Report for the analysis of 16S microbiome data

By Dr. Laurent Falquet, UniFr

Feb 17, 2020

Experimental design and goals:

The aim of the project is to identify the presence and composition of the microorganisms isolated from compost and compost tee samples using a 16S rRNA sequencing strategy in order to investigate the reproducibility of EDAPRO's brewing system as well as the diversity of the microbiome across different biological conditions (i.e. different nutrients).

- We have 3 brewing systems (Treatment: nutrient A, nutrient B, no nutrient C)
- Two different composts as starting material: Compost charge 1 = K2, charge 2 = K3
- We have 5 replicates for each brewing system processed over several weeks producing 30 tee compost samples (K=2 x T=3 x R=5)
- Some control samples (starting Compost K2 and K3 in 5 replicates, MCStd, NTC)

The different questions we would like to answer:
- determine the reproducibility of the brewing system
- diversity and identification / relative abundance of the microbiome of the tee compost samples upon addition of nutrients A and B.
- diversity and identification / relative abundance of the microbiome of tee compost samples versus the microbiome of the compost samples (charge 1 and 2)
- diversity and identification / relative abundance of the microbiome between the compost samples (charge 1 and 2)

Sequencing

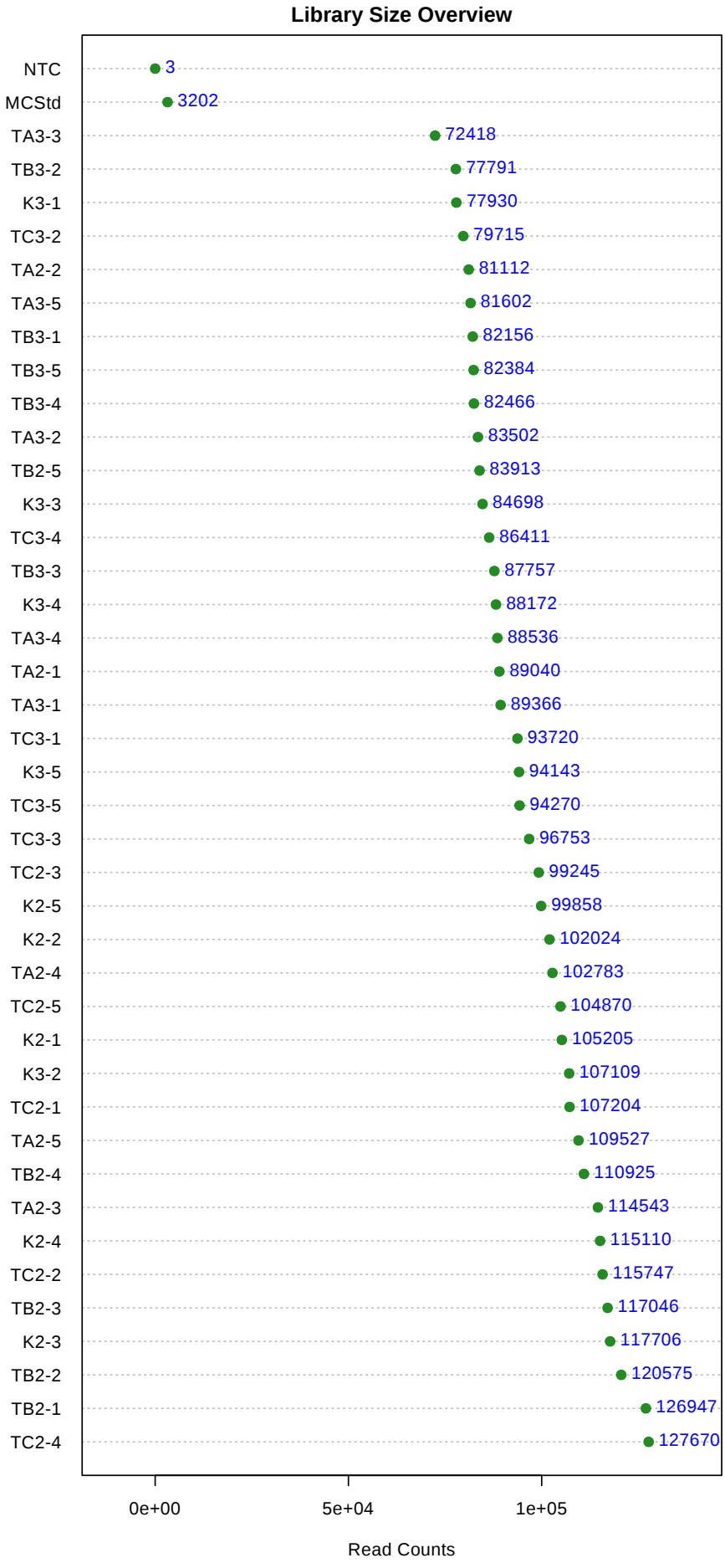
The 42 samples were amplified with primers for 16S V3-V4 region (Klindworth et al, 2013), the library prepared following Illumina recommendations and sequenced on one lane of a MiSeq (2x 300bp PE).

>Forward_Primer
TCGTCGGCAGCGTCAGATGTGTATAAGAGACAGCCTACGGGNGGCWGCAG
>Reverse_Primer
GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAGGACTACHVGGGTATCTAATCC

Analysis

1) QC of reads (FastQC)

The reads were quality controlled with FastQC, revealing no particular issues.


Library Size Overview
NTC 3
MCStd 3202
TA3-3 72418
TB3-2 77791
K3-1 77930
TC3-2 79715
TA2-2 81112
TA3-5 81602
TB3-1 82156
TB3-5 82384
TB3-4 82466
TA3-2 83502
TB2-5 83913
K3-3 84698
TC3-4 86411
TB3-3 87757
K3-4 88172
TA3-4 88536
TA2-1 89040
TA3-1 89366
TC3-1 93720
K3-5 94143
TC3-5 94270
TC3-3 96753
TC2-3 99245
K2-5 99858
K2-2 102024
TA2-4 102783
TC2-5 104870
K2-1 105205
K3-2 107109
TC2-1 107204
TA2-5 109527
TB2-4 110925
TA2-3 114543
K2-4 115110
TC2-2 115747
TB2-3 117046
K2-3 117706
TB2-2 120575
TB2-1 126947
TC2-4 127670
0e+00
5e+04
1e+05
Read Counts

2) DADA2 processing -> ASVs (script dada2.R)

Reads were corrected for errors, merged to form ASV, chimera removed and annotated with Silva database 132 to the genus level following DADA2 protocol.

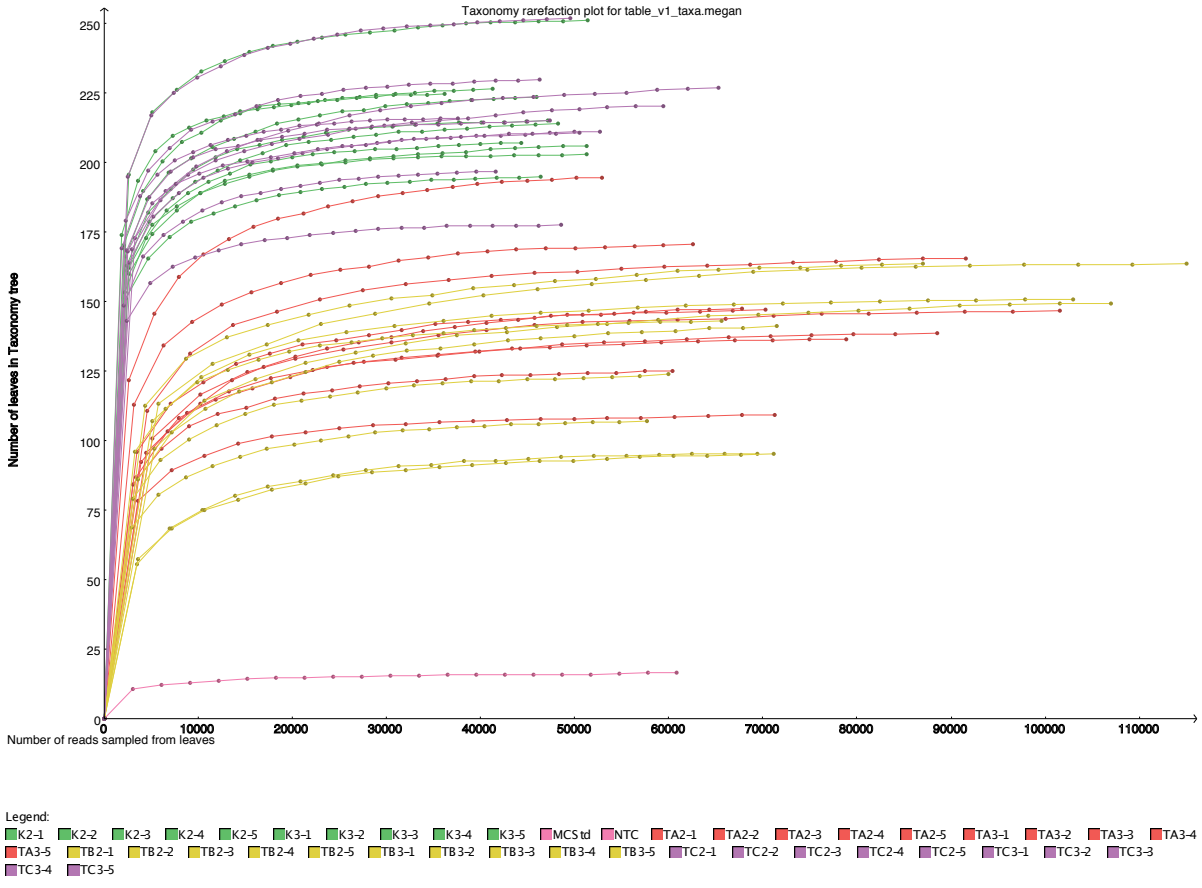
Export files: rep-seqs.fna, seqtab-nochim.txt, seqtab-taxa.txt

3) Convert to biom format with Qiime2

```
source activate qiime2-2019.10
#for conversion of dada2 to biom
qiime tools import --input-path rep-seqs.fna --type 'FeatureData[Sequence]' --output-path rep-seqs.qza
echo -n "#OTU Table" | cat - seqtab-nochim.txt > biom-table.txt
biom convert -i biom-table.txt -o table_v1.biom --table-type="OTU table" --to-json
#invert seqtab-taxa.txt with in-house perl (much faster than R):
./inverter.pl seqtab-taxa.txt > seqtab-taxa_inv.txt
#add "ASV" to first line manually into seqtab-taxa_inv.txt
biom add-metadata -i table_v1.biom -o table_v1_taxa.biom --observation-metadata-fp seqtab-taxa_inv.txt --observation-header OTUID,taxonomy --sc-separated taxonomy
```

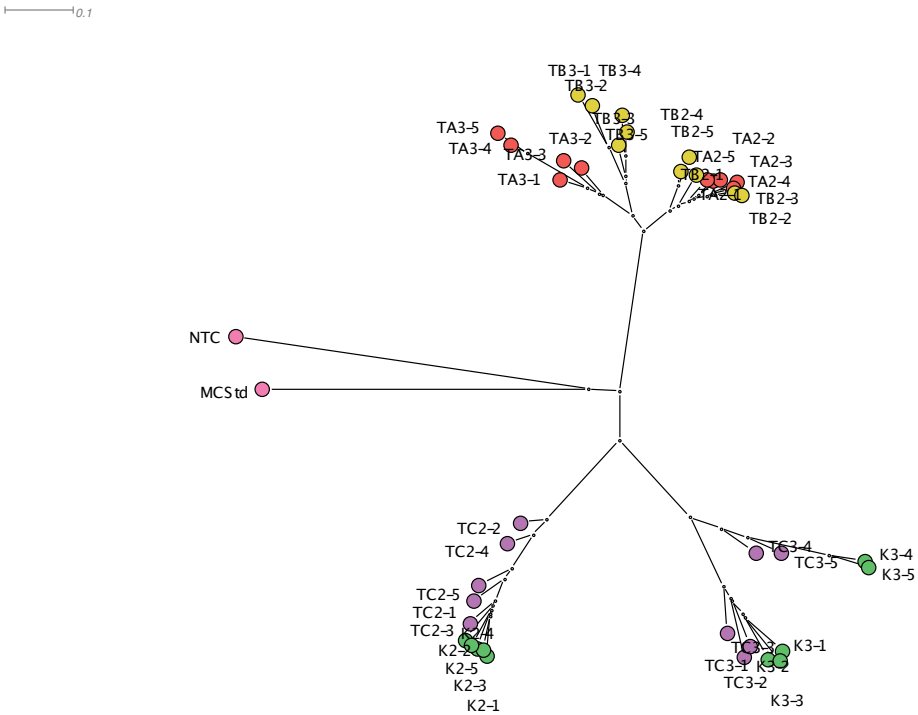
4) Analysis with MEGAN

Import table_v1_taxa.biom and metadata.txt, then calculate rarefaction curves

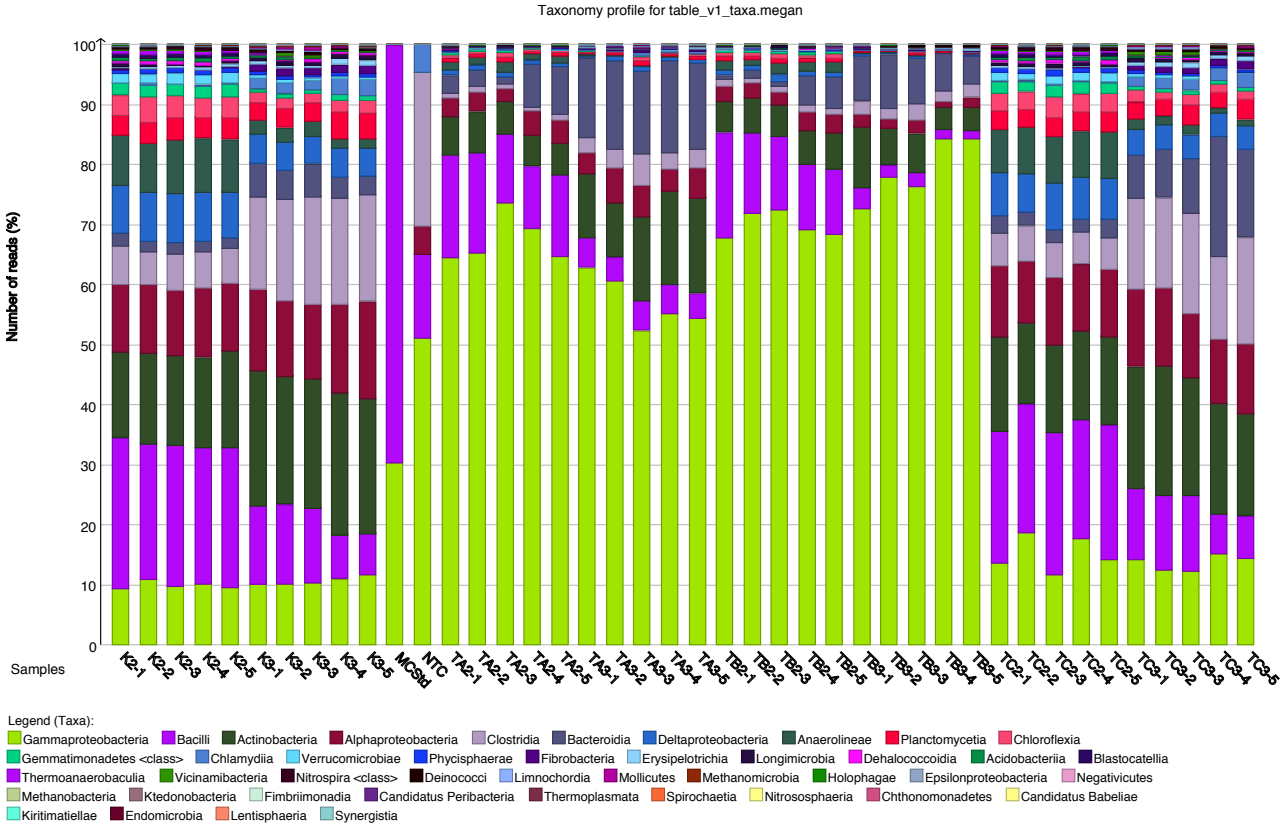


Calculate the PCoA, with Bray-Curtis distance

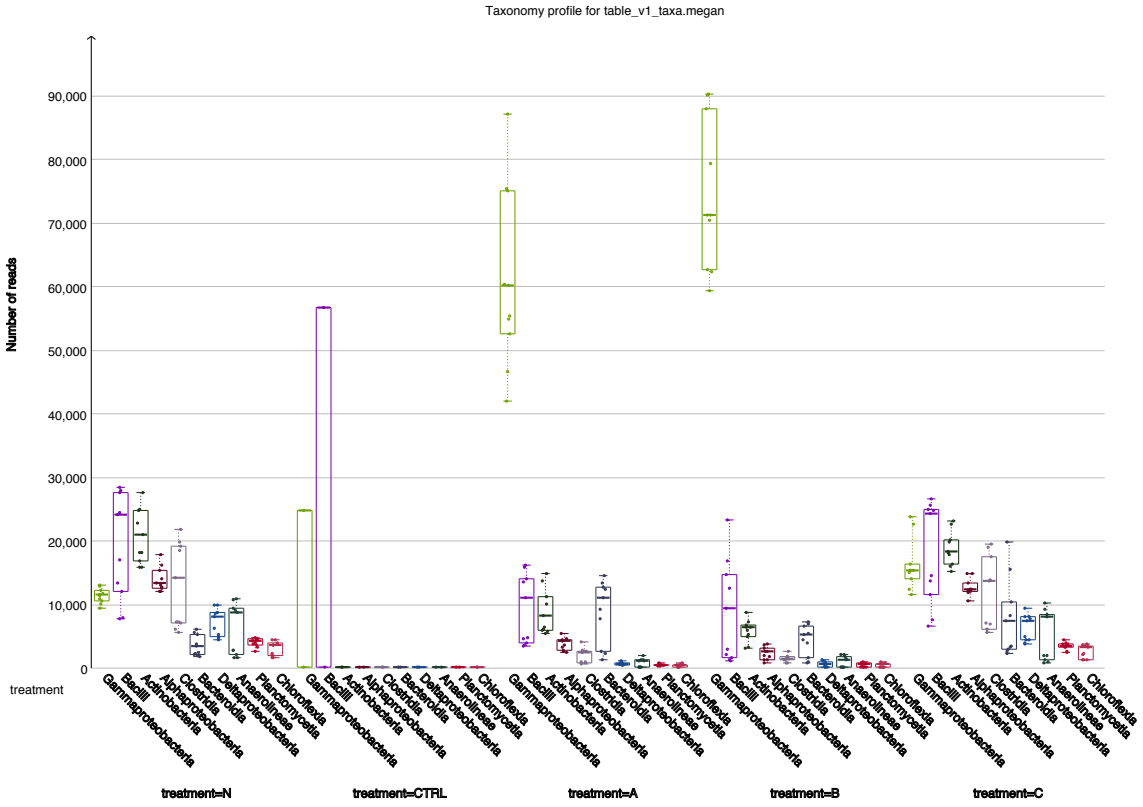
The Neighbour-Joining tree.



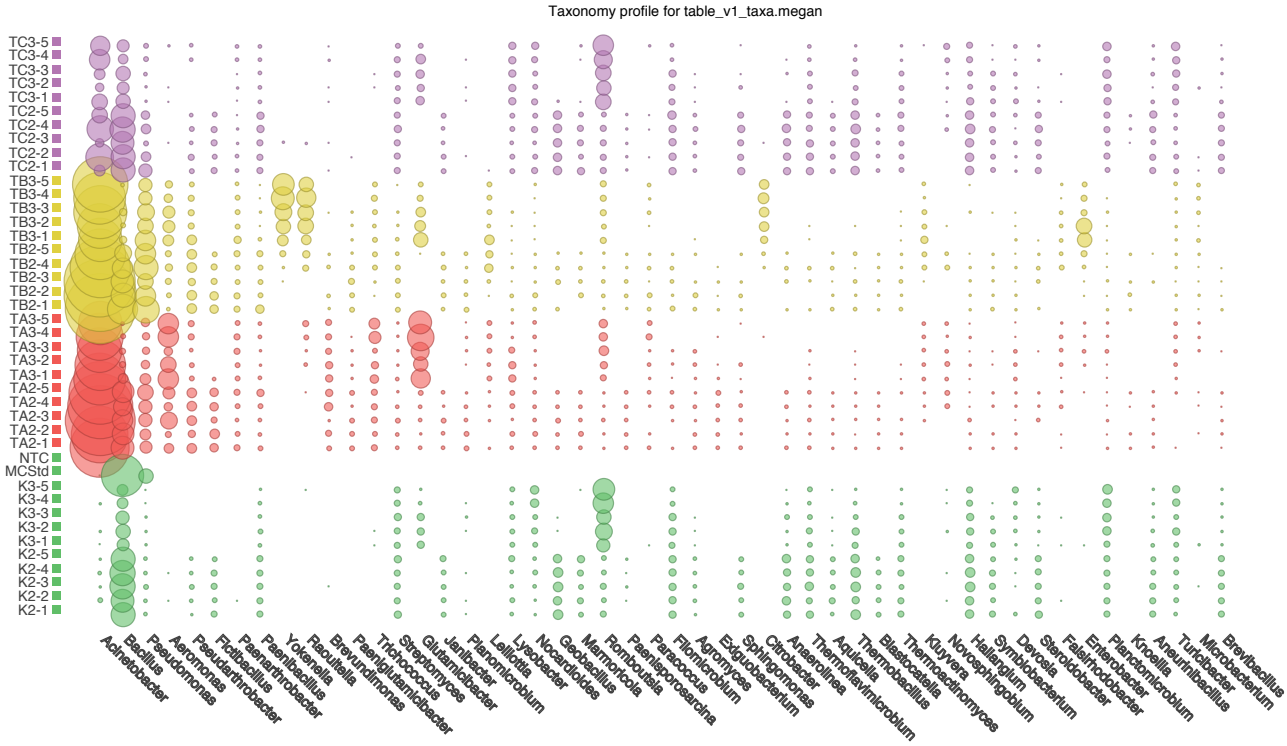
The stacked percentage by taxonomy class



The class profile by treatment



The bubble plot of the top50 most abundant genus overall samples (from left to right):



5) Qiime2 was used to build tree of the representative sequences (ASV)

source activate qiime2-2019.10

qiime alignment mafft --i-sequences rep-seqs.qza --o-alignment mafft-rep-seqs.gza

#Then, eliminate the highly variable positions, to avoid overestimate distances:

qiime alignment mask --i-alignment mafft-rep-seqs.qza --o-masked-alignment masked-msa-rep-seqs.qza

#Finaly, build the ML tree with FastTree:

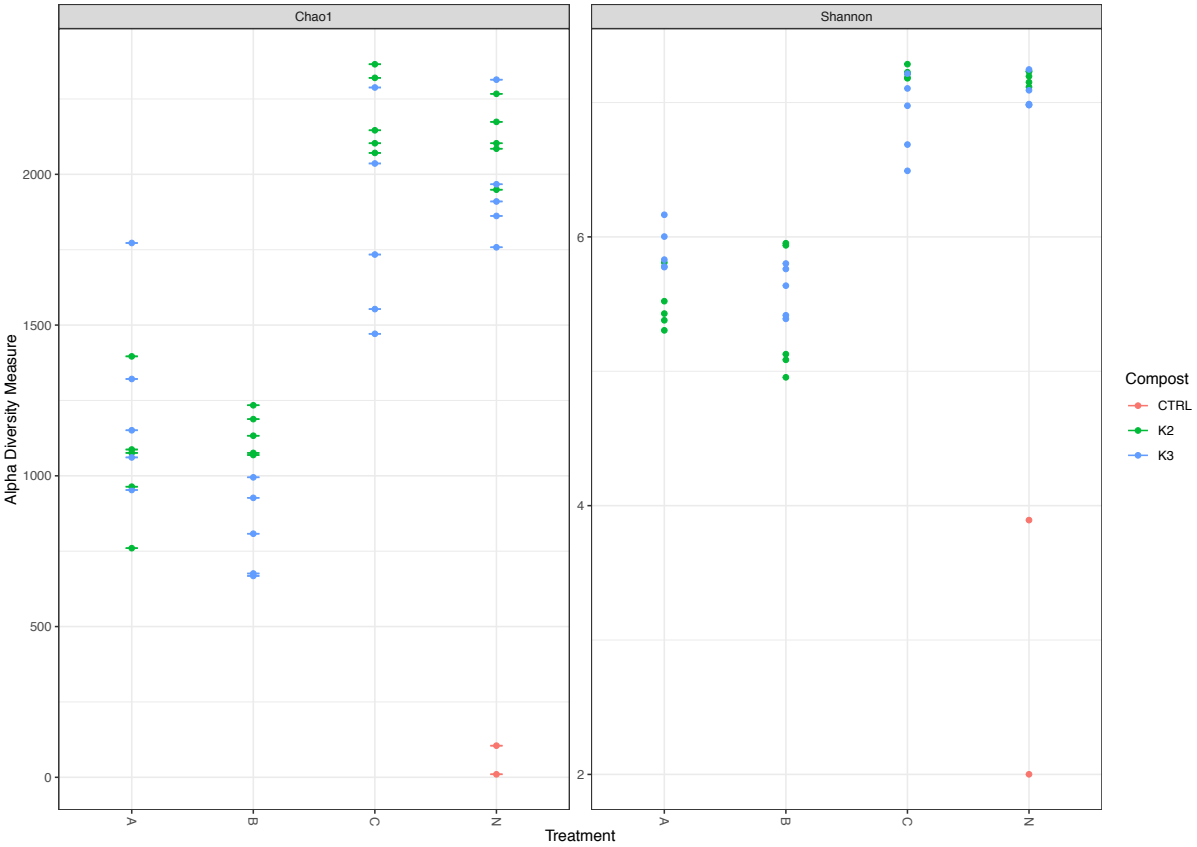
qiime phylogeny fasttree --i-alignment masked-msa-rep-seqs.qza --o-tree unroot-ml-tree-masked.qza

#Additionally, root your unrooted tree based on midpoint rooting method:

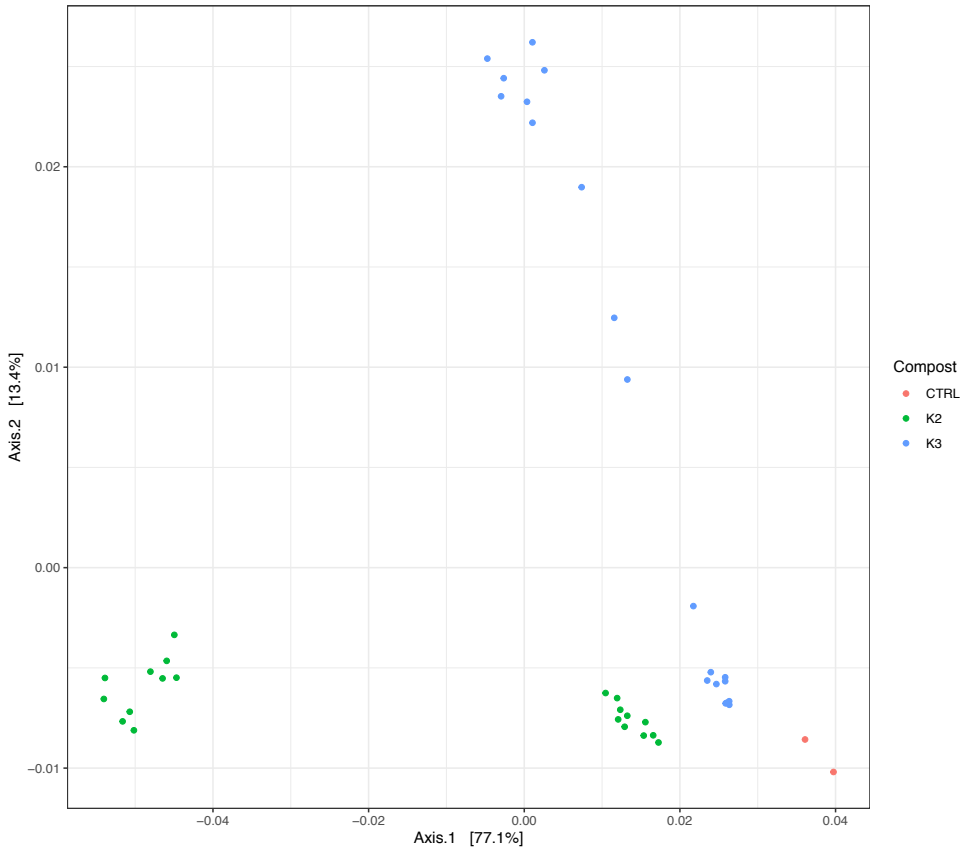
qiime phylogeny midpoint-root --i-tree unroot-ml-tree-masked.qza --o-rooted-tree root-ml-tree.qza

qiime tools export --input-path root-ml-tree.qza --output-path root-ml-tree.nwk

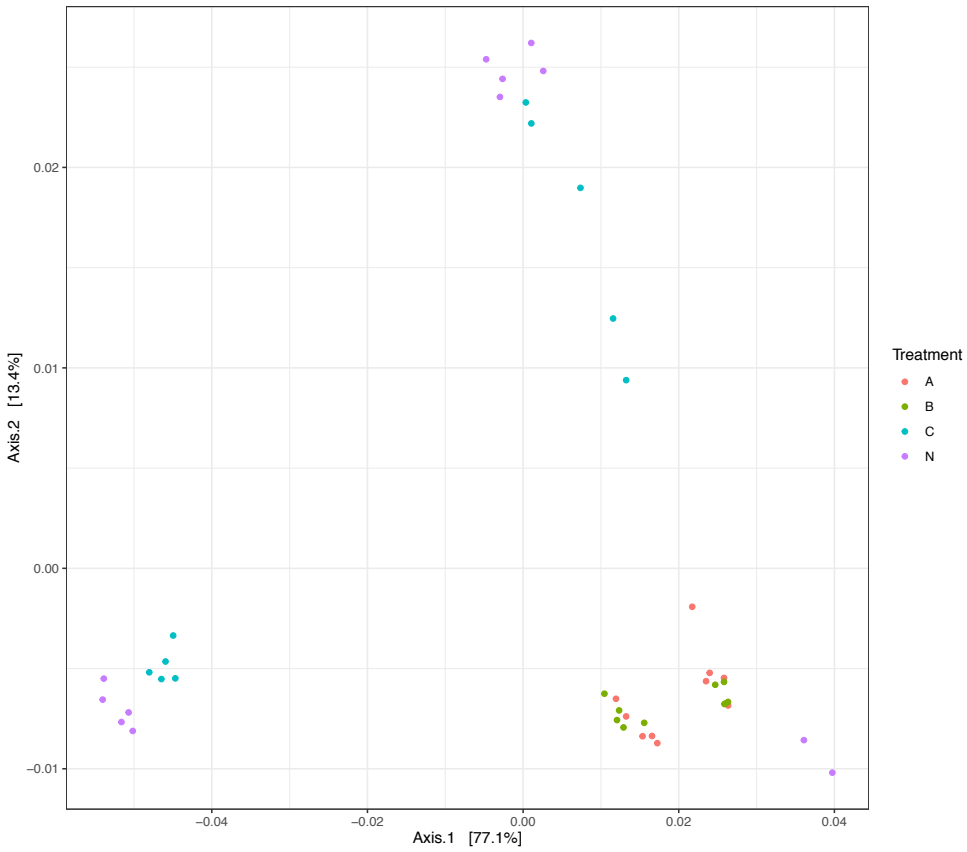
6) Import into phyloseq object (script phyloseq.R)
7) Calculate plot richness alpha diversity



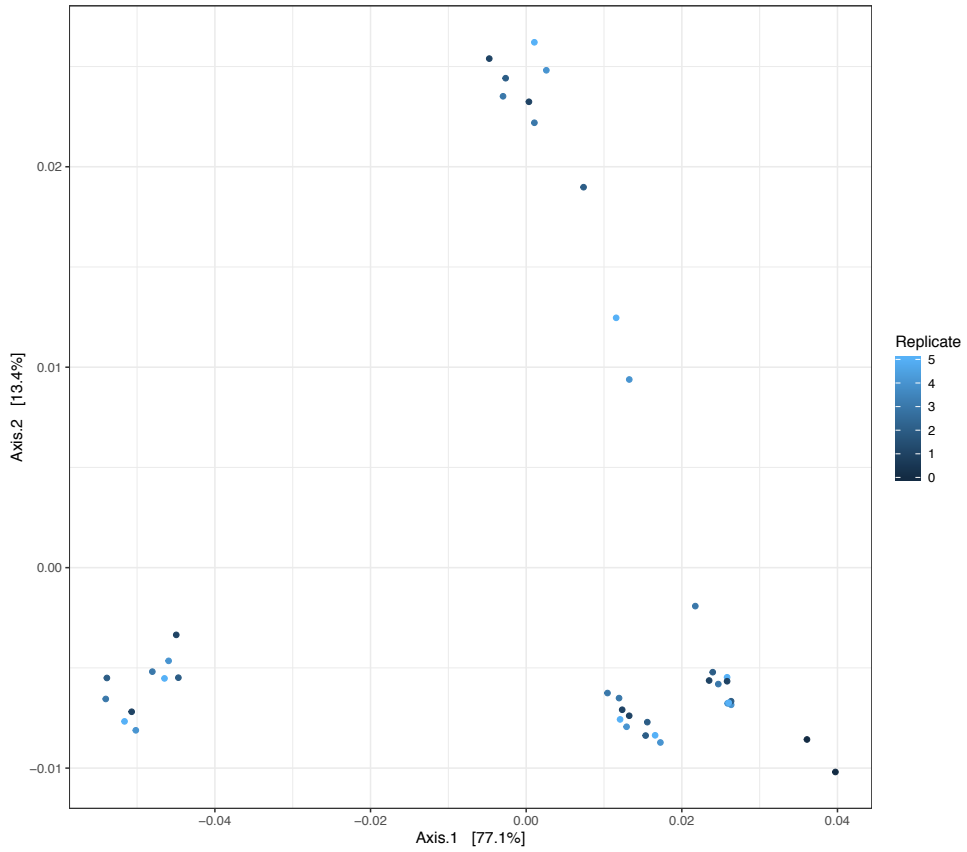
8) Calculate beta diversity (wunifrac, PCoA) and comparisons with PERMANOVA Composts



Treatments



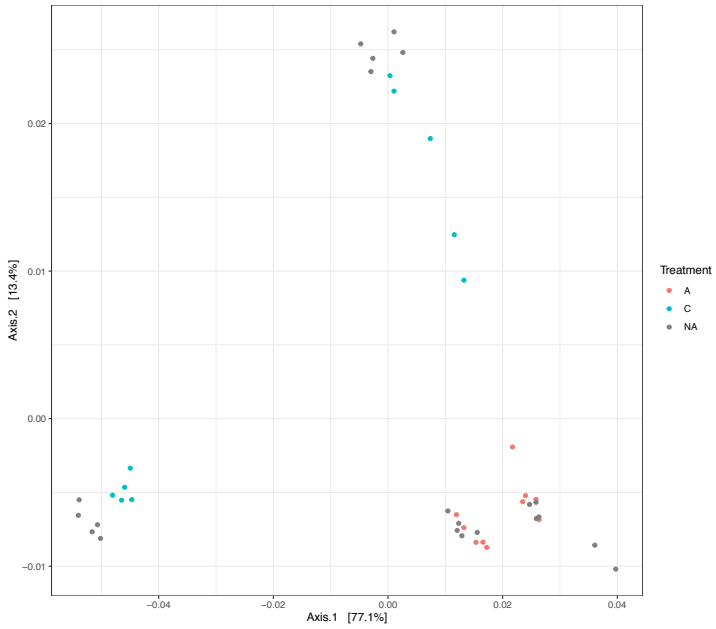
Replicates



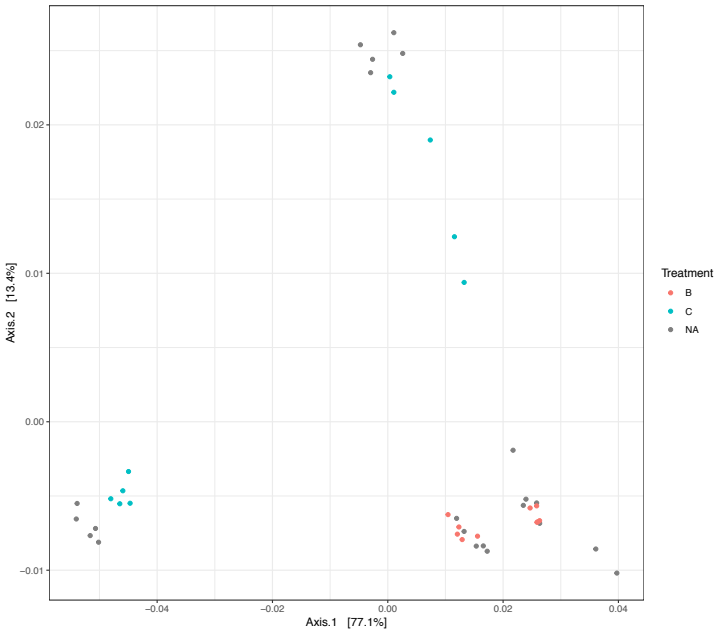
PERMANOVA comparison

Sample	Df	SumOfSqs	MeanSqs	F.Model	R2	Pr(>F)	Signif
sample_data(sichh)$Replicate	1	0.000504	0.00050394	0.44034	0.01089	0.612	
Residuals	40	0.045777	0.00114442		0.98911		
Total	41	0.046281			1.00000		
sample_data(sichh)$Treatment	3	0.016564	0.0055213	7.0603	0.3579	0.001	***
Residuals	38	0.029717	0.0007820		0.6421		
Total	41	0.046281			1.0000		
sample_data(sichh)$Compost	2	0.015341	0.0076704	9.6685	0.33147	0.001	***
Residuals	39	0.030940	0.0007933		0.66853		
Total	41	0.046281			1.00000		
Compare only A, B, C							
sample_data(AvsB)$Treatment	1	0.00014621	1.4621e-04	1.5106	0.07742	0.17	
Residuals	18	0.00174229	9.6794e-05		0.92258		
Total	19	0.00188850			1.00000		
sample_data(AvsC)$Treatment	1	0.0086708	0.0086708	14.67	0.44904	0.001	***
Residuals	18	0.0106387	0.0005910		0.55096		
Total	19	0.0193095			1.00000		
sample_data(BvsC)$Treatment	1	0.0085147	0.0085147	14.329	0.44323	0.001	***
Residuals	18	0.0106959	0.0005942		0.55677		
Total	19	0.0192106			1.00000		
Signif. codes: 0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1							

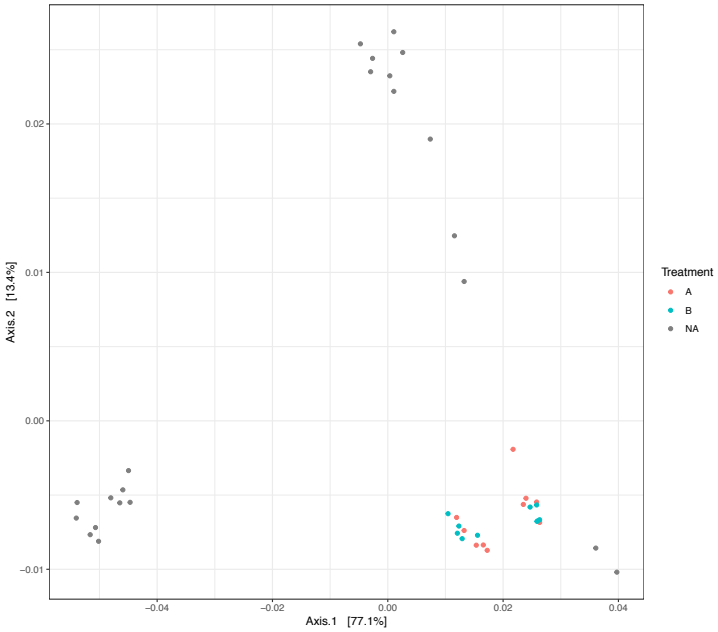
A vs C



B vs C



A vs B



Conclusion

The results explain a clear separation of the microbiomes of the 2 starting composts K2 and K3. Both treatments with nutrient A or nutrient B create a shift of the microbiome leading to a reduction of the diversity with a convergence of the treated compost microbiomes. The genus Acinetobacter (class gammaproteobacteria) is highly dominant in the treated composts. The 5 replicates show a good reproducibility in all cases.

References

FastQC
URL: https://www.bioinformatics.babraham.ac.uk/projects/fastqc/

Dada2
Callahan B, McMurdie P, Rosen M et al. DADA2: High-resolution sample inference from Illumina amplicon data. Nat Methods 13, 581–583 (2016). https://doi.org/10.1038/nmeth.3869

Phyloseq
McMurdie PJ, Holmes S (2013) phyloseq: An R Package for Reproducible Interactive Analysis and Graphics of Microbiome Census Data. PLoS ONE 8(4): e61217. https://doi.org/10.1371/journal.pone.0061217

Qiime2
Bolyen E, Rideout JR, et al. 2019. Reproducible, interactive, scalable and extensible microbiome data science using QIIME 2. Nature Biotechnology 37: 852–857. https://doi.org/10.1038/s41587-019-0209-9

MEGAN
Huson, D, Beier, S, Flade, I, Gorska, A, El-Hadidi, M, Mitra, S, Ruscheweyh, H, and Rewati Tappu, D (2016). MEGAN Community Edition - Interactive exploration and analysis of large-scale microbiome sequencing data. PLoS Computational Biology, 12(6):e1004957

Silva database
Quast C, Pruesse E, Yilmaz P, Gerken J, Schweer T, Yarza P, Peplies J, Glöckner FO (2013) The SILVA ribosomal RNA gene database project: improved data processing and web-based tools. Opens external link in new window. Nucl. Acids Res. 41 (D1): D590-D596.

Primers V3-V4 (kindworth)
Klindworth A, Pruesse E, Schweer T, Peplies J, Quast C, Horn M, Glöckner FO. Evaluation of general 16S ribosomal RNA gene PCR primers for classical and next-generation sequencing-based diversity studies, Nucleic Acids Research, Volume 41, Issue 1, 1 January 2013, Page e1, https://doi.org/10.1093/nar/gks808