Environmental Population Analysis

Fernando Freire

May 21, 2019

Contents

1 Environmental analysis **2**
1.1 Responses . . . 3
1.2 Pipeline . . . 5
1.2.1 Preprocessing and quality check . . . 5
1.2.2 Determination of ASV using Deblur . . . 13
1.2.3 Phylogenetic distances determination using FastTree . . . 15
1.2.4 Taxonomic assignment . . . 16
1.2.5 Diversity analysis . . . 18

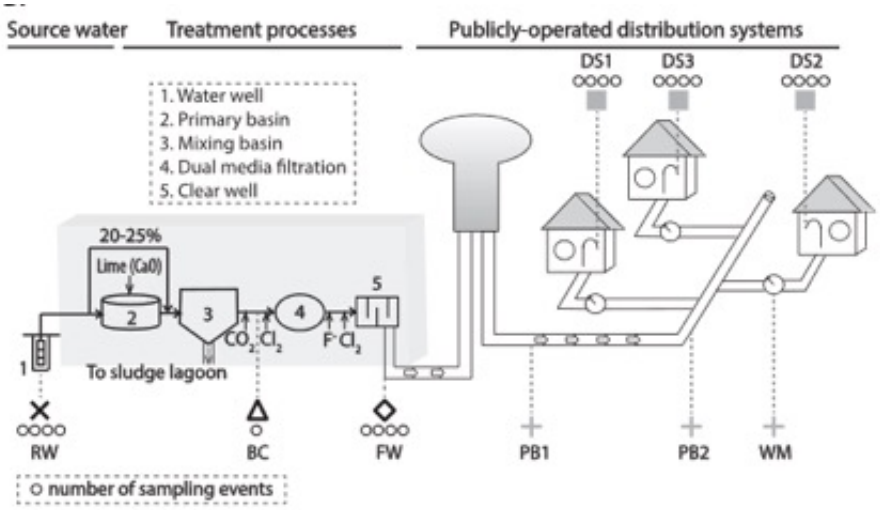

Figure 1: scheme

1 Environmental analysis

RW: Raw Water before treatment
BC: Before Chlorination
FW: Finished Water
DS1-DS3: tap water
PB1-PB2: biofilm in pipes
WM: biofilm in water meters

We are going to study the variation of microbial populations before entering the circuit (RW), at the end of the treatment (FW) and tap water samples in three houses (DS1, DS2 and DS3). We will also study samples taken in June and July.

I propose you to run the pipeline I showed you in Unit 2 whit this dataset and answer the following questions:

1) How many reads have we got for each sample?
2) Which is the trimming length you are using for the denoising step?
3) How many ASVs do you have before filtering deblur-table? And after filtering?
4) Which is the average frequency of sequences per sample before filtering deblur table?
5) Which is the sample with the lower number of sequences after filtering deblur table? And the one with the highest number?
6) Which is the most abundant phylum in each sample?
7) Has the study enough coverage to allow us to make any statistical inference on communities' diversity?
8) Studying the Unifrac Weighted PCoA plot. Is there any effect of water treatment or sampling moment on the bacteria communities?
9) If we compare Untreated samples (RW) vs Treated samples (FW, DS1, DS2, DS3) which are the Phyla or Classes explaining the differences among both groups? (Hint: use LEfSe)

Dataset Contents:

1) fastq folder: raw sample sequences
2) 85_otus.fasta and 85_otus_taxonomy.txt: taxonomy database
3) metadata.txt: sample metadata. I have included some columns useful for diversity analyses.
4) primers.txt: information on primers used for 16S PCR amplification
5) quiz.docx: this file
6) samplemanifest: manifest file with information of the ubication of fastq files and their corresponding tags.

1.1 Responses

1) How many reads have we got for each sample?

   From *grep* against fastq files and verified after import on *qiime*.

File	#reads
SRR3593621_1.fastq	52945
SRR3593621_2.fastq	52945
SRR3593622_1.fastq	62218
SRR3593622_2.fastq	62218
SRR3593623_1.fastq	92740
SRR3593623_2.fastq	92740
SRR3593625_1.fastq	70366
SRR3593625_2.fastq	70366
SRR3593627_1.fastq	100615
SRR3593627_2.fastq	100615
SRR3593628_1.fastq	78495
SRR3593628_2.fastq	78495
SRR3593631_1.fastq	97332
SRR3593631_2.fastq	97332
SRR3593632_1.fastq	84361
SRR3593632_2.fastq	84361
SRR3593664_1.fastq	101827
SRR3593664_2.fastq	101827
SRR3593665_1.fastq	84850
SRR3593665_2.fastq	84850

2) Which is the trimming length you are using for the denoising step?

   The qualiy of the reads is incremented substantially after the merge step by VSEARCH, consequence of a great overlap within forward and reverse sequences. In fact the filtering by means of 'qiime quality-filter q-score-joined' doesn't drop any read.

index	Unknown	Actinobacteria	Bacteroidetes	Caldiserica	Chlorobi	Chloroflexi	Cyanobacteria	Firmicutes	Nitrospirae	OD1	OP11	OP3	Planctomycetes	Proteobacteria	Spirochaetes	WS3	Month	Group
SRR3593622	0.0	257.0	0.0	0.0	0.0	0.0	2847.0	0.0	0.0	0.0	0.0	0.0	183.0	**29274.0**	0.0	0.0	June	DS1
SRR3593623	0.0	65.0	0.0	0.0	0.0	0.0	5817.0	0.0	0.0	0.0	0.0	0.0	114.0	**20791.0**	0.0	0.0	July	DS1
SRR3593627	0.0	94.0	756.0	0.0	0.0	0.0	13014.0	0.0	0.0	0.0	3.0	0.0	4235.0	**22101.0**	0.0	0.0	June	DS2
SRR3593628	3.0	27.0	96.0	0.0	0.0	0.0	5388.0	0.0	0.0	0.0	0.0	2.0	1081.0	**20020.0**	0.0	0.0	July	DS2
SRR3593631	0.0	0.0	2499.0	0.0	0.0	0.0	**32169.0**	0.0	0.0	0.0	0.0	0.0	662.0	6601.0	3.0	0.0	June	DS3
SRR3593632	2.0	0.0	51.0	0.0	0.0	0.0	**21105.0**	0.0	0.0	0.0	0.0	0.0	124.0	11733.0	0.0	0.0	July	DS3
SRR3593664	0.0	211.0	30.0	0.0	0.0	0.0	10267.0	0.0	0.0	0.0	0.0	0.0	160.0	**24818.0**	3.0	0.0	June	FW
SRR3593665	3.0	0.0	0.0	0.0	0.0	0.0	9009.0	0.0	0.0	0.0	0.0	0.0	143.0	**23019.0**	0.0	0.0	July	FW
SRR3593621	0.0	10.0	169.0	35.0	153.0	703.0	24.0	167.0	36.0	10.0	3365.0	337.0	0.0	868.0	**10652.0**	57.0	June	RW
SRR3593620	0.0	25.0	218.0	40.0	527.0	1376.0	28.0	808.0	289.0	20.0	5803.0	879.0	0.0	**4992.0**	45.0	246.0	July	RW

Figure 2: Taxonomy by philum

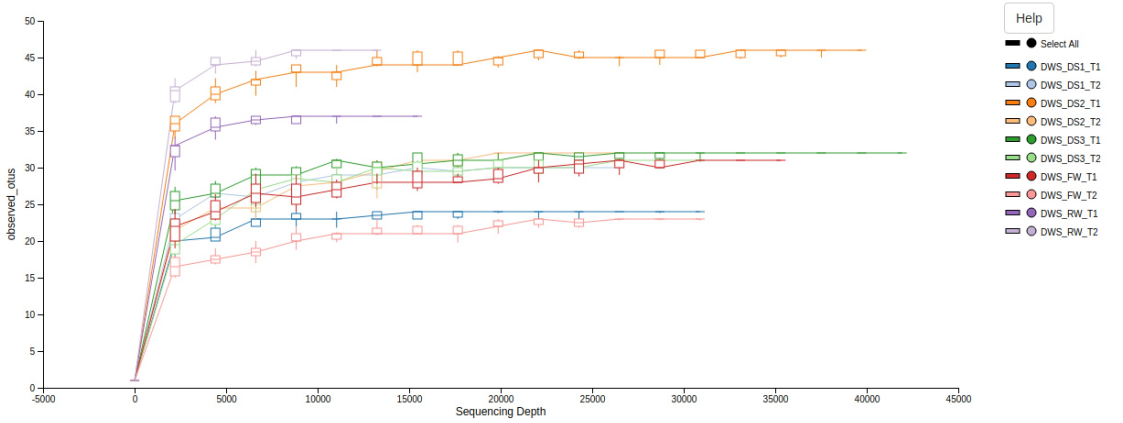

Figure 3: Rarefaction curves

I choose 372 that retains $Q >= 38$.

3) How many ASVs do you have before filtering deblur-table? And after filtering?

   Before filtering I've 765 features after 104.

4) Which is the average frequency of sequences per sample before filtering deblur table? >30,070

5) Which is the sample with the lower number of sequences after filtering deblur table? And the one with the highest number?

The sample *SRR*3593622 has the lower number of sequences : 15,296 The sample *SRR*3593631 has the higher number of sequences : 41,934

6) Which is the most abundant phylum in each sample?

The marked in the table of figure 2.

7) Has the study enough coverage to allow us to make any statistical inference on communities diversity? Yes, the rarefaction curves reach saturation asymptotic pattern, no more reads would give us more observed OTUS (figure 3).

8) Studying the Unifrac Weighted PCoA plot. Is there any effect of water treatment or sampling moment on the bacteria communities?

The water treatment groups cluster together nearly over axis2-axis3 plane, and the not treated water groups far along the axis1. As axis1 explains 43% of differences, we conclude that treated groups are most in common between than between the not treated groups (figure 4).

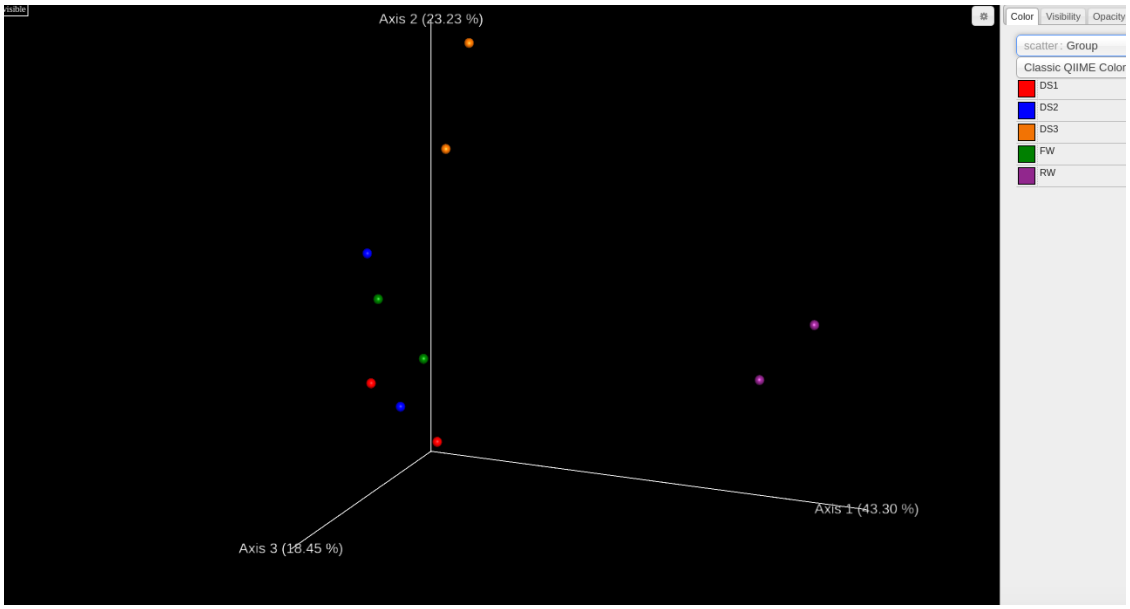

Figure 4: PCoA groups

Also the graph evidences the dissimilarity between DS3 groups and the other treated groups, above on the axis2. The biofilm presence in the distribution, surely is the explaining factor.

The difference in months are less relevant as seen in figure 5. In the case of treated groups month juny groups cluster nearer to the axis2.

9) If we compare Untreated samples (RW) vs Treated samples (FW, DS1, DS2, DS3) which are the Phyla or Classes explaining the differences among both groups? (Hint: use LEfSe)

The classes are drawn in the figure 6 LDA representation (TT treated and TN untrated). The most explanatory respectively are cyanobacteria and sphirocaetes + OP11, coherently with the table of question 6.

Also the graph evidences the higher diversity in the not treated water, as expected.

1.2 Pipeline

1.2.1 Preprocessing and quality check

Script 1.2.1 (python)

```python
import warnings
warnings.filterwarnings('ignore')
import pandas as pd
import matplotlib.pyplot as plt

FILE_ID = "SRR"
FASTQ_STR = "@SRR"
```

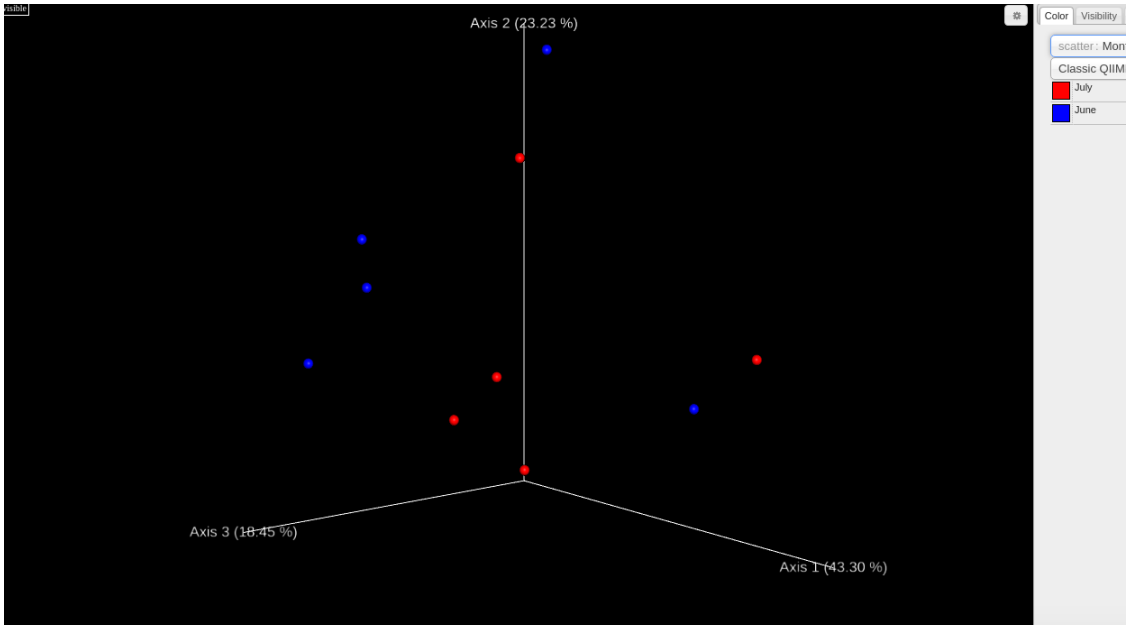

Figure 5: PCoA months

Script 1.2.2 (bash)

```bash
%%bash -s "$FILE_ID" "$FASTQ_STR"
ssh microbioinf@192.168.56.101 env FILE_ID=$1 FASTQ_STR=$2 /bin/bash <<"EOT"
export PATH=$PATH:/home/microbioinf/miniconda3/bin
cd Documentos/Tema_2/fastq
head -4 ${FILE_ID}*fastq
grep -c $FASTQ_STR ${FILE_ID}*fastq
EOT
```

Output

```
==> SRR3593621_1.fastq <==
@SRR3593621.1 1 length=300
TACGTAGGGTGCAAGCGTTATCCGGATTCACTGTTCTTCCCTATTCGTTTAGTTTTTTTTTTCCTTCTCACTTCACAGCCCTTTGCTTTACCT
↪ CGTCCTTCTTTTCTTCTTTACTATACTCGTTTTTTATATTTGTACGTGGTTCTCCTTTTTGTTCTGTGCCTTGCGTTGTGTTCTTGTTG
↪ TACCCCAATTGCCATTGCTCCTTTCTTCTTCATTCCTGTCACTCTTCCACGCAAGCTATCGTACTCCATCAGTTTAGTCCCCTCCTTTT
↪ TTCTAGCCCTCAACTCTTCCCTGCTAGTT
+SRR3593621.1 1 length=300
<6BCCGGGGGGGGGGGGGG7CEE6+,8@@,;C,,,<,<,,,,,,;6,;,6,,,;-,:+8+86,,,<9,:,6CE,,,,,,66,,,,,5:,,669
↪ 6,,8,89,,,9,<,,,,9,,,:,<,,95+,+++9+,9,,,,,,,,+,,+++:>;,74,4,+,,5,,8,,,,8,,+66+++,66,7,,,,
↪ ,,:,:5++,,+2,,,,+3+74<2=2@,,+22,5,5*4+,4*/5++++3*.)*)(((//++(-(((*+*))+)))3.)))++,/;*/(/)/
↪ /587))6((,(*),-.).)-44)43)).-

==> SRR3593621_2.fastq <==
@SRR3593621.1 1 length=300
```

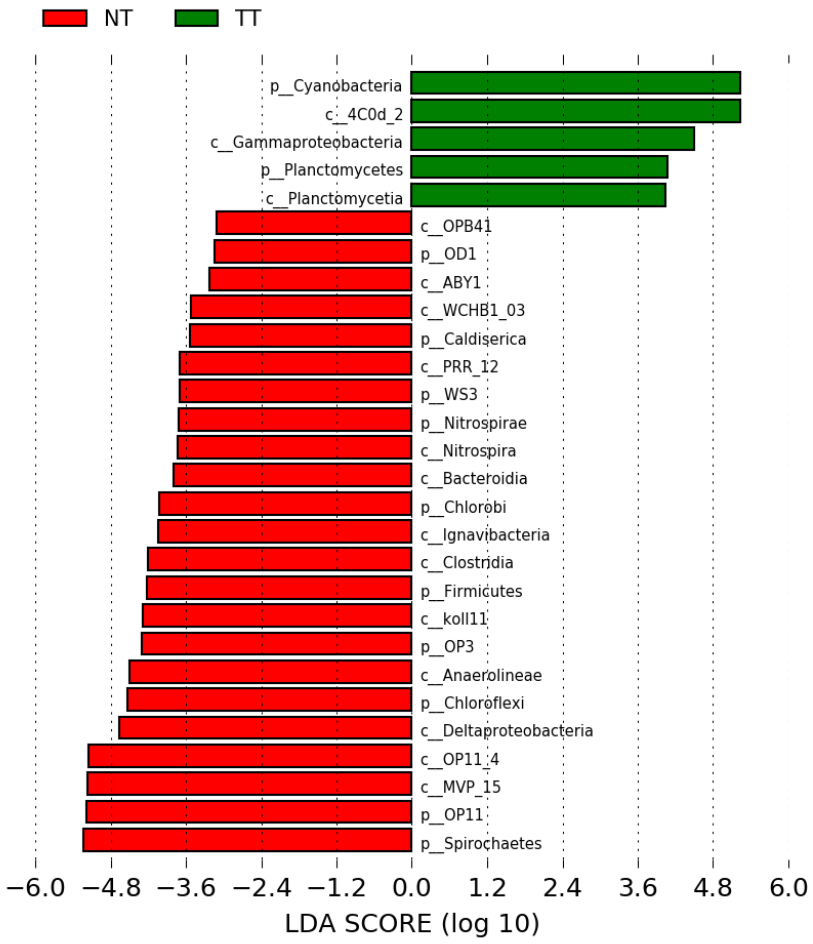

Figure 6: LDA treated/not treated

```
TTTCCTCTTTCTCTCTTCCTCCCCACCCTTCCTCCTTCCCTCTTTCTCTTCCTTCTCTCCTCACATTCTCCTCCCTCCCCTGCTTCCCTTTAT
↪  TTCTCCCTCTACTCCTCTCTCTTCTCTGCTTCTCTCCTCTACCTTTCGTTTCTCAGTGTCATCTATTTCCCTGCCCGTTGCCTTCTCCT
↪  TTCTTCTTCCTCCTCATCTCCACTCATTTCACTTCTCCTCCATGCTTTCCCCTTTCCTCTATCACCCTCCACTCTCTTCCTCTCCTCCC
↪  CCACGCCCCCTTTACGCCCCGTCCTTTCC
+SRR3593621.1 1 length=300
88@--;<C@,C,-6;,6,<C@E@C,+,8,,,;;,B@C,,65;,4>,,@A?,5,6;,;,C6,,,,,,6,,,,6,,8;,,8,,,4,:,5,,94,,
↪  ,,99,,,,,9,9,4,,,,,,,5+,3>9+934459526389+,923<9*0*09;@**3*26*++6*6++20+*))++**4)10008*03>
↪  *0*+0*3:;+:9):)**320**3*))*3;;*205:;=D*80)1*)****001010:8;87**3*;018>5))>@E):7A*19/))*)-/
↪  -)-5(-((((((()))(((((.(((,//./)

==> SRR3593622_1.fastq <==
@SRR3593622.1 1 length=300
TACGAAGGGGGCTAGCGTTGTTCGGATTTACTGGGCGTAAAGCGCACGTAGGCGGTTTTTTAAGTCAGGGGTGAAATCCCAAGGCTCAACCTT
↪  GGAACTGCCTTTGATACTGGAAGTCTCGAGTCCGGGAGAGGTGAGTGGAATTGCTAGTGTAGAGGTGAAATTCGTAGATATTCGCAAGA
↪  ACACCAGTGGCGAAGGCGGCTCACTGGCCCGGTACTGACGCTGAGGTGCGAAAGCGTGGGGAGCAAACAGGATTAGATACCCTGGTAGT
↪  CCACGCCGTAAACTATGGATGCTAGCCGT
+SRR3593622.1 1 length=300
CCCC@FFGGGGGGGGGGGGGGGGGGGGFFCFFFGECFFGGGFG:+8@EFFG@FF+>BDDEE<EFC,C,,@FEGFGEE?=,,:?DFAFFGEEE
↪  9<,,4A?=<EFD<,A<?F9,,,::AF;?<7DF?>F+8@==FF7,,9=+F,,:,@,,8@FDFE9=:D*>D,@3=CDGGDFFGFAD>>7CC
↪  ,>,>FE,C<7<=E=@EDB75>5+2A8+*8*/*211:48CE1:*775C)4>?>>B73;@DB66>>*;D5(<<07@?F66<CF86*546<A
↪  AB6<BF?0>?::<<2<62).,3994(,,4

==> SRR3593622_2.fastq <==
@SRR3593622.1 1 length=300
TTTAACCTTGCGACCGTACTCCCCAGGCGGGATGCTTAATGCGTTAGCTGCGCCACTGAACAGCTAGCTGCCCAACGGCTAGCATCCATAGTT
↪  TACGGCGTTGACTACCACGCTATCTAATCCTCTTTTCTCCCCACGCTTTCGCACCTCAGCGTCAGTACCTTGCCAGTGTGCCGCCTTCG
↪  CCACTGTTGTTCTCCCGACTCTCTACGCCTTTCACCTCTCCACTCCCAGTTCCACTCTCCCCTCCCGTACTCGAGACTTCCATTTTCAA
↪  AGGCAGTTCCAAGTTTTCGCCTTGGCATT
+SRR3593622.1 1 length=300
AC<8-;FFFFGC+7@:FFGGEGEGGF@C@+++8,BFG@,,:@>CE@6EE,@76,+CC6,,:,,4,,,:6,99C,,8=+84+,,,:94,9,,9C
↪  @,54+>68+++B;,4A,,++8+<>;,+53?;,5>@,24;+;@8>+3=0=0+0*@)1A)*389**)53+4*+/)0+***,*14)-55-7(
↪  -624.))/.;17/)),(,(((/6)676((/,=2)63,4<(,-343(.))5<6))66:).2(-.:(33,,(-.-1(((..4<A)))))))
↪  ))(((((,.4))),).))),-3,()((,).

==> SRR3593623_1.fastq <==
@SRR3593623.1 1 length=300
TACGGAGGGAGCTAGCGTTGTTCGGAATTACTGGGCGTAAAGCGCACGTAGGCGGCTTTGTCAGTTAGAGGTGAAAGCCTTGAGCTCACCTCC
↪  AGAATTGCCTTTAAGACTGCATCTCTTGAATCCAGGAGAGGTGAGTGGAATTCCTAGTGTAGAGGTGAAATTCGTAGATATTCGGATGA
↪  ACACCAGTGGCGAAGGCGGCTCACTGGACTGTTATTGACGCTGCGGTGCGAAAGCTTGGGGAGCAACCAGGATTAGATACCCTGGTAGT
↪  CCACGCCGTAACCGATGATAACTAGCTGT
+SRR3593623.1 1 length=300
C@CCCGGGGGGGGGGGGGGDFFG7@;,CE<;FC<8C@C@,CFC77+8@C:E@6@:+8=C,C,,<CC,,,,6C,,CE,,:C,,,:@E=C,CC==
↪  ,9,,<<,<?EFA,,,<:F,,,:A,9E@,,5A@,,5,++,8:A+,,98+A,CB>=,,9CA@==,,8@+,,,@=D+@A?6@,ECC,,++53
↪  ,:,@F*4@3,;55*>*9>):::,4B8,*22/+314<11<*(6*)/):)421*72:*15)3(413*0*)(0(,8=>?2.0<<<?*/...4
↪  9@*8(4,(42:549(3,).)).6).(46:

==> SRR3593623_2.fastq <==
@SRR3593623.1 1 length=300
```

```
TTTACTCTTGCGTCCGTACTCCCCAGGCGGTTCTCTTCTTGCGTTAGCTTCGCCCCCACACTTCTTTCCACCCTCTCCCCTCTTTTTCATCCT
↪ TTTCCCCCTTTACTTCCCCCTTACCTCATCCTCTTTTCTCCCCCCTCTTTCGCTCCTCCTCTTCATTTCCTCTCCCGTTATCCTCCTTC
↪ GCCTCTCTTCTTCTTCCGAATATCTACCTATTTCACCTCTACCCCCGTAATTCCACTCTCCTCTCCCGCCTTCAAGCTCTGCACTCTTA
↪ CATGCACTTCTGCATTTCAGCCCCATCCT
+SRR3593623.1 1 length=300
8@6--6CFC;E++7@+C6BF@EEEB,,98+++,,;EE,,;,66>B,,EF<@+,6,8+++,,,5C<,,,,,69,,,,:,,,4,,99,99,4,,9
↪ ?@,9,,,+4,,,4,,9A,,++,*,)9+*4<,4,9?;++0+368+++03=<+*0+380)*+)+42*+4+20+*5:7+*3,))10**359*
↪ 15>*10)****1*2//*)))))*2/*/****10*/00075*9*/((,0((//)6)743)),.6631((((-(((((((-)))(.(-5-
↪ ))))),)-4))-))))..)))-(-(((((,

==> SRR3593625_1.fastq <==
@SRR3593625.1 1 length=300
TACGTAGGGTCCAAGCGTTAATCGGAATTACTGGGCGTACAGCGTTCGCAGGCGGTTTTTCAAGTCCGATGTGAAATCCCCTTGCTTAACTTT
↪ GGTATTGCTTTTGTTACTGCTCGTCTTGAGTTTGTCAGAGGTGGCTAGAATTCCTCGTGTAGCAGTGAAATGCGTAGTGATGTGGTGGA
↪ ATACCGTTGGCGAAGGCAGCCCCCTGGGATAACACTGACGCTCATGCACGAAAGCTTGGGGAGCAACCAGGTTTAGTTACCCTGGTAGT
↪ CCTCGCCCTAAACGATGTCAACTAGTTGT
+SRR3593625.1 1 length=300
BBCCCGGGGGFGFGGGGGG>,<FDGC7DAC6ECEFC@F@+@FCC@,88+88@+86@@,9,,,69,:66+88E,,69<C666,,:@F<:,CCC,
↪ ,,,,9<55,9A,,:,55C,,,9++6A+,,,C,:,B,4,,:+,++,8,,C,C?4@,7@F:@=,7,7C8,5,@,@+6=C+,,7,7,6,+3+
↪ ,3,:>:+@*,9:***13,3*1*1;C***)0,,4,,868>*(67)C+1*7(,*):3*2):0((.4*2*0(04)5756)4+<<06**.46.
↪ 6=*639((4.,+5((2(.))).62)64)6

==> SRR3593625_2.fastq <==
@SRR3593625.1 1 length=300
TTTACTCTTGCGACCGTACTCCCCAGGCGGTCTACTTCTCGCGTTAGCTTCGTTACTACCCTCTTTCCTCCCCCACACCTACTTTTCATCCTT
↪ TTTGGCGTTGACTACCCCTCTCTCTATTCCTCTTTTCTCCCCACTCTTTCGTCCATGACCGTCACTTTTATCCCACCTCGCTCCCTTCT
↪ CCATCTGTATTCCTCCTCATCTCTACGCCTTTCTCTGCTACACCTCGTATTCTACCCCCCTCTGCCCCCCTCTACCCCTTCTGTCTCCA
↪ ATTCAACTCCCTAGTTATCCTCCTCGCTT
+SRR3593625.1 1 length=300
8B<--6CDDCE@+7@+B6BF@E@C,,,68+86,,CFFF,6,84>C,4EF,C496,@6,,,,,,,6,,,,66?:++7,,89,,59,,5,4,,59
↪ C,,++,48+++:9,,9,+++9,543,+2,47,5>?,33;+9@,=+5=+?++3*1)2+**6):9+)6+62*+303++**,)0*)1))1)*
↪ :6*1:**3*2;****-)76/8AE76;)/(0(4**1**0088*5(-(()((./.)173(.3(63)))(,((,-,))6((,(()).))--)
↪ (-)))))).)),()(.)))-.).,((((,(

==> SRR3593627_1.fastq <==
@SRR3593627.1 1 length=300
TACGTAGGATGCAAGCGTTGTCCGGATTTATTGGGCGTAAAGAGTTCGTAGGTGGTTTGTTAAGTTTGGTGTTAAAGATTGGGGCTCAACCCT
↪ GATACTGCTCTTAATACTGTCAGACTCGAGTGTGGTAGAGGCTAGTGGAATTCCCAGTGTAGCGGTGAATTGCGTAGTTATTGGGATGA
↪ TCACCGGTGGCGTAGGCGACTTCCTGGGCCATTACTGACACTGAGGAACGAAAGCCAGGGGAGCGAATGGGATTAGATACCCCTGTAGT
↪ CCTGGCCGTAAACGTTGTCTACTAGTCGT
+SRR3593627.1 1 length=300
CCCCCGGGDGGGFGFFGGGDCCFEG@@FE<CFECEC@F+,6C,;<CC@B,CCCCCBCC,:C9,EE9C<,:CFE,CC,,:9,,,@CG<CCFC==
↪ ,,,,6<,6,,<,,,<<5C,,5,,,9A,4,+C+<57A;,,,=+4,,9,,C,CA>>5,,8,C@=;>:C775,@==7@6C+@,@=,,3,,6,
↪ ,:,:>8+@<+;@3**5***4;*,4=;,*/4*4,,,9+14<0>**2**055@752(0);.((2(3=*/).92(4??2.62<<0)*1.).6
↪ 5=@).3<(4(-/66(41,).).6).))-4

==> SRR3593627_2.fastq <==
@SRR3593627.1 1 length=300
```

```
TTCAACCTTGCGCCCGTACTACTCAGGCGGGATACTTATCGCGTTAGCTGCGTCACATCACTGTTCGATACCCCATACGCCTAGTATCCATCG
↪  TTTACCGCCAGCACTACTGCGGTATCTAATCCCTTTCGCTCCCCTCGCTTTCGTTCCTCCGTGTCAGTTATGCCCCAGCTCCTCGCCTA
↪  CGCCACCTTTGTTCTTCCCCATATCTACGCATTTCACCGCTACACTGGGATTTCCACTAGCCTCTACCCCCCTCGATTCTGCCCGTTTC
↪  CAGTTCAGTTTCAGGTTTGACCCCCAATC
+SRR3593627.1 1 length=300
@C@8-;CFFFFC+6@:F@FGAF9E?66C@+++8,BFGC@698BEE7AEE,C+,6,8,9C,,,,,95,,6,86C,,4,:,,8?,,:,594,9,+
↪  :A<+A,++6+++,94,A,,,++3+3>3+,47=5,:=+0::DBDC8*+*=9900418ACB+0:*40++51*+**3;?**10)*00)-7>(
↪  7256(-50)+*109/96;3((6(1?557(,(247C37;22(49).)))(((..:<)44().44:72:)(((-,(((((.64-5)((-(.)
↪  6)-.)6))..--))).,-(-).3,,((.(

==> SRR3593628_1.fastq <==
@SRR3593628.1 1 length=300
TACGTAGGATGCAAGCGTTGTCCGGATTTATTGGGCGTAAAGCGTTCGTAGGTGGTTTTTTAAGTTTGGTGTTAAATCTTTTGGCTCAACCCT
↪  GAAACTGCTCTTAATACTGTCTGACTCGAGTGTGGTAGAGGCTATTGGATTTCCCAGTGTAGCGGTGAAATGCGTAGTTATTGGGAAGA
↪  TCACCGGTGGCGTAGGCGACTTCCTGGGCCTTTACTGACACTGAGGAACGAAAGCCTGGGGAGCGAATCGGATTAGATACCCCTGTAGT
↪  CCTGGCCTTACACGTTGGCTACTAGTCGT
+SRR3593628.1 1 length=300
-CCCCGGGFGGGFGGGGGG>BCFCC,6CE;<FEFC@6C6,CC,888668,:CC,CBCC,:@,,CC9C,,:CFF66,,,<9,,,8@F5:,C9,6
↪  ,,,,,9,5,,<,,,955C,,9,,,4:44,+:+<,+9,,,98+4,,9,,9,:A4=7,7A:C@,;++B++,,@9=36+@+@,@=,,,,,+@
↪  ,,,:>++@339@3*5*9**11*,3=,,**/*4,519+4<,0>*9***)/6(1)438).)2(1(:**9.(0(-8@28)6+3<+.**.6).
↪  6=/..36)4),<.8(4(((341:)5))(4

==> SRR3593628_2.fastq <==
@SRR3593628.1 1 length=300
TTCAACCTTGCGCCCGTACTACTCAGGCGGGATACTTATCGCGTTAGCTTCGCCACTTCACTTGTCTATACCCGATACCCCTAGTATCCATCG
↪  TTTACCGCCAGGACTACTGCGGTATCTACTCCCTTTCGCTCCCCTGTCTTTCGTTCCTCAGTGTCAGTTATCGCCCAGCTACTCGCCTT
↪  CCCCACCTTTGTTCTTCCCTATATCTACGCATTTCACCTCTACCCTGGGAATTCCACTTCCCTCTTCCACACTCGAGTCTGCCAGTTTC
↪  CCGTGCACTTTCAGGTTTCACCCCCACTC
+SRR3593628.1 1 length=300
AC---;CFFFFC+6@@FCEG,C,B9;66@:++8,BFFDA;9@CEG?6ED,C+,646,6:,,,,,:5,,9,<<6++68>,9:A,,:,5@=,9,+
↪  :A<+9,448++++99,85,,++34:><+,47@,,:@3*::FEFEC*+5=CD;5681AEA+0<+5;++53*;*)8;C**4)**00)/7<)
↪  :)09)(/()4*17CC?5>=*.0*9@C6701(.,9C))7.2)76)/)()(((//7?)))))-44?2)65>4(4<?:((.4).53)).)-)
↪  .)-((,().,-.)))).)).).((-((,,

==> SRR3593631_1.fastq <==
@SRR3593631.1 1 length=300
TACGGAGGGAGCTAGCGTTGTTCGGAATTACTGGGCGTAAAGCGCCCGTAGGCGGCTTTTTAAGTCTGGTGTGACATCCCTGGGCTCAACCCC
↪  GGTACTGCCCTTGATACTGTCTATCTTGATTCTTGTAGTGGCGATTGGAATTCCTAGTGTAGAGGTGAAATTCGTAGTTATTCGGAAGA
↪  TCACCAGTGGCGAAGGCGACTCCCTGGACCTGTCTTGCCGCTGAGGTTCGACAGCTTGGGGAGCCACCCGGTTTAGTTACCCTGTTAGT
↪  CCTCGCCGTACACGTTGTTACCTATCTTT
+SRR3593631.1 1 length=300
B@CCC@FGD7CFEFGFGGG>C@F@B@,CCC6EC@C@8C+,,CC@6+77@68@,+8+8=CCE,,:CC,,,,9C,9,<95,4,,,4@E=,:F44+
↪  ++++4<,5,,:=,,,,5C,,,,,,8C,,,,<7@@,,,7,,,++++83,3,:>=@,,3C,@3,,,5@+,8,7>;35;@+7,@<=,,***,
↪  ,3,8>,,67,34***/8*)*4*+2;:**.***5+49+/0*(4))/*2)4(.*.,1)1)04(-(3*+0.(04(45.:).+3<+1**.5).
↪  <?*.,22(-4(*64(4((.))-.).).)6

==> SRR3593631_2.fastq <==
@SRR3593631.1 1 length=300
```

```
TTTACTCTTGCGACCGTACTCCCCCGGCGGTTCACTTATTGCGTTAGCTTCGCCACCCCCCTTCTTTCCACCCCGCCCCCTTTTTATCCTCTT
↪ TTTCCCCCTTTCCTTCCCCCGTTTCTTATCCTCTTTTCTCCCCACGCTTTCGCTCCCTCTCTTCCTTTCTTTTCCCGCTATTCTCCTTC
↪ GCCCCTCTTTTTCTTCCCACTTTCTCCCCCTTTCCCCTCTTCACTCGGACTTCCCCTCCCCCCCCCCTCATTCCCGCTTCCCCCTTTCC
↪ ACCTCCCTTCCCGCTTTTCCCCCCTCTCT
+SRR3593631.1 1 length=300
@B@--6CEDCE@+6@+B,BF@F@E+++77++6,,:FF,,6966>C7,EE,9+,6,@6=+++449C,5,,,59=+++7+++4,,95,59,7,,7
↪ 8?,@,,,+4,,,37,9=,,++++,),++3573,7::+2049>C,0*+033+*0*11*)**0*3+*+4+4/2*1:2**)*((+1*05.>*
↪ ((5)3,(()+/.(...)()(/))).)/(((,0..))4((,).).).()((,.).(.(,(-,(-,((((((.))))-,-,,,(((-.)))
↪ ))((.(,(..((((,-()))32(((((,

==> SRR3593632_1.fastq <==
@SRR3593632.1 1 length=300
TACGTAGGATGCAAGCGTTGTCCGGATTTATTGGGCGTAAAGAGTTCGTAGGTGGTTTGTTAAGTTTGGTGTTAAAGATTGGGGCTCAACCCT
↪ GATACTGCACTGAATACTGGCAGACTCGAGTGTGGTAGAGGCTAGTGGAATTCCCAGTGTAGCGGTGAAATGCGTAGATATTGGGAAGA
↪ ACACCGGTGGCGTAGGCGACTAGCTGGGCCATAACTGACACTGAGGAACGAAAGCCAGGGGAGCGAATGGGATTAGATACCCCAGTAGT
↪ CCTGGCCGTAAACGATGGATACTAGGCGT
+SRR3593632.1 1 length=300
-ACCCFGGFFFGGGFFEFGEFEGGGFEGEFCFGGGGGGGGFFCGEE@FCCFGEGGFED,CECCFG<CFCFGGGGGGF,6C6,CCEG<CFG<8C
↪ <?,,6C@9,,C,:,?C,FE,,4?,9A=D<FF+BFFFFGFGG:9=,CFFG9FE=>5,EFGEEEDFGGGFF,@,=E@CE+@=DA,BEE,AF
↪ ,@@FGG6EEEDE:@FEG8C9;*,0BFF*04*=,>8C@BF60509FC*):??EFFF)08>?66(./:>5:FF??FFF2AGDC8585>44:
↪ <FFB:::2906CA:6<8?(.6:AA2029<

==> SRR3593632_2.fastq <==
@SRR3593632.1 1 length=300
TTCAACCTTGCGGCCGTACTTCTCAGGCGGTATACTTATTGCGTTAGCTGCGCCACTCCACCTTTCGATCCCCCACACCCCTTGTTCCCATCT
↪ TTTACCGCCCGGACTACTGCGGTATCTAACCCCTTTCGCTCCCCTCGCTTTCGTTCCTCCCTGTCTTTTTTCCCCCAGCTCCCCCCTTT
↪ CGCCTCCTGTTTTCTTCCCTCTATCTCCGTTTTTCCCCTCTACCCGGTGATTTCCTCTCCCCTCTCCCACACTCGATTCTGCCATTTTA
↪ CAGTCCTTTTTCGTTGTTCACCCCCCATC
+SRR3593632.1 1 length=300
AC6A-,=CDFFF+7F:CFGG,E,CFC<68@++6,CFF@6,;,>BE7@EF,67+,+8,,:,,,,,6,,,,,,96+++,6,4+,,,5,,4,,44,
↪ <A,9:,+++++++8+49,,,+83+8>2+,+38++8@=*1*;==A8***6*3304)1;C=+)5*3)++5+*3)*3*?45,))*)))-(-)
↪ 1((/((,()2)1()).)/()(/)16.5)(()0(((()/8(-().)((()(().--6).-((,((-3(,,41((81(.)))))).))--)
↪ ..4.).))).).-.((-(--.-4(((4,.

==> SRR3593664_1.fastq <==
@SRR3593664.1 1 length=300
TACGTAGGATGCAAGCGTTGTCCGGATTTATTGGGCGTAAAGCGTTCGTAGGTGGTTTTTTAAGTTTGGTGTTAAATCTTCCGGCTCCACCCT
↪ GATACTGCTCTTAATACTGTCAGACTCGAGTGTGGTAGAGGCTATTGGATTTCCCAGTGTAGCGGTGAATTGCGTAGTTTTTTGGATTA
↪ TCCCCGTTGGCGTAGGCGACTTTCTGGTCCTTTACTGACACTGAGGAACGAAAGCCTGGGTAGCGAATCGGATTAGTTCCCCCTTTAGT
↪ CCTGGCCGTACACGATGGTTCCTAGTCGT
+SRR3593664.1 1 length=300
CCCCCGGG?FFGCGGGGGG7C@FCF86CC,<ECCEC@F@,,C,88686@,:::,98C@,9@,,:<6@,,::EF<@9,,99,,,8@F4,6C94=
↪ ,,,,5<,9,,<,,,9<9C,,5,,,9A4+,+C+<,+995,,,+,,,9,,5,:A447,,8,8@,4++@++5,@,@3@+@+@,@=+++,,4>
↪ ,7,:@**6*,3**1**3**14*,4=,,,10,4,419+4>908*20**)/61*573)).2((/,3/*)0(03(4;<<).+3<:)**.-).
↪ 6@6)(36(-3(*.4(3-()-),.:))-,4

==> SRR3593664_2.fastq <==
@SRR3593664.1 1 length=300
```

```
TTCAACCTTGCGCCCGTACTACTCAGGCGGTATACTTATCGCGTTAGCTTCGCCACTTCCCCCTTCTATACCCCATACCCCTTGTATCCTTCT
↪ TTTACCGCCTGCACTACTGCGGTATCTATTCCCTTTCGCTCCCCTCGCTTTCGTTCCTCAGTGTCTTTTCTTCCCCCGCTCCTCTCCTT
↪ CGCCTCCTGTTTTATTCCCAATATCTTCGCTTTTCCCCGCTACACTTGGTTTTCCTCTCGCCTCTACCACCCTCGATTCTCCCATTTTC
↪ CAGTGCAGTTTCAGTTTTTCTCCCCCCTT
+SRR3593664.1 1 length=300
@B6--6CFECEF+,@+@+BF9F,EF;6C8+++:5EFG<@,:6@CE76ED,C+,6,?,,:,,,++64,,6,669,,6,9,,:?,,:,,44,9,,
↪ <:9,A,,+4++,,94,9,,,++3+3>++,444,,9=338:?C=C8*+*60=0*151@DD+)5+60++21*2**5=D**+))*03*308*
↪ :)85))/))5*10/1/*/7*(7)98**1/((055@*.5(4..2)6.)+))(,/)7))1((,4,4()6)).(--((((.)))..()-).)
↪ ))...6)))--))))-))(,)66,((-,(

==> SRR3593665_1.fastq <==
@SRR3593665.1 1 length=300
TACGGAGGGAGCTAGCGTTGTTCGGAATTACTGGGCGTAAAGCGCACGTAGGCGGCTTTGTAAGTTAGAGGTGAAAGCCTGGAGCTCAACTCC
↪ AGAATTGCCTTTAAGACTGCATCGCTTGAATCCAGGAGAGGTGAGTGGAATTCCGAGTGTAGAGGTGAAATTCGTAGATATTCGGAAGA
↪ ACACCAGTGGCGAAGGCGGCTCACTGGACTGGTATTGACGCTGAGGTGCGAAAGCGTGGGGAGCAAACAGGATTAGATACCCTGGTAGT
↪ CCACGCCGTAAACGATGATAACTAGCTGT
+SRR3593665.1 1 length=300
CCCCCGGGGGGGGGGGGGGGGGGGGDFFFEGGGGFGGFFFGEF@@@@FFGGCCG7CCC,C<<FFF,C9,CFFGEF9CAE,,=FFGCFFGDFD
↪ 8=,,<?=AAEFE9,,C?F@@9EE+@FD@8=FD?EF,+?<FFF798CDFGFFGFF,+8@@EFC9EGFC:8,ECF@DFG>DFGGEDC85;D
↪ 7@>FC>5DFGGGGDDGF5CCED+;CFC*2=55>;=CFCGE1?DBF:D3?B:5<9?)8>>B0<>;CC>;6F?<BF7>95+3<8F6D84<<
↪ >F4>?9:69??:6?9:365)6<A<7(64<

==> SRR3593665_2.fastq <==
@SRR3593665.1 1 length=300
TTTAATCTTGCGACCGTACTCCCCAGGCGGATAACTTAATGCGTTAGCTTCGCCACCAAAGTTCTGTGAACCCTGACAGCTAGTTATCATCGT
↪ TTTCGGCGTGGACTACCAGGGTATCTTATCCTGTTTGCTCCCCACGCTTTCGCACCCCAGCGTCAATACCAGTCCAGTTATCCGCCTTC
↪ GCCACTCTTGTTCTTCCGAATATCTACGAATTTCACCTCTACACTCGGAATTCCACCCACCTCTCCTGGATTCATGCGATCCAGTCTTA
↪ AAGGCAATTCTGGATTTTAGCTCCATGCT
+SRR3593665.1 1 length=300
CCCC-CFFFFFGC@CCFFGGDFEEEFDCCC7@,5FFG<,<E<EEGFEGF,C@78,?8<,,,9AEE,<9,,:C?,4,:,,95,,CA,:@,@;,A
↪ ED,9++4+8+++4,,AB,,+?439:C++7===,@EE?79??@EF8*60@F+*:+15*)*+05::)+;+3)+)3:?7*?,;*4;5):=<3
↪ 49A*:.)*/*18**976(---/*/*5@((((27796@3273>)67@3)(233)9)7(/(462,),46.(())<<)))((-(-))..642
↪ <))),((.)...)).)4))(.).4().()
SRR3593621_1.fastq:52945
SRR3593621_2.fastq:52945
SRR3593622_1.fastq:62218
SRR3593622_2.fastq:62218
SRR3593623_1.fastq:92740
SRR3593623_2.fastq:92740
SRR3593625_1.fastq:70366
SRR3593625_2.fastq:70366
SRR3593627_1.fastq:100615
SRR3593627_2.fastq:100615
SRR3593628_1.fastq:78495
SRR3593628_2.fastq:78495
SRR3593631_1.fastq:97332
SRR3593631_2.fastq:97332
SRR3593632_1.fastq:84361
SRR3593632_2.fastq:84361
SRR3593664_1.fastq:101827
SRR3593664_2.fastq:101827
```

```
SRR3593665_1.fastq:84850
SRR3593665_2.fastq:84850
```

Script 1.2.3 (bash)

```bash
%%bash
ssh microbioinf@192.168.56.101 /bin/bash <<"EOT"
export PATH=$PATH:/home/microbioinf/miniconda3/bin
source activate qiime2-2018.11
cd Documentos/Tema_2
qiime tools import --type 'SampleData[PairedEndSequencesWithQuality]' \
                   --input-path samplemanifest \
                   --output-path paired-end-demux.qza \
                   --input-format PairedEndFastqManifestPhred33
EOT
```

Output

```
Imported samplemanifest as PairedEndFastqManifestPhred33 to paired-end-demux.qza
```

Script 1.2.4 (bash)

```bash
%%bash
ssh microbioinf@192.168.56.101 /bin/bash <<"EOT"
export PATH=$PATH:/home/microbioinf/miniconda3/bin
source activate qiime2-2018.11
cd Documentos/Tema_2
qiime demux summarize --i-data paired-end-demux.qza --o-visualization paired-end-demux.qzv
EOT
```

Output

```
Saved Visualization to: paired-end-demux.qzv
```

1.2.2 Determination of ASV using Deblur

Sample pre-processing

Script 1.2.5 (bash)

```bash
%%bash
ssh microbioinf@192.168.56.101 /bin/bash <<"EOT"
export PATH=$PATH:/home/microbioinf/miniconda3/bin
source activate qiime2-2018.11
cd Documentos/Tema_2

# Deblur does not currently support unpaired paired-end readings,
#so we have to use the VSEARCH algorithm to merge the readings:
qiime vsearch join-pairs --i-demultiplexed-seqs paired-end-demux.qza \
```

```
                          --o-joined-sequences joined-reads.qza
qiime demux summarize --i-data joined-reads.qza --o-visualization joined-reads.qzv

# Filtering of the readings according to their quality.
qiime quality-filter q-score-joined --i-demux joined-reads.qza \
                                    --o-filter-stats filt_stats.qza \
                                    --o-filtered-sequences joined-filt-reads.qza

qiime demux summarize --i-data joined-filt-reads.qza --o-visualization joined-filt-reads.qzv

EOT
```

Output

```
Saved SampleData[JoinedSequencesWithQuality] to: joined-reads.qza
Saved Visualization to: joined-reads.qzv
Saved SampleData[JoinedSequencesWithQuality] to: joined-filt-reads.qza
Saved QualityFilterStats to: filt_stats.qza
Saved Visualization to: joined-filt-reads.qzv
```

Determination of ASV/features

Script 1.2.6 (bash)

```
%%bash
ssh microbioinf@192.168.56.101 /bin/bash <<"EOT"
export PATH=$PATH:/home/microbioinf/miniconda3/bin
source activate qiime2-2018.11
cd Documentos/Tema_2

rm -rf deblur_output
qiime deblur denoise-16S --i-demultiplexed-seqs joined-filt-reads.qza \
                         --p-trim-length 372 \
                         --p-sample-stats \
                         --p-jobs-to-start 2 \
                         --p-min-reads 1 \
                         --output-dir deblur_output

qiime feature-table summarize --i-table deblur_output/table.qza --o-visualization
 ↪  deblur_output/deblur_table_summary.qzv


EOT
```

Output

```
Saved FeatureTable[Frequency] to: deblur_output/table.qza
Saved FeatureData[Sequence] to: deblur_output/representative_sequences.qza
Saved DeblurStats to: deblur_output/stats.qza
Saved Visualization to: deblur_output/deblur_table_summary.qzv
```

Script 1.2.7 (bash)

```
%%bash
ssh microbioinf@192.168.56.101 /bin/bash <<"EOT"
export PATH=$PATH:/home/microbioinf/miniconda3/bin
source activate qiime2-2018.11
cd Documentos/Tema_2

# Exclude sOTUS that have a frequency below 0.1% of the mean depth.
# This threshold would exclude those sOTUs that are due to Illumina sequencing errors (0.1%
↪  of total reads).

qiime feature-table filter-features --i-table deblur_output/table.qza \
                                   --p-min-frequency 30 \
                                   --p-min-samples 1 \
                                   --o-filtered-table deblur_output/deblur_table_filt.qza

# Exclude low frequency sOTUS
qiime feature-table filter-seqs --i-data deblur_output/representative_sequences.qza \
                                --i-table deblur_output/deblur_table_filt.qza \
                                --o-filtered-data deblur_output/rep_seqs_filt.qza


# Summarize
qiime feature-table summarize --i-table deblur_output/deblur_table_filt.qza
↪  --o-visualization deblur_output/deblur_table_filt_summary.qzv

EOT
```

Output

```
Saved FeatureTable[Frequency] to: deblur_output/deblur_table_filt.qza
Saved FeatureData[Sequence] to: deblur_output/rep_seqs_filt.qza
Saved Visualization to: deblur_output/deblur_table_filt_summary.qzv
```

1.2.3 Phylogenetic distances determination using FastTree

Script 1.2.8 (bash)

```
%%bash
ssh microbioinf@192.168.56.101 /bin/bash <<"EOT"
export PATH=$PATH:/home/microbioinf/miniconda3/bin
source activate qiime2-2018.11
cd Documentos/Tema_2
rm -rf tree_out
mkdir tree_out
#Sequence alignment
qiime alignment mafft --i-sequences deblur_output/rep_seqs_filt.qza \
                      --p-n-threads 3 \
                      --o-alignment tree_out/rep_seqs_filt_aligned.qza
```

```

#Mask hypervariable regions
qiime alignment mask --i-alignment tree_out/rep_seqs_filt_aligned.qza \
                     --o-masked-alignment tree_out/rep_seqs_filt_aligned_masked.qza

#Calculate phylogenie

qiime phylogeny fasttree --i-alignment tree_out/rep_seqs_filt_aligned_masked.qza \
                         --p-n-threads 2 \
                         --o-tree tree_out/rep_seqs_filt_aligned_masked_tree

#Root the tree

qiime phylogeny midpoint-root --i-tree tree_out/rep_seqs_filt_aligned_masked_tree.qza \
                              --o-rooted-tree
                               ↪  tree_out/rep_seqs_filt_aligned_masked_tree_rooted.qza

EOT
```

Output

```
Saved FeatureData[AlignedSequence] to: tree_out/rep_seqs_filt_aligned.qza
Saved FeatureData[AlignedSequence] to: tree_out/rep_seqs_filt_aligned_masked.qza
Saved Phylogeny[Unrooted] to: tree_out/rep_seqs_filt_aligned_masked_tree.qza
Saved Phylogeny[Rooted] to: tree_out/rep_seqs_filt_aligned_masked_tree_rooted.qza
```

1.2.4 Taxonomic assignment

Assignment database training

Script 1.2.9 (bash)

```
%%bash
ssh microbioinf@192.168.56.101 /bin/bash <<"EOT"
export PATH=$PATH:/home/microbioinf/miniconda3/bin
source activate qiime2-2018.11
cd Documentos/Tema_2

#Import reference sequences
qiime tools import \
  --type 'FeatureData[Sequence]' \
  --input-path 85_otus.fasta \
  --output-path 85_otus.qza

#Import reference taxonomy
qiime tools import \
  --type 'FeatureData[Taxonomy]' \
  --input-format HeaderlessTSVTaxonomyFormat \
  --input-path 85_otu_taxonomy.txt \
  --output-path ref-taxonomy.qza
```

```

EOT
```

Output

```
Imported 85_otus.fasta as DNASequencesDirectoryFormat to 85_otus.qza
Imported 85_otu_taxonomy.txt as HeaderlessTSVTaxonomyFormat to ref-taxonomy.qza
```

Script 1.2.10 (bash)

```bash
%%bash
ssh microbioinf@192.168.56.101 /bin/bash <<"EOT"
export PATH=$PATH:/home/microbioinf/miniconda3/bin
source activate qiime2-2018.11
cd Documentos/Tema_2

#Trim reference sequences to the region intra-primers according to primers.txt
qiime feature-classifier extract-reads \
  --i-sequences 85_otus.qza \
  --p-f-primer GTGCCAGCMGCCGCGGTAA \
  --p-r-primer CCGTCAATTCMTTTRAGTTT \
  --p-min-length 100 \
  --p-max-length 400 \
  --o-reads ref-seqs.qza

#Generate classifier naive-bayes
qiime feature-classifier fit-classifier-naive-bayes \
  --i-reference-reads ref-seqs.qza \
  --i-reference-taxonomy ref-taxonomy.qza \
  --o-classifier classifier.qza

EOT
```

Output

```
Saved FeatureData[Sequence] to: ref-seqs.qza
Saved TaxonomicClassifier to: classifier.qza
```

Taxonomic assignment of representative sequences

Script 1.2.11 (bash)

```bash
%%bash
ssh microbioinf@192.168.56.101 /bin/bash <<"EOT"
export PATH=$PATH:/home/microbioinf/miniconda3/bin
source activate qiime2-2018.11
cd Documentos/Tema_2

rm -rf taxa
```

```
#Taxonomic assignment of the representative sequences of each sOTU
qiime feature-classifier classify-sklearn --i-reads deblur_output/rep_seqs_filt.qza \
--i-classifier classifier.qza \
--p-n-jobs 2 \
--output-dir taxa

#Export to tabular file
qiime tools export --input-path taxa/classification.qza --output-path taxa

#Obtain interactive graph to visualize the abundance in each sOTU per sample
qiime taxa barplot --i-table deblur_output/deblur_table_filt.qza \
--i-taxonomy taxa/classification.qza \
--m-metadata-file metadata.txt \
--o-visualization taxa/taxa_barplot.qzv

EOT
```

Output

```
Saved FeatureData[Taxonomy] to: taxa/classification.qza
Exported taxa/classification.qza as TSVTaxonomyDirectoryFormat to directory taxa
Saved Visualization to: taxa/taxa_barplot.qzv
```

1.2.5 Diversity analysis

Alpha diversity

Script 1.2.12 (bash)

```
%%bash
ssh microbioinf@192.168.56.101 /bin/bash <<"EOT"
export PATH=$PATH:/home/microbioinf/miniconda3/bin
source activate qiime2-2018.11
cd Documentos/Tema_2

# Obtain rarefaction curves with max-depht = num of reads of the richest sample
qiime diversity alpha-rarefaction --i-table deblur_output/deblur_table_filt.qza \
                                  --p-max-depth 41934 \
                                  --p-steps 20 \
                                  --i-phylogeny
                                  ↪ tree_out/rep_seqs_filt_aligned_masked_tree_rooted.qza \
                                  --m-metadata-file metadata.txt \
                                  --o-visualization rarefaction_curves.qzv

EOT
```

Output

```
Saved Visualization to: rarefaction_curves.qzv
```

Beta diversity

Script 1.2.13 (bash)

```
%%bash
ssh microbioinf@192.168.56.101 /bin/bash <<"EOT"
export PATH=$PATH:/home/microbioinf/miniconda3/bin
source activate qiime2-2018.11
cd Documentos/Tema_2
rm -rf diversity
# Obtain metrics curves with sampling-depth 19845 = num of reads of the poorest sample
qiime diversity core-metrics-phylogenetic --i-table deblur_output/deblur_table_filt.qza \
                                          --i-phylogeny tree_out/rep_seqs_filt_aligned_maske
                                          ↪  d_tree_rooted.qza
                                          ↪  \
                                          --p-sampling-depth 15296 \
                                          --m-metadata-file metadata.txt \
                                          --p-n-jobs 2 \
                                          --output-dir diversity

EOT
```

Output

```
Saved FeatureTable[Frequency] to: diversity/rarefied_table.qza
Saved SampleData[AlphaDiversity] % Properties(['phylogenetic']) to:
↪  diversity/faith_pd_vector.qza
Saved SampleData[AlphaDiversity] to: diversity/observed_otus_vector.qza
Saved SampleData[AlphaDiversity] to: diversity/shannon_vector.qza
Saved SampleData[AlphaDiversity] to: diversity/evenness_vector.qza
Saved DistanceMatrix % Properties(['phylogenetic']) to:
↪  diversity/unweighted_unifrac_distance_matrix.qza
Saved DistanceMatrix % Properties(['phylogenetic']) to:
↪  diversity/weighted_unifrac_distance_matrix.qza
Saved DistanceMatrix to: diversity/jaccard_distance_matrix.qza
Saved DistanceMatrix to: diversity/bray_curtis_distance_matrix.qza
Saved PCoAResults to: diversity/unweighted_unifrac_pcoa_results.qza
Saved PCoAResults to: diversity/weighted_unifrac_pcoa_results.qza
Saved PCoAResults to: diversity/jaccard_pcoa_results.qza
Saved PCoAResults to: diversity/bray_curtis_pcoa_results.qza
Saved Visualization to: diversity/unweighted_unifrac_emperor.qzv
Saved Visualization to: diversity/weighted_unifrac_emperor.qzv
Saved Visualization to: diversity/jaccard_emperor.qzv
Saved Visualization to: diversity/bray_curtis_emperor.qzv
```

Script 1.2.14 (bash)

```
%%bash
ssh microbioinf@192.168.56.101 /bin/bash <<"EOT"
export PATH=$PATH:/home/microbioinf/miniconda3/bin
source activate qiime2-2018.11
cd Documentos/Tema_2
```

```
qiime taxa collapse --i-table deblur_output/deblur_table_filt.qza \
                    --o-collapsed-table deblur_output/L3_collapse_table.qza \
                    --p-level 3 \
                    --i-taxonomy taxa/classification.qza

qiime tools export --input-path deblur_output/L3_collapse_table.qza \
                   --output-path lefse_table/

biom convert -i lefse_table/feature-table.biom \
             -o lefse_table/feature-table.txt \
             --header-key ‘‘taxonomy’’ --to-tsv

EOT
```

Output

```
Saved FeatureTable[Frequency] to: deblur_output/L3_collapse_table.qza
Exported deblur_output/L3_collapse_table.qza as BIOMV210DirFmt to directory lefse_table/
```