

Sequence Pretreatment and Quality Analysis (FastQC & Trimmomatic)

domingo, 25 de julio de 2021 14:17

1. Rawdata fastqc (R1 and R2)
2. Pretreatment with Trimmomatic
3. Join of R1 and R2
4. FastQC results visualization
5. Count of readings at each step

USED PARAMETERS:

- **SLIDINGWINDOW:** Perform a sliding window trimming, cutting once the average quality within the window falls below a threshold 6:20
- **LEADING:** Cut bases off the start of a read, if below a threshold quality
- **TRAILING:** Cut bases off the end of a read, if below a threshold quality

A) GUT

TRIMMING1:

```
java -jar
/Users/fernanda/verano2021/Trimmomatic-0.39/trimmomatic-0.39.jar PE
gut_sample1_pond1_R1.fastq.gz gut_sample1_pond1_R2.fastq.gz
gut_sample1_pond1_R1_paired.fastq.gz gut_sample1_pond1_R1
_unpaired.fastq.gz gut_sample1_pond1_R2_paired.fastq.gz gut_sample1
_pond1_R2_unpaired.fastq.gz LEADING:20 TRAILING:20
SLIDINGWINDOW:6:20
```

QUALITY:

Quality encoding detected as phred33
Input Read Pairs: 46133 Both Surviving: 28319 (61.39%) Forward Only
Surviving: 17736 (38.45%) Reverse Only Surviving: 5 (0.01%)
Dropped: 73 (0.16%)

JOIN:

```
fastq-join gut_sample1_pond1_R1_paired.fastq.gz gut_sample1_pond1
_R2_paired.fastq.gz -o gjoinedsample1p1.fastq
Total reads: 28319
Total joined: 13515
Average join len: 65.32
Stdev join len: 32.34
```

TRIMMING2:

```
java -jar
/Users/fernanda/verano2021/Trimmomatic-0.39/trimmomatic-0.39.jar PE
gut_sample1_pond2_R1.fastq.gz gut_sample1_pond2_R2.fastq.gz
gut_sample1_pond2_R1_paired.fastq.gz gut_sample1_pond2_R1
_unpaired.fastq.gz gut_sample1_pond2_R2_paired.fastq.gz gut_sample1
_pond2_R2_unpaired.fastq.gz LEADING:20 TRAILING:20
SLIDINGWINDOW:6:20
```

QUALITY:

Quality encoding detected as phred33

Input Read Pairs: 56184 Both Surviving: 37159 (66.14%) Forward Only
Surviving: 18908 (33.65%) Reverse Only Surviving: 12 (0.02%)

Dropped: 105 (0.19%)

JOIN:

```
fastq-join gut_sample1_pond2_R1_paired.fastq.gz gut_sample1_pond2  
_R2_paired.fastq.gz -o gjoinedsample1p2.fastq
```

Total reads: 37159

Total joined: 20069

Average join len: 66.88

Stdev join len: 42.87

TRIMMING3:

```
java -jar
```

```
/Users/fernanda/verano2021/Trimmomatic-0.39/trimmomatic-0.39.jar PE  
gut_sample1_pond3_R1.fastq.gz gut_sample1_pond3_R2.fastq.gz  
gut_sample1_pond3_R1_paired.fastq.gz gut_sample1_pond3_R1  
_unpaired.fastq.gz gut_sample1_pond3_R2_paired.fastq.gz gut_sample1  
_pond3_R2_unpaired.fastq.gz LEADING:20 TRAILING:20  
SLIDINGWINDOW:6:20
```

QUALITY:

Quality encoding detected as phred33

Input Read Pairs: 85007 Both Surviving: 52925 (62.26%) Forward Only
Surviving: 31933 (37.57%) Reverse Only Surviving: 14 (0.02%)

Dropped: 135 (0.16%)

JOIN:

```
fastq-join gut_sample1_pond3_R1_paired.fastq.gz gut_sample1_pond3  
_R2_paired.fastq.gz -o gjoinedsample1p3.fastq
```

Total reads: 52925

Total joined: 24730

Average join len: 65.57

Stdev join len: 25.12

TRIMMING4:

```
java -jar
```

```
/Users/fernanda/verano2021/Trimmomatic-0.39/trimmomatic-0.39.jar PE  
gut_sample2_pond1_R1.fastq.gz gut_sample2_pond1_R2.fastq.gz  
gut_sample2_pond1_R1_paired.fastq.gz gut_sample2_pond1_R1  
_unpaired.fastq.gz gut_sample2_pond1_R2_paired.fastq.gz gut_sample2  
_pond1_R2_unpaired.fastq.gz LEADING:20 TRAILING:20  
SLIDINGWINDOW:6:20
```

QUALITY:

Quality encoding detected as phred33

Input Read Pairs: 106896 Both Surviving: 71991 (67.35%) Forward Only Surviving: 34609 (32.38%) Reverse Only Surviving: 21 (0.02%) Dropped: 275 (0.26%)

JOIN:

```
fastq-join gut_sample2_pond1_R1_paired.fastq.gz gut_sample2_pond1_R2_paired.fastq.gz -o gjoinedsample2p1.fastq
          Total reads: 71991
          Total joined: 43176
    Average join len: 57.24
      Stdev join len: 20.06
```

TRIMMING5:

```
java -jar
/Users/fernanda/verano2021/Trimmomatic-0.39/trimmomatic-0.39.jar PE
gut_sample2_pond2_R1.fastq.gz gut_sample2_pond2_R2.fastq.gz
gut_sample2_pond2_R1_paired.fastq.gz gut_sample2_pond2_R1_unpaired.fastq.gz
gut_sample2_pond2_R2_paired.fastq.gz gut_sample2_pond2_R2_unpaired.fastq.gz
LEADING:20 TRAILING:20
SLIDINGWINDOW:6:20
```

QUALITY:

Quality encoding detected as phred33

Input Read Pairs: 63686 Both Surviving: 39779 (62.46%) Forward Only Surviving: 23788 (37.35%) Reverse Only Surviving: 5 (0.01%) Dropped: 114 (0.18%)

JOIN:

```
fastq-join gut_sample2_pond2_R1_paired.fastq.gz gut_sample2_pond2_R2_paired.fastq.gz -o gjoinedsample2p2.fastq
          Total reads: 39779
          Total joined: 20099
    Average join len: 57.78
      Stdev join len: 20.54
```

TRIMMING6:

```
java -jar
/Users/fernanda/verano2021/Trimmomatic-0.39/trimmomatic-0.39.jar PE
gut_sample2_pond3_R1.fastq.gz gut_sample2_pond3_R2.fastq.gz
gut_sample2_pond3_R1_paired.fastq.gz gut_sample2_pond3_R1_unpaired.fastq.gz
gut_sample2_pond3_R2_paired.fastq.gz gut_sample2_pond3_R2_unpaired.fastq.gz
LEADING:20 TRAILING:20
SLIDINGWINDOW:6:20
```

QUALITY:

Quality encoding detected as phred33

Input Read Pairs: 150685 Both Surviving: 95123 (63.13%) Forward Only Surviving: 55371 (36.75%) Reverse Only Surviving: 22 (0.01%)

Dropped: 169 (0.11%)

JOIN:

```
fastq-join gut_sample2_pond3_R1_paired.fastq.gz gut_sample2_pond3
_R2_paired.fastq.gz -o gjoinedsample2p3.fastq
      Total reads: 95123
      Total joined: 46294
Average join len: 54.78
Stdev join len: 19.34
```

TRIMMING7:

```
java -jar
/Users/fernanda/verano2021/Trimmomatic-0.39/trimmomatic-0.39.jar PE
gut_sample3_pond1_R1.fastq.gz gut_sample3_pond1_R2.fastq.gz
gut_sample3_pond1_R1_paired.fastq.gz gut_sample3_pond1_R1
_unpaired.fastq.gz gut_sample3_pond1_R2_paired.fastq.gz gut_sample3
_pond1_R2_unpaired.fastq.gz LEADING:20 TRAILING:20
SLIDINGWINDOW:6:20
```

QUALITY:

Quality encoding detected as phred33
Input Read Pairs: 219766 Both Surviving: 137019 (62.35%) Forward
Only Surviving: 82121 (37.37%) Reverse Only Surviving: 47 (0.02%)
Dropped: 579 (0.26%)

JOIN:

```
fastq-join gut_sample3_pond1_R1_paired.fastq.gz gut_sample3_pond1
_R2_paired.fastq.gz -o gjoinedsample3p1.fastq
      Total reads: 137019
      Total joined: 60769
Average join len: 67.13
Stdev join len: 25.77
```

TRIMMING8:

```
java -jar
/Users/fernanda/verano2021/Trimmomatic-0.39/trimmomatic-0.39.jar PE
gut_sample3_pond2_R1.fastq.gz gut_sample3_pond2_R2.fastq.gz
gut_sample3_pond2_R1_paired.fastq.gz gut_sample3_pond2_R1
_unpaired.fastq.gz gut_sample3_pond2_R2_paired.fastq.gz gut_sample3
_pond2_R2_unpaired.fastq.gz LEADING:20 TRAILING:20
SLIDINGWINDOW:6:20
```

QUALITY:

Quality encoding detected as phred33
Input Read Pairs: 69709 Both Surviving: 45117 (64.72%) Forward Only
Surviving: 24406 (35.01%) Reverse Only Surviving: 6 (0.01%)
Dropped: 180 (0.26%)

JOIN:

```
fastq-join gut_sample3_pond2_R1_paired.fastq.gz gut_sample3_pond2
```

```
_R2_paired.fastq.gz -o gjoinedsample3p2.fastq
      Total reads: 45117
      Total joined: 25190
Average join len: 57.24
      Stdev join len: 21.31
```

TRIMMING9:

```
java -jar
/Users/fernanda/verano2021/Trimmomatic-0.39/trimmomatic-0.39.jar PE
gut_sample3_pond3_R1.fastq.gz gut_sample3_pond3_R2.fastq.gz
gut_sample3_pond3_R1_paired.fastq.gz gut_sample3_pond3_R1
_unpaired.fastq.gz gut_sample3_pond3_R2_paired.fastq.gz gut_sample3
_pond3_R2_unpaired.fastq.gz LEADING:20 TRAILING:20
SLIDINGWINDOW:6:20
```

QUALITY:

```
Quality encoding detected as phred33
Input Read Pairs: 38466 Both Surviving: 25000 (64.99%) Forward Only
Surviving: 13338 (34.67%) Reverse Only Surviving: 15 (0.04%)
Dropped: 113 (0.29%)
```

JOIN:

```
fastq-join gut_sample3_pond3_R1_paired.fastq.gz gut_sample3_pond3
_R2_paired.fastq.gz -o gjoinedsample3p3.fastq
      Total reads: 25000
      Total joined: 13215
Average join len: 58.82
      Stdev join len: 25.84
```

B) HEP

TRIMMING1:

```
java -jar
/Users/fernanda/verano2021/Trimmomatic-0.39/trimmomatic-0.39.jar PE
hep_sample1_pond1_R1.fastq.gz hep_sample1_pond1_R2.fastq.gz
hep_sample1_pond1_R1_paired.fastq.gz hep_sample1_pond1_R1
_unpaired.fastq.gz hep_sample1_pond1_R2_paired.fastq.gz hep_sample1
_pond1_R2_unpaired.fastq.gz LEADING:20 TRAILING:20
SLIDINGWINDOW:6:20
```

QUALITY:

```
Quality encoding detected as phred33
Input Read Pairs: 49816 Both Surviving: 47095 (94.54%) Forward Only
Surviving: 2699 (5.42%) Reverse Only Surviving: 13 (0.03%) Dropped:
9 (0.02%)
```

JOIN:

```
fastq-join hep_sample1_pond1_R1.fastq.gz hep_sample1_pond1
_R2.fastq.gz -o hjoinedsample1p1.fastq
      Total reads: 49816
      Total joined: 36876
Average join len: 44.88
      Stdev join len: 26.60
```

TRIMMING2:

```
java -jar
/Users/fernanda/verano2021/Trimmomatic-0.39/trimmomatic-0.39.jar PE
hep_sample1_pond2_R1.fastq.gz hep_sample1_pond2_R2.fastq.gz
hep_sample1_pond2_R1_paired.fastq.gz hep_sample1_pond2_R1
_unpaired.fastq.gz hep_sample1_pond2_R2_paired.fastq.gz hep_sample1
_pond2_R2_unpaired.fastq.gz LEADING:20 TRAILING:20
SLIDINGWINDOW:6:20
```

QUALITY:

Quality encoding detected as phred33
Input Read Pairs: 42005 Both Surviving: 40217 (95.74%) Forward Only
Surviving: 1778 (4.23%) Reverse Only Surviving: 9 (0.02%) Dropped:
1 (0.00%)

JOIN:

```
fastq-join hep_sample1_pond2_R1.fastq.gz hep_sample1_pond2
_R2.fastq.gz -o hjoinedsample1p2.fastq
Total reads: 42005
Total joined: 30773
Average join len: 40.75
Stdev join len: 17.40
```

TRIMMING3:

```
java -jar
/Users/fernanda/verano2021/Trimmomatic-0.39/trimmomatic-0.39.jar PE
hep_sample1_pond3_R1.fastq.gz hep_sample1_pond3_R2.fastq.gz
hep_sample1_pond3_R1_paired.fastq.gz hep_sample1_pond3_R1
_unpaired.fastq.gz hep_sample1_pond3_R2_paired.fastq.gz hep_sample1
_pond3_R2_unpaired.fastq.gz LEADING:20 TRAILING:20
SLIDINGWINDOW:6:20
```

QUALITY:

Quality encoding detected as phred33
Input Read Pairs: 128116 Both Surviving: 91429 (71.36%) Forward
Only Surviving: 36452 (28.45%) Reverse Only Surviving: 30 (0.02%)
Dropped: 205 (0.16%)

JOIN:

```
fastq-join hep_sample1_pond3_R1.fastq.gz hep_sample1_pond3
_R2.fastq.gz -o hjoinedsample1p3.fastq
Total reads: 128116
Total joined: 70032
Average join len: 66.79
Stdev join len: 23.87
```

TRIMMING4:

```
java -jar
/Users/fernanda/verano2021/Trimmomatic-0.39/trimmomatic-0.39.jar PE
```

```
hep_sample2_pond1_R1.fastq.gz hep_sample2_pond1_R2.fastq.gz
hep_sample2_pond1_R1_paired.fastq.gz hep_sample2_pond1_R1
_unpaired.fastq.gz hep_sample2_pond1_R2_paired.fastq.gz hep_sample2
_pond1_R2_unpaired.fastq.gz LEADING:20 TRAILING:20
SLIDINGWINDOW:6:20
```

QUALITY:

Quality encoding detected as phred33
Input Read Pairs: 47095 Both Surviving: 45134 (95.84%) Forward Only
Surviving: 1949 (4.14%) Reverse Only Surviving: 8 (0.02%) Dropped:
4 (0.01%)

JOIN:

```
fastq-join hep_sample2_pond1_R1.fastq.gz hep_sample2_pond1
_R2.fastq.gz -o hjoinedsample2p1.fastq
Total reads: 47095
Total joined: 32689
Average join len: 47.65
Stdev join len: 22.53
```

TRIMMING5:

```
java -jar
/Users/fernanda/verano2021/Trimmomatic-0.39/trimmomatic-0.39.jar PE
hep_sample2_pond2_R1.fastq.gz hep_sample2_pond2_R2.fastq.gz
hep_sample2_pond2_R1_paired.fastq.gz hep_sample2_pond2_R1
_unpaired.fastq.gz hep_sample2_pond2_R2_paired.fastq.gz hep_sample2
_pond2_R2_unpaired.fastq.gz LEADING:20 TRAILING:20
SLIDINGWINDOW:6:20
```

QUALITY:

Quality encoding detected as phred33
Input Read Pairs: 44696 Both Surviving: 42630 (95.38%) Forward Only
Surviving: 2060 (4.61%) Reverse Only Surviving: 4 (0.01%) Dropped:
2 (0.00%)

JOIN:

```
fastq-join hep_sample2_pond2_R1.fastq.gz hep_sample2_pond2
_R2.fastq.gz -o hjoinedsample2p2.fastq
Total reads: 44696
Total joined: 32977
Average join len: 44.71
Stdev join len: 20.75
```

TRIMMING6:

```
java -jar
/Users/fernanda/verano2021/Trimmomatic-0.39/trimmomatic-0.39.jar PE
hep_sample2_pond3_R1.fastq.gz hep_sample2_pond3_R2.fastq.gz
hep_sample2_pond3_R1_paired.fastq.gz hep_sample2_pond3_R1
_unpaired.fastq.gz hep_sample2_pond3_R2_paired.fastq.gz hep_sample2
_pond3_R2_unpaired.fastq.gz LEADING:20 TRAILING:20
SLIDINGWINDOW:6:20
```

QUALITY:

Quality encoding detected as phred33
Input Read Pairs: 81647 Both Surviving: 78422 (96.05%) Forward Only
Surviving: 3163 (3.87%) Reverse Only Surviving: 31 (0.04%) Dropped:
31 (0.04%)

JOIN:

```
fastq-join hep_sample2_pond3_R1.fastq.gz hep_sample2_pond3  
_R2.fastq.gz -o hjoinedsample2p3.fastq
```

Total reads: 81647
Total joined: 64527
Average join len: 56.10
Stdev join len: 39.23

TRIMMING7:

```
java -jar  
/Users/fernanda/verano2021/Trimmomatic-0.39/trimmomatic-0.39.jar PE  
hep_sample3_pond1_R1.fastq.gz hep_sample3_pond1_R2.fastq.gz  
hep_sample3_pond1_R1_paired.fastq.gz hep_sample3_pond1_R1  
_unpaired.fastq.gz hep_sample3_pond1_R2_paired.fastq.gz hep_sample3  
_pond1_R2_unpaired.fastq.gz LEADING:20 TRAILING:20  
SLIDINGWINDOW:6:20
```

QUALITY:

Quality encoding detected as phred33
Input Read Pairs: 36514 Both Surviving: 35281 (96.62%) Forward Only
Surviving: 1229 (3.37%) Reverse Only Surviving: 4 (0.01%) Dropped:
0 (0.00%)

JOIN:

```
fastq-join hep_sample3_pond1_R1.fastq.gz hep_sample3_pond1  
_R2.fastq.gz -o hjoinedsample3p1.fastq
```

Total reads: 36514
Total joined: 26580
Average join len: 45.62
Stdev join len: 21.03

TRIMMING8:

```
java -jar  
/Users/fernanda/verano2021/Trimmomatic-0.39/trimmomatic-0.39.jar PE  
hep_sample3_pond2_R1.fastq.gz hep_sample3_pond2_R2.fastq.gz  
hep_sample3_pond2_R1_paired.fastq.gz hep_sample3_pond2_R1  
_unpaired.fastq.gz hep_sample3_pond2_R2_paired.fastq.gz hep_sample3  
_pond2_R2_unpaired.fastq.gz LEADING:20 TRAILING:20  
SLIDINGWINDOW:6:20
```

QUALITY:

Quality encoding detected as phred33
Input Read Pairs: 48415 Both Surviving: 46103 (95.22%) Forward Only
Surviving: 2300 (4.75%) Reverse Only Surviving: 9 (0.02%) Dropped:
3 (0.01%)

JOIN:

```
fastq-join hep_sample3_pond2_R1.fastq.gz hep_sample3_pond2_R2.fastq.gz -o hjoinedsample3p2.fastq
```

```
Total reads: 48415
Total joined: 34946
Average join len: 42.93
Stdev join len: 18.29
```

TRIMMING9:

```
java -jar
/Users/fernanda/verano2021/Trimmomatic-0.39/trimmomatic-0.39.jar PE
hep_sample3_pond3_R1.fastq.gz hep_sample3_pond3_R2.fastq.gz
hep_sample3_pond3_R1_paired.fastq.gz hep_sample3_pond3_R1_unpaired.fastq.gz
hep_sample3_pond3_R2_paired.fastq.gz hep_sample3_pond3_R2_unpaired.fastq.gz
LEADING:20 TRAILING:20
SLIDINGWINDOW:6:20
```

QUALITY:

```
Quality encoding detected as phred33
Input Read Pairs: 43833 Both Surviving: 41177 (93.94%) Forward Only
Surviving: 2642 (6.03%) Reverse Only Surviving: 11 (0.03%) Dropped:
3 (0.01%)
```

JOIN:

```
fastq-join hep_sample3_pond3_R1.fastq.gz hep_sample3_pond3_R2.fastq.gz -o hjoinedsample3p3.fastq
```

```
Total reads: 43833
Total joined: 30528
Average join len: 50.42
Stdev join len: 29.45
```

QIIME 2 (Quantitative Insights into Microbial Ecology, 2016)

sábado, 24 de julio de 2021

14:36

1. The first thing we have to do is get the data into an artifact. We can import the data with the import action from the tools. For that we have to give QIIME 2 a manifest (list of raw files) and tell it what *type of data* we are importing and what type of artifact we want.

```
qiime tools import --type SampleData[SequencesWithQuality] --  
input-path se-33-manifest2.0.csv --output-path   
se-33demux2.0.qza --input-format SingleEndFastqManifestPhred33
```



We selected *type* **[SequencesWithQuality]** due to we did a trimming and filtering to our joined samples in the previous step.

Artifact: se-33demux2.0.qza (output-path)

2. We will run the DADA2 plugin which will do 3 things:

- Filter and trim the reads
- Find the most likely set of unique sequences in the sample (ASVs)
- Remove chimeras
- Count the abundances of each ASV

```
Qiime dada2 denoise-single --i-demultiplexed-seqs   
se-33demux2.0.qza --p-trim left 0 --p-trunc-len 106 --o-  
representative-sequences rep_seqs_dada2.qza --o-table   
table_da2.qza --o-denoising-stats stats_da2.qza
```

--i-... = input files
--o-... = output files
--p-... = parameters
--m-... = metadata



3. We generate visual and tabular summaries of a feature table.

```
qiime feature-table summarize --i-table table_da2.qza --o-  
visualization table2.qzv --m-metadata-file metadata_2.txt
```

4. Now we are going to assign taxonomy to our ASVs using a Naive-Bayes approach implemented in the scikit learn Python library and the SILVA database.

```
qiime feature-classifier classify-sklearn --i-classifier   
silva-138-99-nb-classifier.qza --i-reads rep_seqs_dada2.qza --o-  
classification taxonomy_fernanda2.qza --p-n-jobs 64
```

5. QIIME 2 provides a visualizer for viewing metadata in an interactive table, the resulting table of the next command shows the joined metadata files with a column of the feature IDs, a column of the representative sequences, a column of the taxonomic assignments, and lastly, a column of the assignment confidence.

- ```
qiime metadata tabulate --m-input-file rep_seqs_da2.qza --m-input-
file taxonomy_fernanda2.qza --o-visualization tabulated-feature-
metadata2.qzv
```
- ```
qiime tools view metadata2.qzv
```

6. We also are going to use *qiime taxa barplot*. This visualizer generates an interactive barplot visualization of our taxonomies. Interactive features include multi-level sorting, plot recoloring, sample relabeling, and SVG figure export.

```
qiime taxa barplot --i-table table_da2.qza --i-taxonomy
taxonomy_fernanda2.qza --m-metadata-file metadata_2.txt --o-
visualization taxa-bar-plots2.qzv
```

1. Galaxy (LDA SCORE), page 13

2. **Diversity metrics:** In microbial community analysis we are usually interested in two different families of diversity metrics, alpha diversity (ecological diversity within a sample) and beta diversity (ecological differences between samples). FastTree is able to construct phylogenies from large sequence alignments quite rapidly.

```
qiime phylogeny align-to-tree-mafft-fasttree --i-sequences
rep_seqs_da2.qza --o-alignment aligned_rep_seqs2.qza --o-tree
unrooted-tree2.qza --o-rooted-tree rooted-tree2.qza --p-n-threads
12 --o-masked-alignment masked_rep_seqs2.qza
```

```
qiime diversity core-metrics-phylogenetic --i-phylogeny rooted-
tree2.qza --i-table table_da2.qza --p-sampling-depth 5016 --m-
metadata-file metadata_2.txt --output-dir diversity2
```

- Beta diversity:

```
qiime diversity beta-group-significance --i-distance-matrix
unweighted_unifrac_distance_matrix.qza --m-metadata-file metadata_
2.txt --o-visualization anosim_unweighted_unifrac --m-metadata-
column organ --p-method anosim
```

```
qiime diversity beta-group-significance --i-distance-matrix
weighted_unifrac_distance_matrix.qza --m-metadata-file metadata_
2.txt --o-visualization anosim_weighted_unifrac --m-metadata-
column organ --p-method anosim
```

- Alfa diversity:

```
for index in observed_features shannon goods_coverage chao1
simpson ; do echo qiime diversity alpha-rarefaction --i-phylogeny
rooted-tree2.qza --i-table table_da2.qza --p-max-depth 5016 --m-
metadata-file metadata_2.txt --o-visualization $index
\_alpha_diversity.qzv --p-metrics $index ; done
```

- 1- qiime diversity alpha-rarefaction --i-phylogeny rooted-
tree2.qza --i-table table_da2.qza --p-max-depth 5016 --m-metadata-
file metadata_2.txt --o-visualization
observed_features_alpha_diversity2.qzv --p-metrics
observed_features
- 2- qiime diversity alpha-rarefaction --i-phylogeny rooted-
tree2.qza --i-table table_da2.qza --p-max-depth 5016 --m-metadata-
file metadata_2.txt --o-visualization
shannon_alpha_diversity2.qzv --p-metrics shannon
- 3- qiime diversity alpha-rarefaction --i-phylogeny rooted-
tree2.qza --i-table table_da2.qza --p-max-depth 5016 --m-metadata-
file metadata_2.txt --o-visualization
goods_coverage_alpha_diversity2.qzv --p-metrics goods_coverage

✧ We have to run them individually. We will obtain five qzv files, from five alpha metrics.

- 4- `qiime diversity alpha-rarefaction --i-phylogeny rooted-tree2.qza --i-table table_da2.qza --p-max-depth 5016 --m-metadata-file metadata_2.txt --o-visualization chao1_alpha_diversity2.qzv --p-metrics chao1`
- 5- `qiime diversity alpha-rarefaction --i-phylogeny rooted-tree2.qza --i-table table_da2.qza --p-max-depth 5016 --m-metadata-file metadata_2.txt --o-visualization simpson_alpha_diversity2.qzv --p-metrics simpson`
- 1- Install R Studio package in Visual Studio Code. And remake shannon graph, anosim weighted unifracs, and anosim unweighted unifracs graphs

Galaxy (Huttenhower lab Galaxy server)

domingo, 25 de julio de 2021

14:05

1. We are going to upload the levels 2-7 obtained in the *taxa-bar-plots2.qzv* file, but first we have to download the CSV of the respective level, then we go to the galaxy platform and add a new file by coping the table (CSV) on Galaxy (*Paste/ Fetch Data*). We are going to use the next Lefse tools:

- A) **Format Data for LEfse**
- B) **LDA Effect Size (LEfse)**
- C) **Plot LEfse Results**

- ★ Level 1 = Kingdom
- ★ Level 2 = Phylum (filos)
- ★ Level 3 = Class (clases)
- ★ Level 4 = Order (órdenes)
- ★ Level 5 = Family (familias)
- ★ Level 6 = Genus (género)
- ★ Level 7 = Species (especies)

2. **Format Data for LEfse:** *Columns - #2:organ (class) - #3:pond (subclass) - #1:index (subject) - yes*. These parameters are for all the levels.

Selects the structure of the file (classes, subclasses, subjects) and formats the tabular abundance data for the LDA Effect Size.

3. **LDA Effect Size (LEfse):** *Alpha Value for factorial Krustal-Wallis (0.05) - Alpha Value for the pairwise Wilcoxon (0.05) - Threshold on the logarithmic LDA score (1.0) - No- All-against-all (more strict)*. These parameters are for all the levels.

Performs the analysis using the data and provides input for the visualization in Plot LEfse Results.

4. **Plot LEfse Results:** *Text and label options (Default) - Graphical options (Default) - Output format (png) - dpi resolution (150)*. These parameters are for all the levels.

Graphically reports the discovered biomarkes with their effect sizes

LEfSe (Linear discriminant analysis Effect Size) determines the features (organisms, clades, operational taxonomic units, genes, or functions) most likely to explain differences between classes by coupling standard tests for statistical significance with additional tests encoding biological consistency and effect relevance.