multi-genome mapping and quality control. F1000Res. 2018 Aug 24 [revised 2018 Jan 1];7:1338. doi: 10.12688/f1000research.15931.2. eCollection
[3] Bushnell B. - BBMap - sourceforge.net/projects/bbmap
[4] Joshi NA and Fass JN (2011). Sickle: A sliding-window, adaptive, quality-based trimming tool for FastQ files (Version 1.33). 2011 [Software]. Available atv https://github.com/najoshi/sickle
[5] Jaime Huerta-Cepas, Damian Szklarczyk, Davide Heller, Ana Hernández-Plaza, Sofia K Forslund, Helen Cook, Daniel R Mende, Ivica Letunic, Thomas Rattei, Lars J Jensen, Christian von Mering, Peer Bork (2019). eggNOG 5.0: a hierarchical, functionally and phylogenetically annotated orthology resource based on 5090 organisms and 2502 viruses. Nucleic Acids Res. 2019 Jan 8; 47(Database issue): D309–D314. doi: 10.1093/nar/gky1085
[6] Jaime Huerta-Cepas, Kristoffer Forslund, Luis Pedro Coelho, Damian Szklarczyk, Lars Juhl Jensen, Christian von Mering and Peer Bork (2017). Fast genome-wide functional annotation through orthology assignment by eggNOG-mapper. Mol Biol Evol. 2017. [doi: 10.1093/molbev/msx148] doi.org/10.1093/molbev/msx148
[7] Alexa A, Rahnenfuhrer J (2020). topGO: Enrichment Analysis for Gene Ontology. R package version 2.42.0.
[8] Tange, O. (2020, August 22). GNU Parallel 20200822 ('Beirut'). Zenodo. doi.org/10.5281/zenodo.3996295