ESCLAVO v1.0

16S analysis

Autor: Michael Jackson

Contents

1- Experimental samples read asignation

A total of **3** sequence were recived and processed through Illumina Miseq platform, generating 150bp paired-end sequences for each sample (r1 & r2) with qualified libraries. For full analysis source and files results please download the *projects* folder which contains the following:

- **0-raw**: The initial FASTQ files used in downstream steps, also contains an initial QC report.
- **1-qc**: Filtered reads used to map against DADA2 DB, also contains an QC report.
- **2-taxInsight**: Counts and Plots for mapped reads against DADA2 DB.
- **report.pdf**: static report, a formal way to read a report.
- **report.html**: dynamic report with interactive plots and tables.

2- Read quality control

The first bioinformatic analysis step is the quality filter reads using DADA2[1] software which consider:

- adapter removal from Illumina sequencing.
- remove bad quality bases at the end of reads (min Quality: 20).
- remove reads who in average have Phread < 20.
- remove Phix sequences from samples.

Quality reports for the **3** samples before and after filter were generated using FASTQC[2]. The figures 1 and 2 show a high quality and homogenity distribution along the reads, both represent a post filter quality summary for the **6** files (r1 and r2), which can be analyzed in more details in the following table:

Table 1: Summary of quality control initial reads after dada2 filt, derepli-
cate and merge steps.

Sample	input	filtered	denoisedF	denoisedR	merged	nonchim
ERR3357130	100000	89380	88974	88399	81785	70379
ERR3357150	100000	90656	90352	89911	85476	75759
ERR3357164	100000	93646	93330	93296	89490	77538

3- Taxonomic insight

After non-chimera filtered reads, the next step is map it against a taxonomic database, we used the SILVA[3] v132 database and DADA2 *assignTaxonomy* function to map in a 100% the aligment as their authors recommends. The a summary of results obtained from the assign taxonomy can be checked in the following table:

Table 2: DADA2 top 10 abundances taxas for each sample. The table
include at least 1 read assigned for at least 1 sample.

Family	Genus	Species	ERR3357130	ERR3357150	ERR3357164
Bacteroidaceae	Bacteroides	acidifaciens	4090	1466	4523
Bacteroidaceae	Bacteroides	vulgatus	0	0	7658
Muribaculaceae	NA	NA	482	2112	254
Prevotellaceae	Prevotellaceae_UCG-001	NA	10729	8058	5483
Rikenellaceae	Alistipes	NA	2896	1654	1835
Deferribacteraceae	Mucispirillum	schaedleri	0	0	2931
Helicobacteraceae	Helicobacter	typhlonius	0	0	6153
Helicobacteraceae	Helicobacter	NA	657	2561	47
Lactobacillaceae	Lactobacillus	NA	455	6052	734

Table 2: DADA2 top 10 abundances taxas for each sample. The table
include at least 1 read assigned for at least 1 sample. *(continued)*

Family	Genus	Species	ERR3357130	ERR3357150	ERR3357164
Lactobacillaceae	Lactobacillus	NA	888	12149	1486
Lactobacillaceae	Lactobacillus	NA	971	1894	453
Lactobacillaceae	Lactobacillus	NA	6058	14	2907
Lachnospiraceae	Lachnospiraceae_FCS020_group	NA	292	49	1613
Lachnospiraceae	Tyzzerella_3	NA	0	2350	79
Enterobacteriaceae	Escherichia/Shigella	NA	23	11002	0
Mycoplasmataceae	Mycoplasma	NA	0	5646	11919

graphically it can be shown as a barplot:
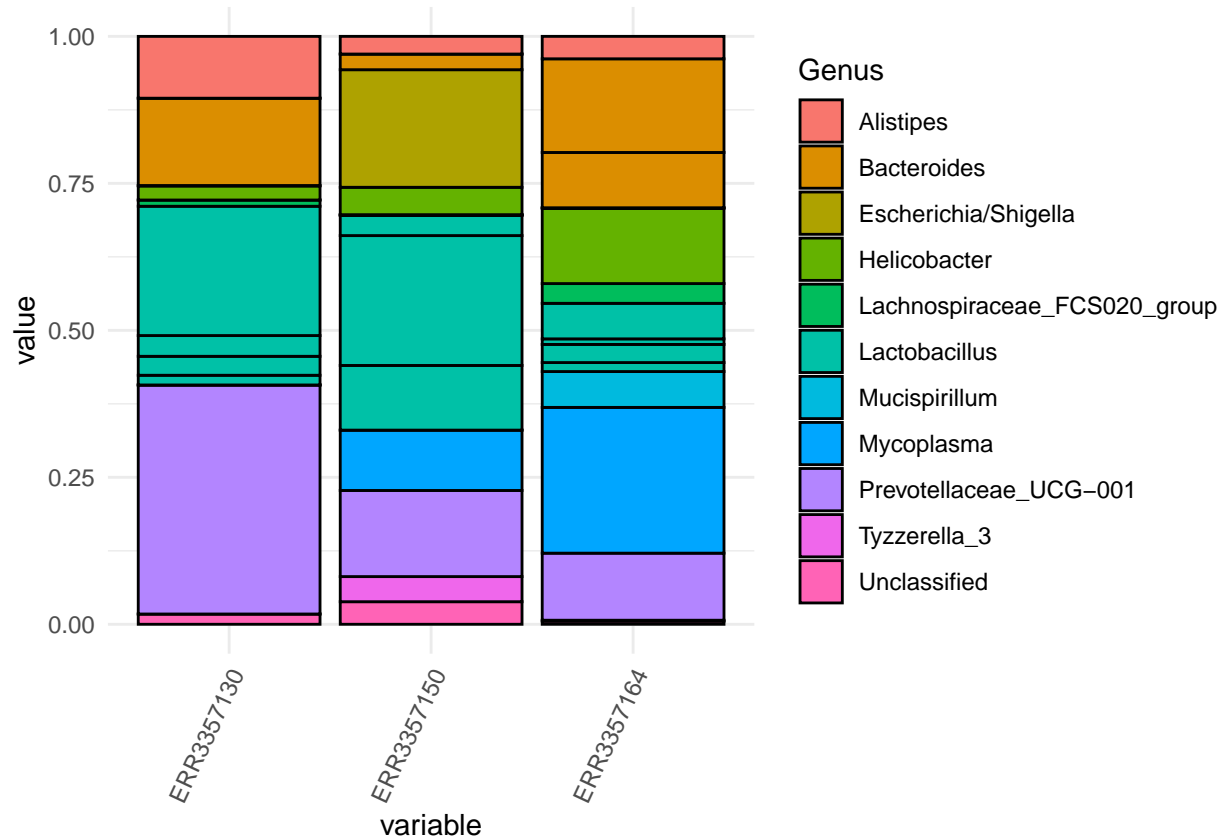
if you want the full table, please see the file 2-taxInsight/abundance.tsv.

References

1. Callahan BJ, McMurdie PJ, Rosen MJ, Han AW, Johnson AJA, Holmes SP (2016). "DADA2: High-resolution sample inference from Illumina amplicon data." Nature Methods, 13, 581-583. doi: 10.1038/nmeth.3869.

2. Philip Ewels, Måns Magnusson, Sverker Lundin and Max Käller. "MultiQC: Summarize analysis results for multiple tools and samples in a single report." Bioinformatics (2016), doi: 10.1093/bioinformatics/btw354, PMID: 27312411

3. Yilmaz P, Parfrey LW, Yarza P, Gerken J, Pruesse E, Quast C, Schweer T, Peplies J, Ludwig W, Glöckner FO (2014) The SILVA and "All-species Living Tree Project (LTP)" taxonomic frameworks. Opens external link in new windowNucl. Acids Res. 42:D643-D648

4.