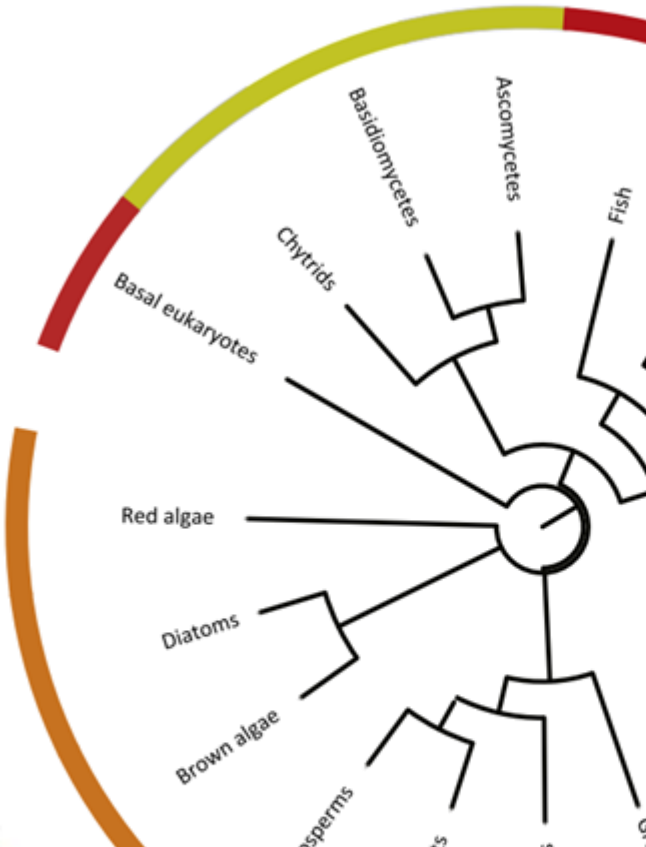
Introduction to Bioinformatics analysis of Metabarcoding data

Practical part

Alvaro Sebastián Yagüe

www.sixthresearcher.com

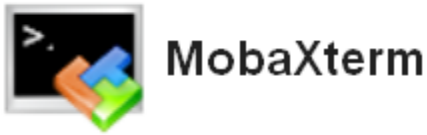
@SixthResearcher

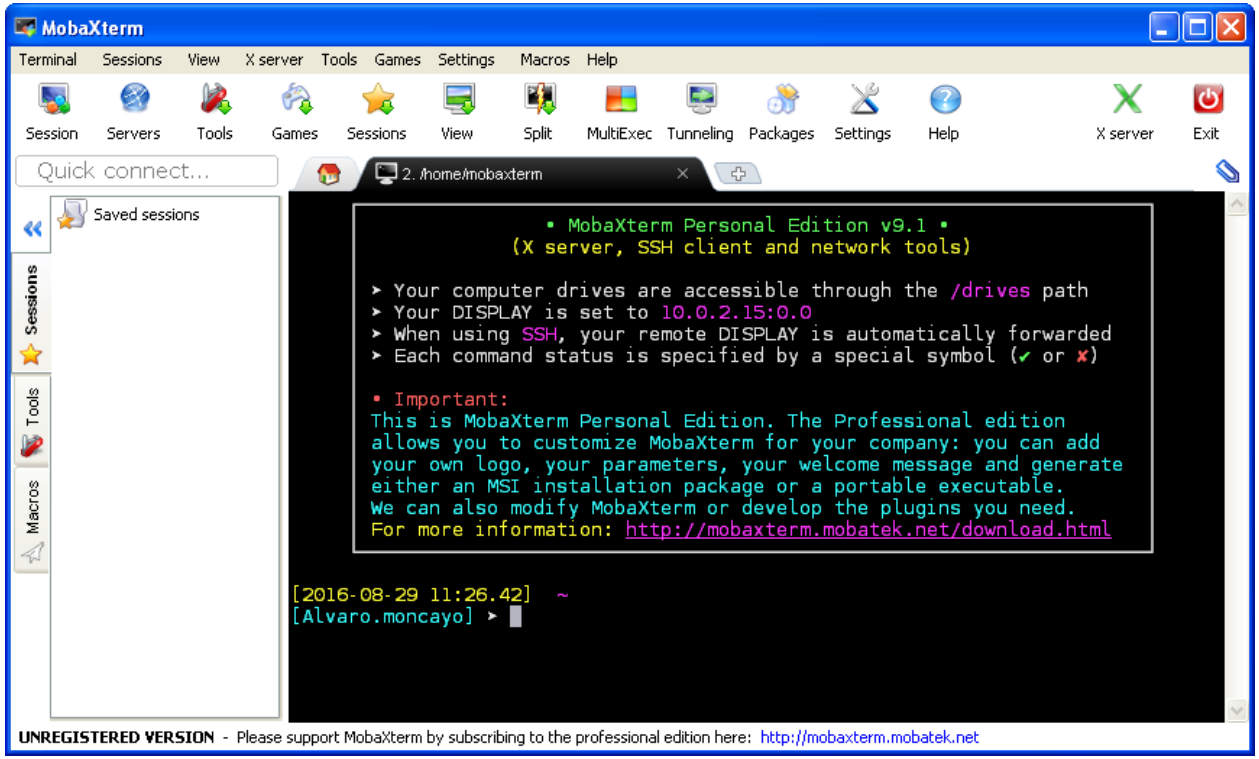


Install a Linux emulator in Windows

We will download and install MobaXterm, a free “Enhanced terminal for Windows”.

http://mobaxterm.mobatek.net/download.html





- Retrieve experimental data
- Checking of reads
- Read pre-processing
- AmpliTAXO analysis
- Results interpretation

www.sixthresearcher.com
@SixthResearcher

- Retrieve experimental data
- Checking of reads
- Read pre-processing
- AmpliTAXO analysis
- Results interpretation

www.sixthresearcher.com
@SixthResearcher

Example data

- The example data has been produced in a experiment about the **variation in the gut microbiome on mice health (Kozich et al.)**.
- The marker/barcode used is the **V4 region of the 16S rRNA** gene. This region is **short length (~250bp)** and allows for fully overlapping forward and reverse reads.
- To make the example easier to analyze, only part of the data is provided: **10 mice fecal samples** (10 times points for one animal).
- In addition, a **mock community** with known composition of genomic DNA from **21 bacterial strains** is also sequenced as **internal control**.
- **Illumina MiSeq technology** was used to produce 2x250bp **paired-end reads in FASTQ files**, with each pair representing the two sets of reads per sample.

Kozich,J.J. *et al.* (2013) Development of a dual-index sequencing strategy and curation pipeline for analyzing amplicon sequence data on the miseq illumina sequencing platform. *Appl. Environ. Microbiol.*, **79**, 5112–5120.

Example data

First, we will download, decompress and list the gut microbiome example files.

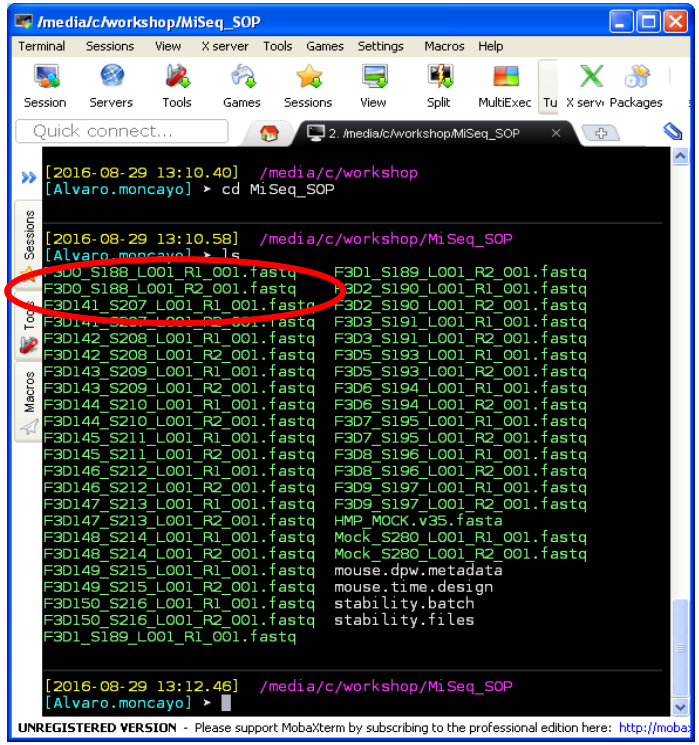
- ✓ Download the example files into a folder called 'workshop' in the C disk: http://www.mothur.org/w/images/d/d6/MiSeqSOPData.zip
- ✓ Move to the 'workshop' folder:

  `> cd /media/c/workshop/`
- ✓ Decompress the file:

  `> unzip MiSeqSOPData.zip`
- ✓ Move to the example folder:

  `> cd MiSeq_SOP`
- ✓ List the example files:

  `> ls`



- Retrieve experimental data
- Checking of reads
- Read pre-processing
- AmpliTAXO analysis
- Results interpretation

www.sixthresearcher.com
@SixthResearcher

Manual checking of reads

Let's check the number and length of the reads.

- ✓ See the contents of the 'F3D0_S188_L001_R1_001.fastq' FASTQ file:

```
> less F3D0_S188_L001_R1_001.fastq
```

- ✓ How many reads contain the 'F3D0_S188_L001_R1_001.fastq' FASTQ file?

```
> cat F3D0_S188_L001_R1_001.fastq | echo $((`wc -l`/4))
```

Counting the number of sequences in a FASTA file is easier:

```
> grep -c "^>" HMP_MOCK.v35.fasta
```

- ✓ How long are the reads? (for example, the sequence at the 2nd line of the file)

```
> awk 'NR==2 {print length($0)}' F3D0_S188_L001_R1_001.fastq
```

Counting the number of reads for each length is more complex:

```
> awk '{ c++; if((c-2)%4==0){ ++a[length()] } } END{ print "len\tseqs";
for (i in a) print i"\t"a[i]}' F3D0_S188_L001_R1_001.fastq
```

Manual checking of reads

- ✓ Show the contents of the ‘F3D0_S188_L001_R1_001.fastq’ FASTQ file:

  `> less F3D0_S188_L001_R1_001.fastq`

  Select and copy the first read into the Notepad. To exit push ‘Q’ key.

- ✓ Show the contents of the ‘F3D0_S188_L001_R1_001.fastq’ FASTQ file:

  `> less F3D0_S188_L001_R2_001.fastq`

  Select and copy the first read into the Notepad. To exit push ‘Q’ key.

COPY

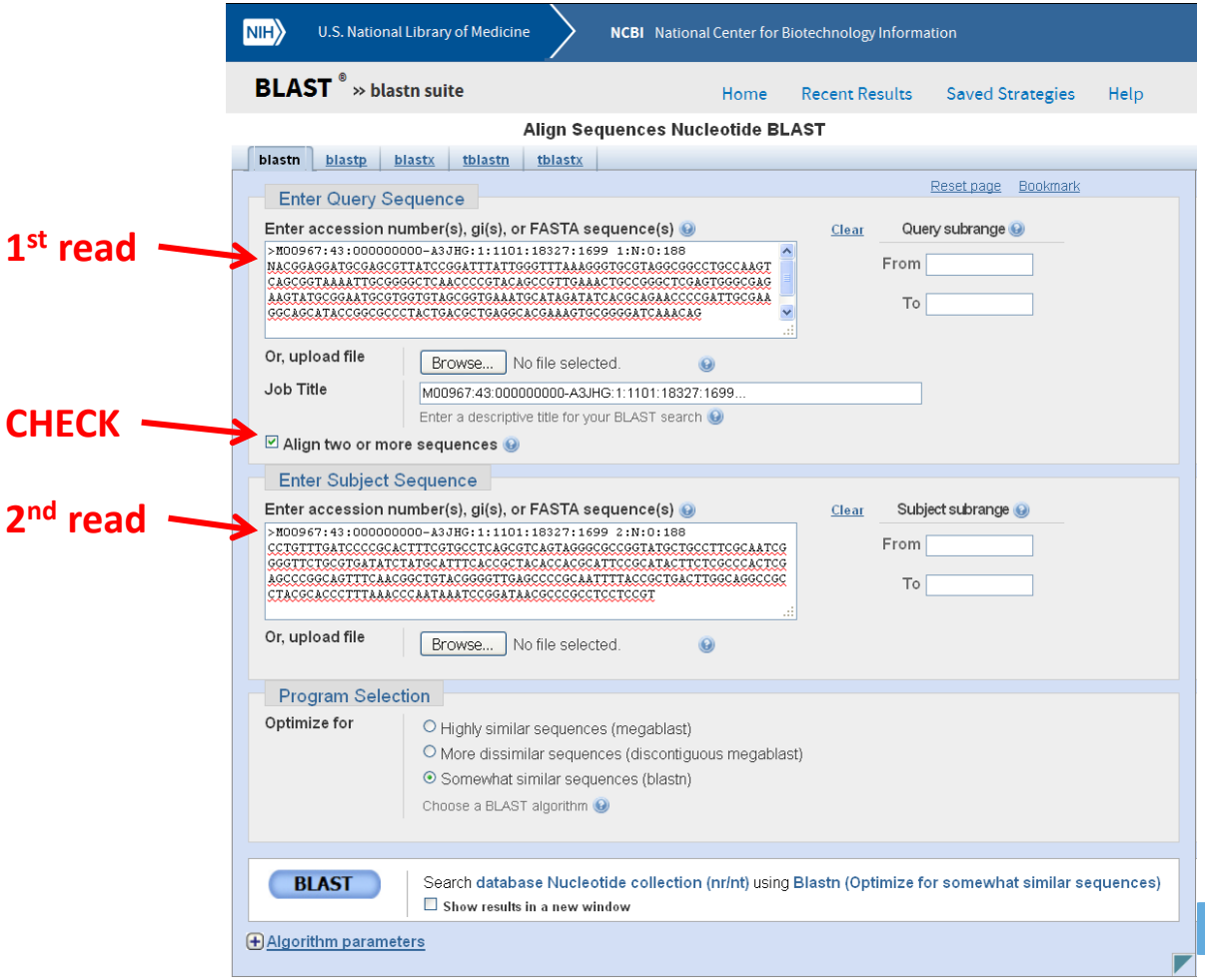
```
@M00967:43:000000000-A3JHG:1:1101:18327:1699 1:N:0:188
NACGGAGGATGCGAGCGTTATCCGGATTTATTGGGTTTAAAGGGTGCGTAGGCGGCCTGCCAAGTCAGCGGTAAAATTGCGGGGCTCAACCCCGTACAGCCGTTGAAACTGCCGGGCTCGAGTGGGCGAGAAGTATGCGGAATGCGTGGTGTAGCGGTGA
AATGCATAGATATCACGCAGAACCCCGATTGCGAAGGCAGCATACCGGCGCCCTACTGACGCTGAGGCACGAAAGTGCGGGGATCAAACAG
+
#>>AABABBFFFGGGGGGGGGGGGGGGHHHHHHHGGGHHHHHGHGGGGGGHGGGGGGHHHHHHHHHHGGGGGHHHHGHGGGGGGHHBGHGDGGGGGHHHGGGGHHHHHHHHGGGGGHG@DHHGHEGGGGGGBFGGEGGGGGGGG.DFEFFFFFFFDCF
FFFFFFFFFFFFFFFFFFFFFFFFFDFDFFFEFFCFF?FDFFFFFFFFAFFFFFFFFFFFBDDFFFFFFEFADFFFFFBAFFFA?EFFFBFF
@M00967:43:000000000-A3JHG:1:1101:14069:1827 1:N:0:188
TACGGAGGATGCGAGCGTTATCCGGATTTATTGGGTTTAAAGGGTGCGTAGGCGGCCTGCCAAGTCAGCGGTAAAATTGCGGGGCTCAACCCCGTACAGCCGTTGAAACTGCCGGGCTCGAGTGGGCGAGAAGTATGCGGAATGCGTGGTGTAGCGGTGA
AATGCATAGATATCACGCAGAACCCCGATTGCGAAGGCAGCATACCGGCGCCCTACTGACGCTGAGGCACGAAAGTGCGGGGATCAAACAG
+
3AA?ABBDBFFBEGGEGGGGAFFGGGGGHHHCGGGGGGHFGHGGCFDEFGGGHGGGEGF1GGFGHHHHHGGEGGHHHHHFGGGGGGHHHHHGGGGCDDGHHGGGFHHHHHHHHCD@CCHGGGGHEHGGG@GFGGGGGGG@BGGGEGCEBFFFBFFB;9@E
FFFEFFFFFFFFFFFFAFBBBFFFFFBBBFFFFBBBFFFFFFFFFFFBBBBBBBFFFFFFFFFDDFAFFFFF.AF9/FBBBBB.EAFFE?F
@M00967:43:000000000-A3JHG:1:1101:18044:1900 1:N:0:188
TACGGAGGATGCGAGCGTTGTCCGGAATCACTGGGCGTAAAGGGCGCGTAGGCGGTTTAATAAGTCAGTGGTGAAAACTGAGGGCTCAACCCTCAGCCTGCCACTGATACTGTTAGACTTGAGTATGGAAGAGGAGAATGGAATTCCTAGTGTAGCGGTG
AAATGCGTAGATATTAGGAGGAACACCAGTGGCGAAGGCGATTCTCTGGGCCAAGACTGACGCTGAGGCGCGAAAGCGTGGGGAGCAAACA
+
BA@BBBABBFFFGGGGGGGGGHGGGGGHHHHGHHHGGGGHHHGGGGGGGGHGGGGGGFHFHHHFHHHHHHGFHHHHHHFHGGFGHHGGHGGHHFHHHHHHHHHHHHHHHGHHHGHDHHHHHHHHHHGHDGFFHHHDHHHHHHHHHHHHFHHHDGHHGGGG
GHHHHHGGGGEGGGGGBF0BGGGGGGGGFFFGGGDGFFFFBCFBFFBFFBEEFFEFFFFFFFFFBDA9EEFFF>DFFFFDDFFFADDFFFF.
@M00967:43:000000000-A3JHG:1:1101:13234:1983 1:N:0:188
TACGGAGGATGCGAGCGTTATCCGGATTTATTGGGTTTAAAGGGTGCGCAGGCGGAAGATCAAGTCAGCGGTAAAATTGAGAGGCTCAACCTCTTCGAGCCGTTGAAACTGGTTTTCTTGAGTGAGCGAGAAGTATGCGGAATGCGTGGTGTAGCGGTGA
AATGCATAGATATCACGCAGAACTCCGATTGCGAAGGCAGCATACCGGCGCTCAACTGACGCTCATGCACGAAAGTGTGGGTATCGAACAG
+
BCBCCCCCCFFFGGGGGGGGGHGGGGGHHHHHHHGGGHHHHHGHGGGGGGGGGGGGHHHHHHHHHHHGGGGGHHHHHHHHHGGHHHHHHHHGHHHGHGGGGGGHHHHHHHHGHHGHHHGHHHHHHHGGGGGHHHHHHHGGGGGHHGGGGHGHHHHGGGG
F3D0_S188_L001_R1_001.fastq
```

Manual checking of reads

- ✓ Change the '@' symbol for '>' at the beginning of the read names.

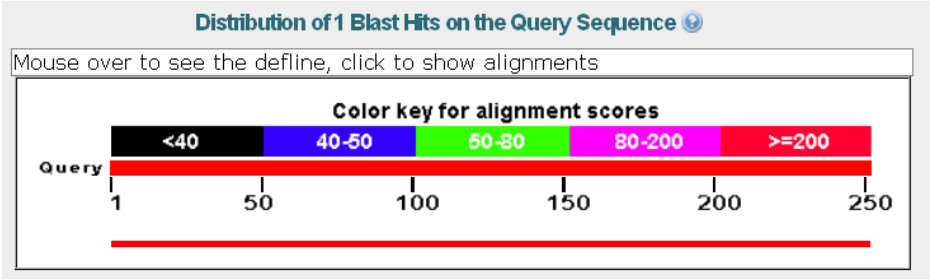
Manual checking of reads

- ✓ Copy name-modified reads and paste them into Blastn web form:
  http://blast.ncbi.nlm.nih.gov/Blast.cgi?PAGE=Blastn&BLAST_SPEC=blast2seq



Manual checking of reads

Blastn alignment results show how **both paired-end reads totally overlap** as we expected because the experimental design.



M00967:43:000000000-A3JHG:1:1101:18327:1699 2:N:0:188

Sequence ID: lcl|Query_156151 **Length:** 251 **Number of Matches:** 1

Range 1: 2 to 251 Graphics ▼ Next Match ▲ Previous Match

Score	Expect	Identities	Gaps	Strand
451 bits(244)	6e-132	248/250(99%)	0/250(0%)	Plus/Minus

1st read →
2nd read →

```
Query  2    ACGGAGGATGCGAGCGTTATCCGGATTTATTGGGTTTAAAGGGTGCGTAGGCGGCCTGCC  61
            |||||||| ||| |||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  251  ACGGAGGAGGCGGGCGTTATCCGGATTTATTGGGTTTAAAGGGTGCGTAGGCGGCCTGCC  192

Query  62   AAGTCAGCGGTAAAATTGCGGGGCTCAACCCCGTACAGCCGTTGAAACTGCCGGGCTCGA  121
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  191  AAGTCAGCGGTAAAATTGCGGGGCTCAACCCCGTACAGCCGTTGAAACTGCCGGGCTCGA  132

Query  122  GTGGGCGAGAAGTATGCGGAATGCGTGGTGTAGCGGTGAAATGCATAGATATCACGCAGA  181
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  131  GTGGGCGAGAAGTATGCGGAATGCGTGGTGTAGCGGTGAAATGCATAGATATCACGCAGA  72

Query  182  ACCCCGATTGCGAAGGCAGCATACCGGCGCCCTACTGACGCTGAGGCACGAAAGTGCGGG  241
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  71   ACCCCGATTGCGAAGGCAGCATACCGGCGCCCTACTGACGCTGAGGCACGAAAGTGCGGG  12

Query  242  GATCAAACAG  251
            ||||||||||
Sbjct  11   GATCAAACAG  2
```

ww.sixthresearcher.com
@SixthResearcher

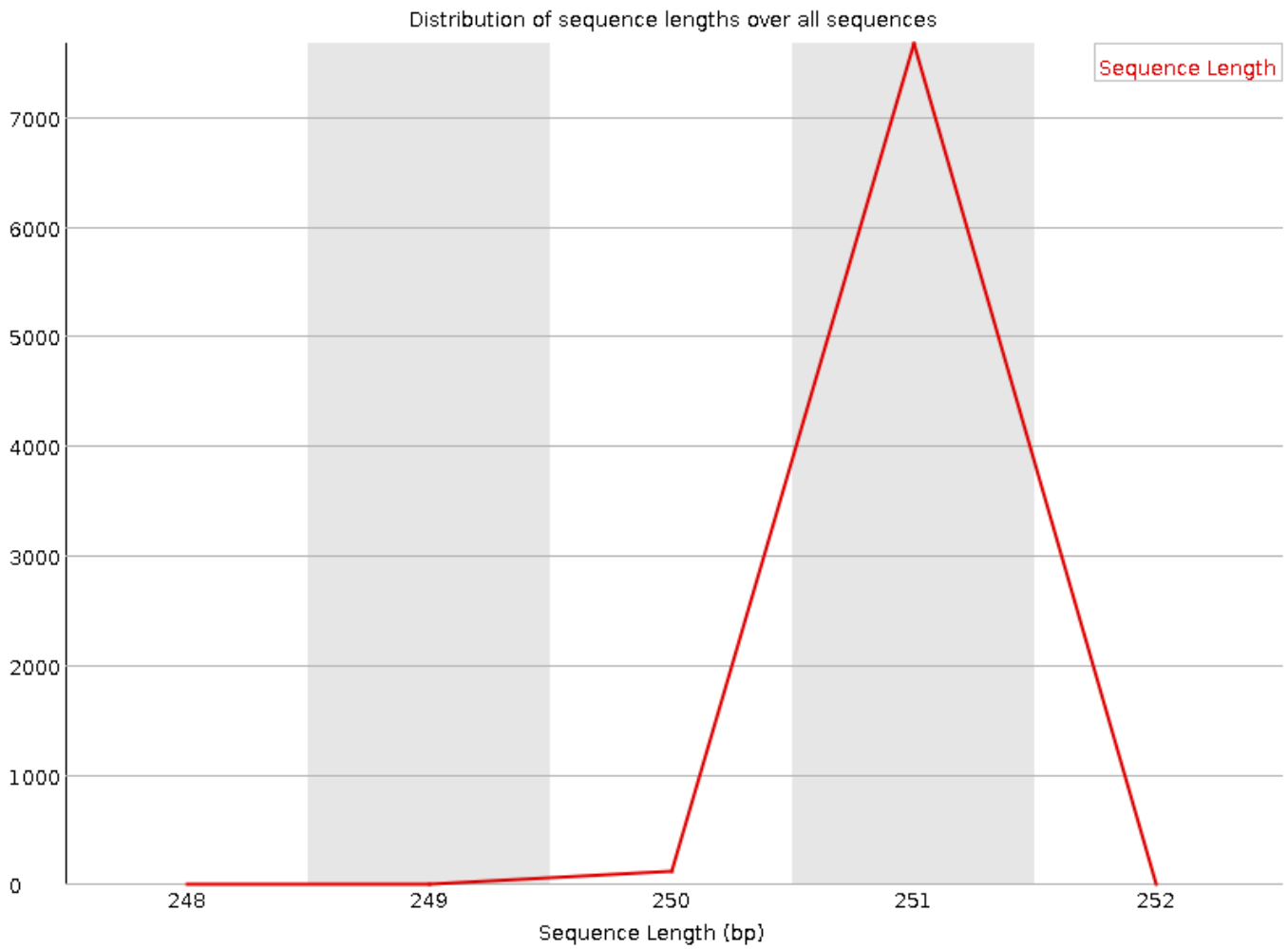
Automatic checking of reads

Let's check automatically the quality of the reads.

- ✓ Download and install the FASTQC software:
  http://www.bioinformatics.babraham.ac.uk/projects/fastqc/
- ✓ Check the 'F3D0_S188_L001_R1_001.fastq' FASTQ file:
  ```
  > fastqc F3D0_S188_L001_R1_001.fastq
  ```
- ✓ Check the 'F3D0_S188_L001_R2_001.fastq' FASTQ file:
  ```
  > fastqc F3D0_S188_L001_R2_001.fastq
  ```
- ✓ Open the resulting files 'F3D0_S188_L001_R1_001_fastqc.html' and 'F3D0_S188_L001_R2_001_fastqc.html' with an internet browser.

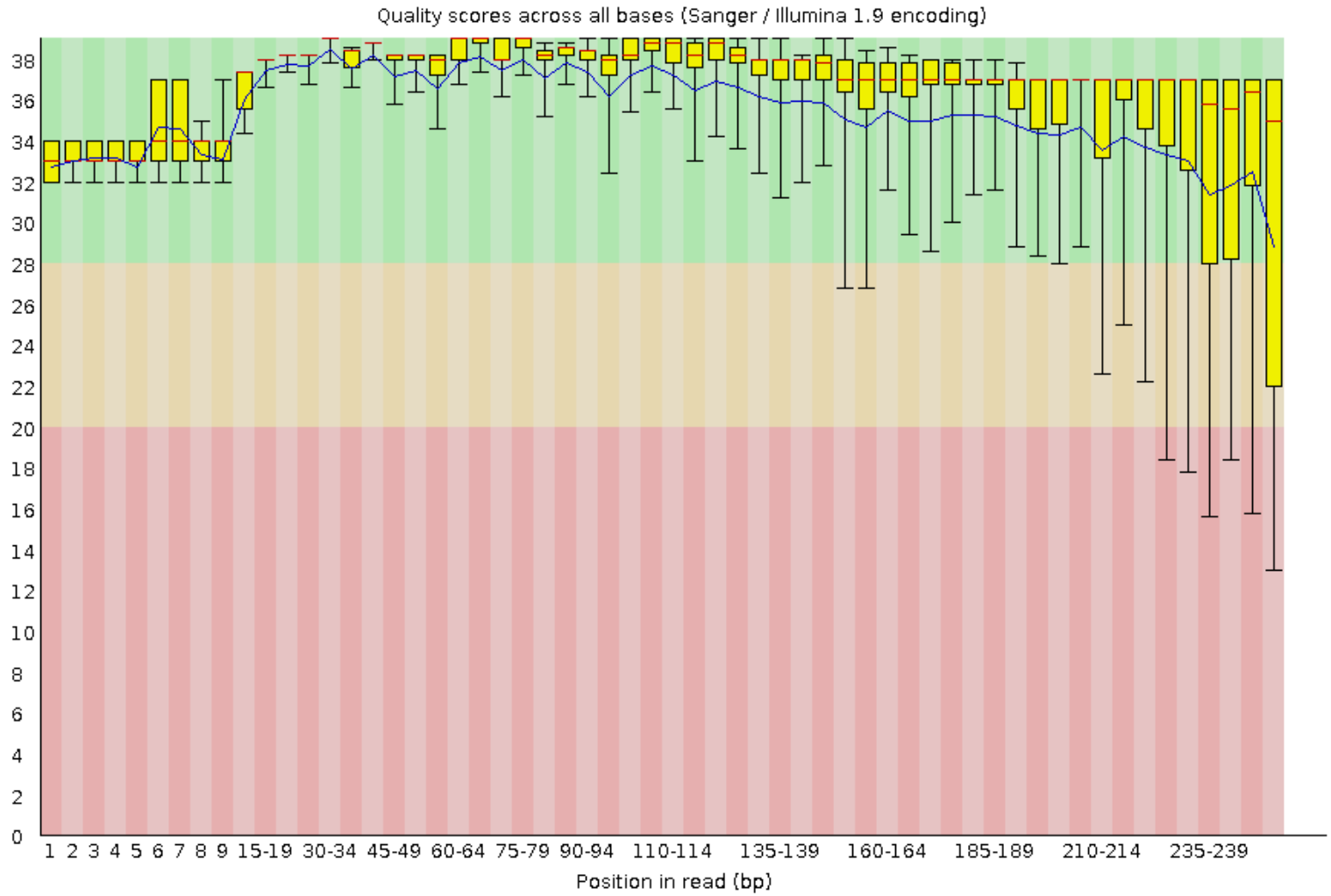
Automatic checking of reads

Sequence length distribution:



Automatic checking of reads

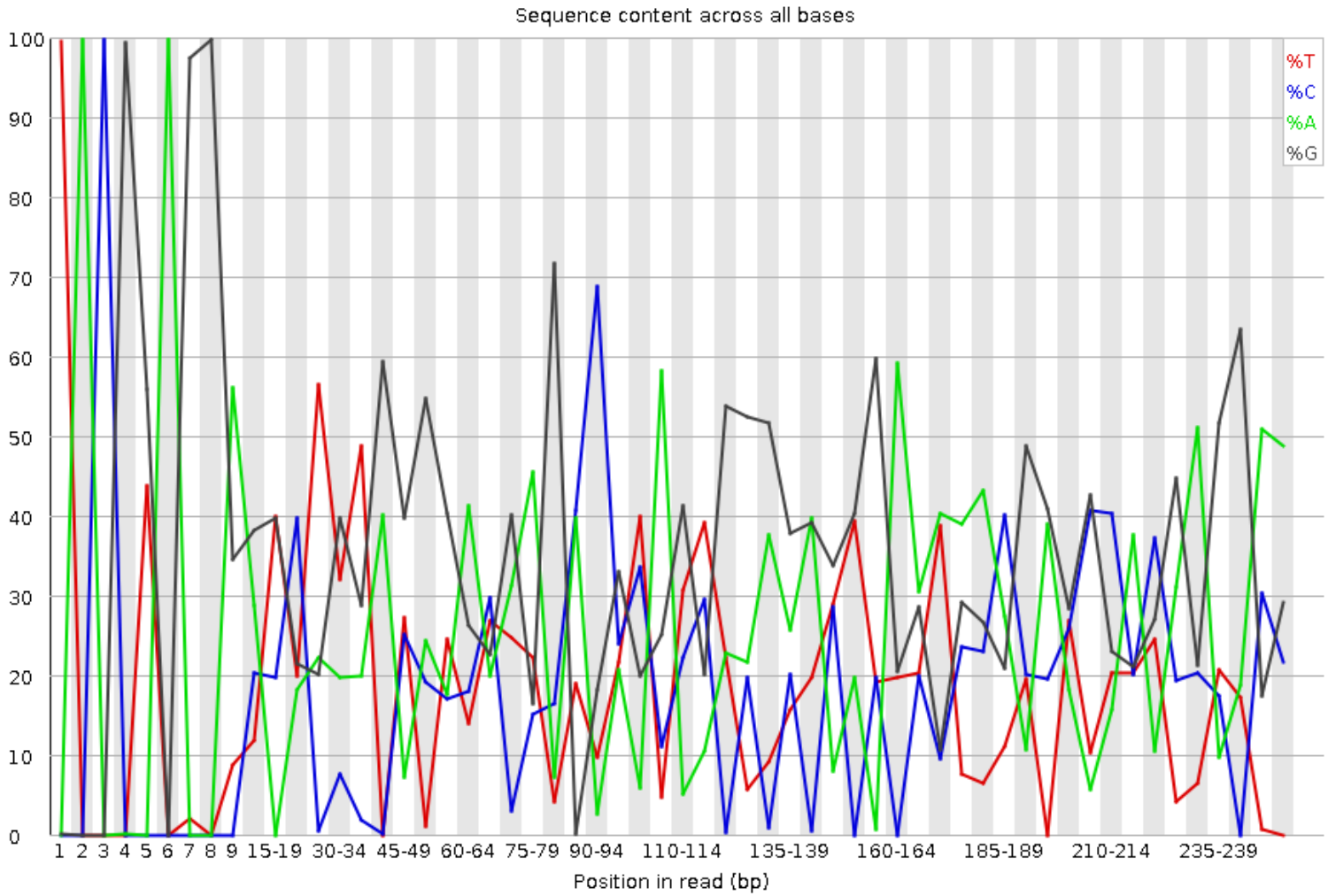
Per base sequence quality:



www.sixthresearcher.com
@SixthResearcher

Automatic checking of reads

Per base sequence content:



www.sixthresearcher.com
@SixthResearcher

Packing and compressing reads

Now we are going to pack and compress the paired-end files into two files: **example_reads_R1.tar.gz** and **example_reads_R2.tar.gz**

- ✓ We will pack and compress together all the R1 reads:

```
> tar -cvzf example_reads_R1.tar.gz *_R1_001.fastq
```

- ✓ We will do the same for all the R2 reads:

```
> tar -cvzf example_reads_R2.tar.gz *_R2_001.fastq
```

```
[Alvaro.moncayo] ➤ tar -cvzf example_reads_R1.tar.gz *_R1_001.fastq
F3D0_S188_L001_R1_001.fastq
F3D1_S189_L001_R1_001.fastq
F3D141_S207_L001_R1_001.fastq
F3D142_S208_L001_R1_001.fastq
F3D143_S209_L001_R1_001.fastq
F3D144_S210_L001_R1_001.fastq
F3D145_S211_L001_R1_001.fastq
F3D146_S212_L001_R1_001.fastq
F3D147_S213_L001_R1_001.fastq
F3D148_S214_L001_R1_001.fastq
F3D149_S215_L001_R1_001.fastq
F3D150_S216_L001_R1_001.fastq
F3D2_S190_L001_R1_001.fastq
F3D3_S191_L001_R1_001.fastq
F3D5_S193_L001_R1_001.fastq
F3D6_S194_L001_R1_001.fastq
F3D7_S195_L001_R1_001.fastq
F3D8_S196_L001_R1_001.fastq
F3D9_S197_L001_R1_001.fastq
Mock_S280_L001_R1_001.fastq
```

The output files will be packed by the TAR Linux command and compressed in GZIP format.

Packing and compressing reads

- ✓ Let's check the size of FASTQ files before compressing:

```
> du -ch *.fastq | grep  total
```

- ✓ And the size of the TAR.GZ files after compressing:

```
> du -ch *.tar.gz | grep  total
```

```
[2016-10-17 16:17.46]  /media/c/workshop/MiSeq_SOP
[Alvaro.aeiou] ➤ du -ch *.fastq | grep  total
163.2M  total

[2016-10-17 16:17.51]  /media/c/workshop/MiSeq_SOP
[Alvaro.aeiou] ➤ du -ch *.tar.gz | grep  total
33.2M   total
```

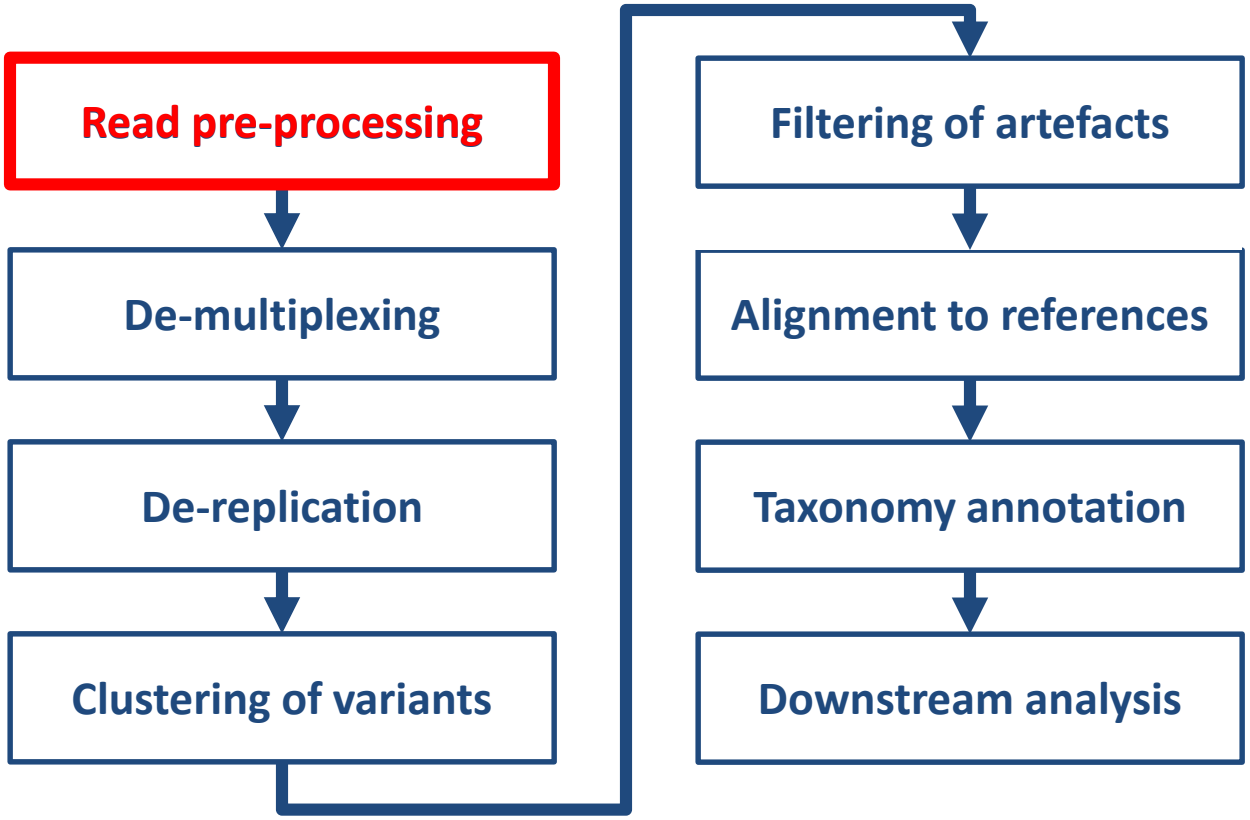
The compressed size is less than 5 times the original!

- ✓ If we want to extract the TAR.GZ files:

```
> tar -xvzf example_reads_R1.tar.gz

> tar -xvzf example_reads_R2.tar.gz
```

- Retrieve experimental data
- Checking of reads
- Read pre-processing
- AmpliTAXO analysis
- Results interpretation

Analysis pipeline

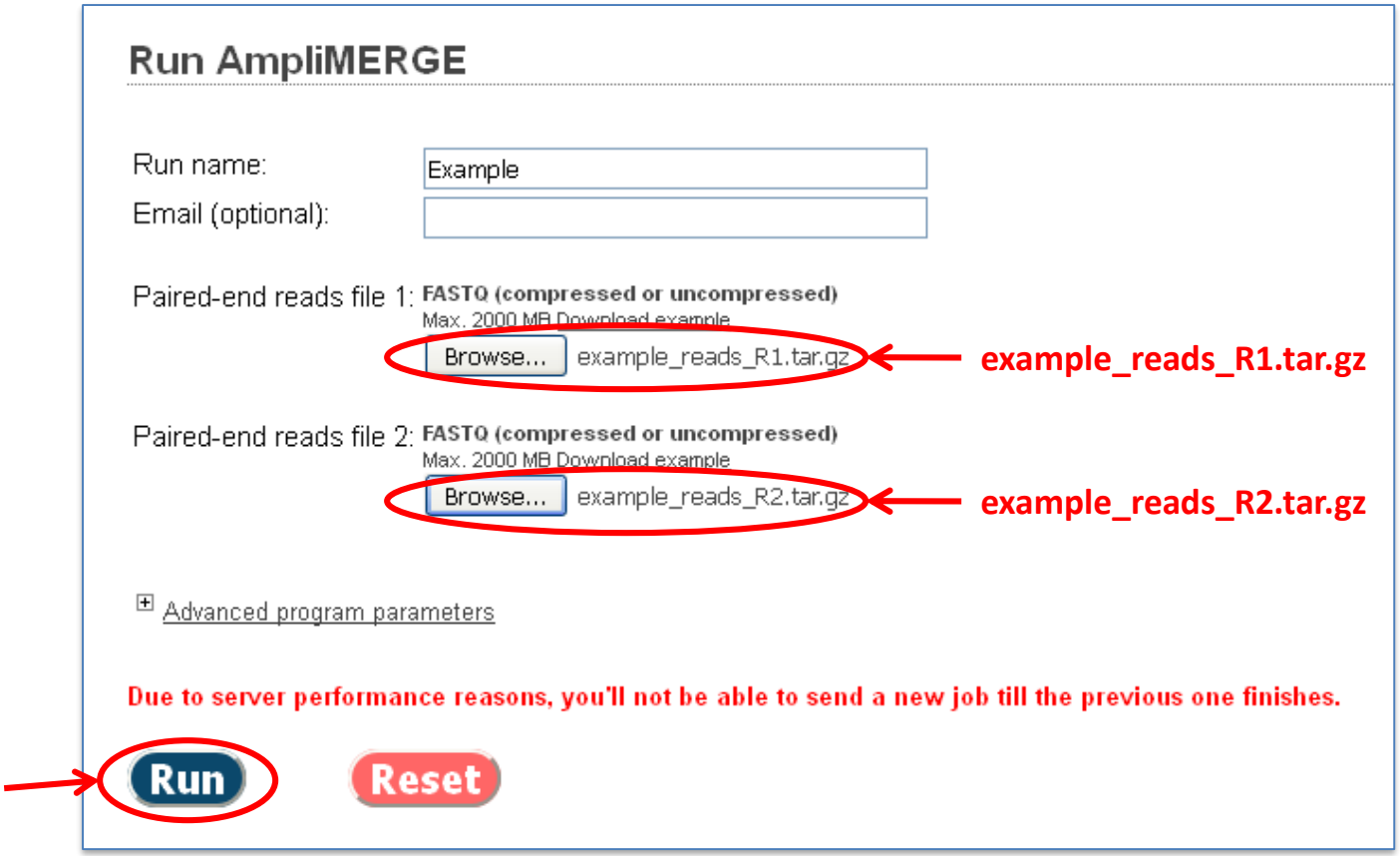


www.sixthresearcher.com
@SixthResearcher

Merging paired-end reads

- ✓ We will use the two compressed paired-end files as input for **AmpliMERGE**:

  http://evobiolab.biol.amu.edu.pl/amplisat/index.php?amplimerge



Merging paired-end reads

✓ We will download AmpliMERGE results into the example folder.

AmpliMERGE results for Example

Download AmpliMERGE analysis results. ← **Example.tar.gz**

Analysis details:

```
Running 'bin/ampliMERGE.pl -gz -i results/5fa3172ce81755a14850b9650d800c5b/example_reads_R1.tar.gz

Merging paired-end files.
        Merging 'F3D0_S188_L001_R1_001.fastq' and 'F3D0_S188_L001_R2_001.fastq'
        Merged 7060 reads from 7793
        Merging 'F3D1_S189_L001_R1_001.fastq' and 'F3D1_S189_L001_R2_001.fastq'
        Merged 5404 reads from 5869
        Merging 'F3D141_S207_L001_R1_001.fastq' and 'F3D141_S207_L001_R2_001.fastq'
        Merged 5317 reads from 5958
        Merging 'F3D142_S208_L001_R1_001.fastq' and 'F3D142_S208_L001_R2_001.fastq'
        Merged 2797 reads from 3183
        Merging 'F3D143_S209_L001_R1_001.fastq' and 'F3D143_S209_L001_R2_001.fastq'
        Merged 2805 reads from 3178
        Merging 'F3D144_S210_L001_R1_001.fastq' and 'F3D144_S210_L001_R2_001.fastq'
        Merged 4078 reads from 4827
        Merging 'F3D145_S211_L001_R1_001.fastq' and 'F3D145_S211_L001_R2_001.fastq'
        Merged 6313 reads from 7377
        Merging 'F3D146_S212_L001_R1_001.fastq' and 'F3D146_S212_L001_R2_001.fastq'
        Merged 4492 reads from 5021
        Merging 'F3D147_S213_L001_R1_001.fastq' and 'F3D147_S213_L001_R2_001.fastq'
        Merged 14103 reads from 17070
        Merging 'F3D148_S214_L001_R1_001.fastq' and 'F3D148_S214_L001_R2_001.fastq'
        Merged 10787 reads from 12405
        Merging 'F3D149_S215_L001_R1_001.fastq' and 'F3D149_S215_L001_R2_001.fastq'
        Merged 11577 reads from 13083
        Merging 'F3D150_S216_L001_R1_001.fastq' and 'F3D150_S216_L001_R2_001.fastq'
        Merged 4822 reads from 5509
        Merging 'F3D2_S190_L001_R1_001.fastq' and 'F3D2_S190_L001_R2_001.fastq'
        Merged 17898 reads from 19620
        Merging 'F3D3_S191_L001_R1_001.fastq' and 'F3D3_S191_L001_R2_001.fastq'
        Merged 5987 reads from 6758
        Merging 'F3D5_S193_L001_R1_001.fastq' and 'F3D5_S193_L001_R2_001.fastq'
        Merged 3928 reads from 4448
        Merging 'F3D6_S194_L001_R1_001.fastq' and 'F3D6_S194_L001_R2_001.fastq'
        Merged 7297 reads from 7989
        Merging 'F3D7_S195_L001_R1_001.fastq' and 'F3D7_S195_L001_R2_001.fastq'
        Merged 4635 reads from 5129
        Merging 'F3D8_S196_L001_R1_001.fastq' and 'F3D8_S196_L001_R2_001.fastq'
        Merged 4880 reads from 5294
        Merging 'F3D9_S197_L001_R1_001.fastq' and 'F3D9_S197_L001_R2_001.fastq'
        Merged 6580 reads from 7070
        Merging 'Mock_S280_L001_R1_001.fastq' and 'Mock_S280_L001_R2_001.fastq'
        Merged 4561 reads from 4779

Saved 20 files with 135321 merged sequences into 'results/5fa3172ce81755a14850b9650d800c5b/Example.tar.gz'.
```

Merging details file by file

Merging paired-end reads

We are going to inspect the merged reads.

- ✓ First, rename the downloaded file to ‘example_reads_merged.tar.gz’:

```
> mv workshop.tar.gz example_reads_merged.tar.gz
```

- ✓ List the files included into ‘example_reads_merged.tar.gz’:

```
> tar -tf example_reads_merged.tar.gz
```

- ✓ Let’s check their lengths :

```
> tar xvf example_reads_merged.tar.gz -O
| awk '{ c++; if((c-2)%4==0){
++a[length()] } } END{ print "len\tseqs";
for (i in a) print i"\t"a[i]}'
```

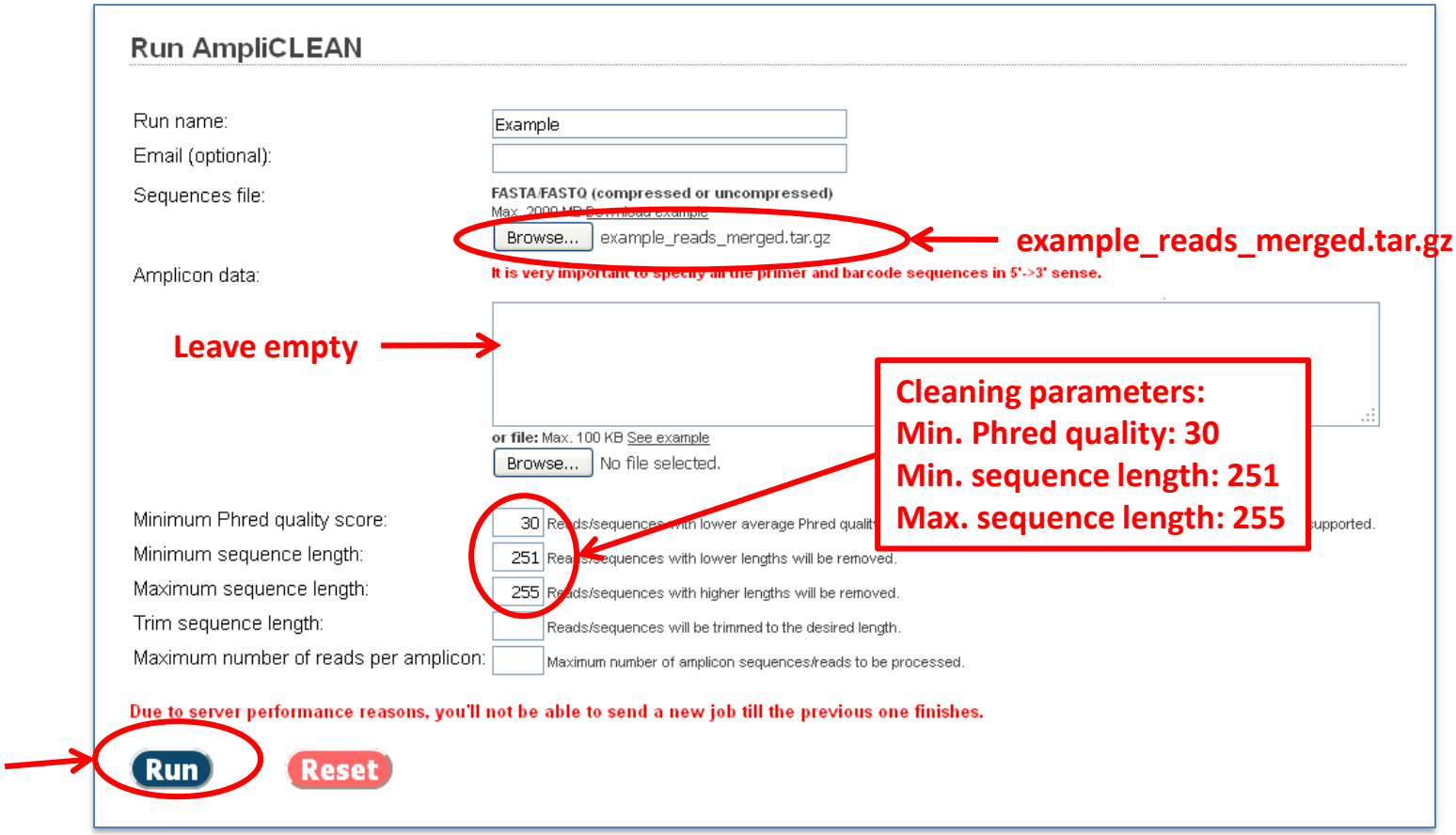
len	seqs
490	5
491	5
485	1
486	3
487	163
488	3
489	8
210	1
107	1
220	1
242	2
250	17
251	291
252	72847
253	60184
254	1576
255	209
168	1
249	3

Correctly merged reads have lengths in the range 251-255 bps.

Cleaning merged reads

- ✓ We will use the file with merged reads as input for **AmpliCLEAN**:

  http://evobiolab.biol.amu.edu.pl/amplisat/index.php?ampliclean



Cleaning merged reads

- ✓ We will download AmpliCLEAN results into the example folder.

AmpliCLEAN results for Example

Download AmpliCLEAN analysis results. ← **Example.tar.gz**

Analysis details:

```
Running 'bin/ampliCLEAN.pl -thr 20 -gz -i results/17766ba626a41e99be51df4f9b34b68b/example_reads_merged.tar.gz

Cleaning 7060 reads from file 'F3D0_S188_L001.fastq'.
        20 reads removed.

Cleaning 5404 reads from file 'F3D1_S189_L001.fastq'.
        23 reads removed.

Cleaning 5317 reads from file 'F3D141_S207_L001.fastq'.
        9 reads removed.

Cleaning 2797 reads from file 'F3D142_S208_L001.fastq'.
        3 reads removed.

Cleaning 2805 reads from file 'F3D143_S209_L001.fastq'.
        7 reads removed.

Cleaning 4078 reads from file 'F3D144_S210_L001.fastq'.
        5 reads removed.

Cleaning 6313 reads from file 'F3D145_S211_L001.fastq'.
        6 reads removed.

Cleaning 4492 reads from file 'F3D146_S212_L001.fastq'.
        10 reads removed.

Cleaning 14103 reads from file 'F3D147_S213_L001.fastq'.
        25 reads removed.

Cleaning 10787 reads from file 'F3D148_S214_L001.fastq'.
        23 reads removed.

Cleaning 11577 reads from file 'F3D149_S215_L001.fastq'.
        24 reads removed.

Cleaning 4822 reads from file 'F3D150_S216_L001.fastq'.
        15 reads removed.

Cleaning 17898 reads from file 'F3D2_S190_L001.fastq'.
        48 reads removed.

Cleaning 5987 reads from file 'F3D3_S191_L001.fastq'.
        8 reads removed.

Cleaning 3928 reads from file 'F3D5_S193_L001.fastq'.
        9 reads removed.
```

Cleaning details file by file

Cleaning merged reads

We are going to inspect the cleaned reads before final analysis.

- ✓ First, rename the downloaded file to ‘example_reads_cleaned.tar.gz’:

```
> mv workshop.tar.gz example_reads_cleaned.tar.gz
```

- ✓ List the files included into ‘example_reads_cleaned.tar.gz’:

```
> tar -tf example_reads_cleaned.tar.gz
```

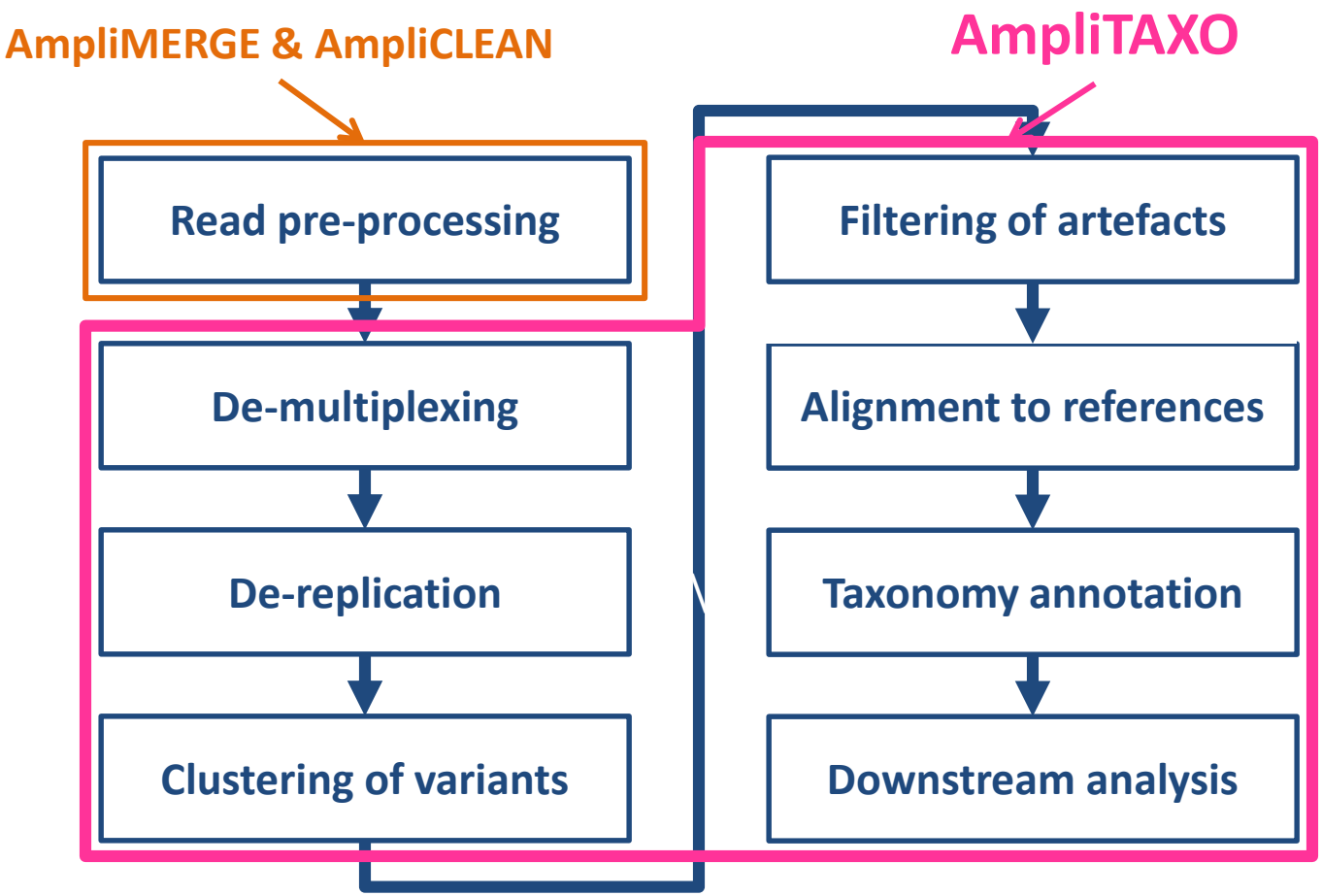
- ✓ Let’s check their lengths :

```
> tar xvf example_reads_cleaned.tar.gz -O
| awk '{ c++; if((c-2)%4==0){
++a[length()] } } END{ print "len\tseqs";
for (i in a) print i"\t"a[i]}'
```

len	seqs
251	289
252	72810
253	60143
254	1575
255	209

- Retrieve experimental data
- Checking of reads
- Read pre-processing
- AmpliTAXO analysis
- Results interpretation

Analysis pipeline

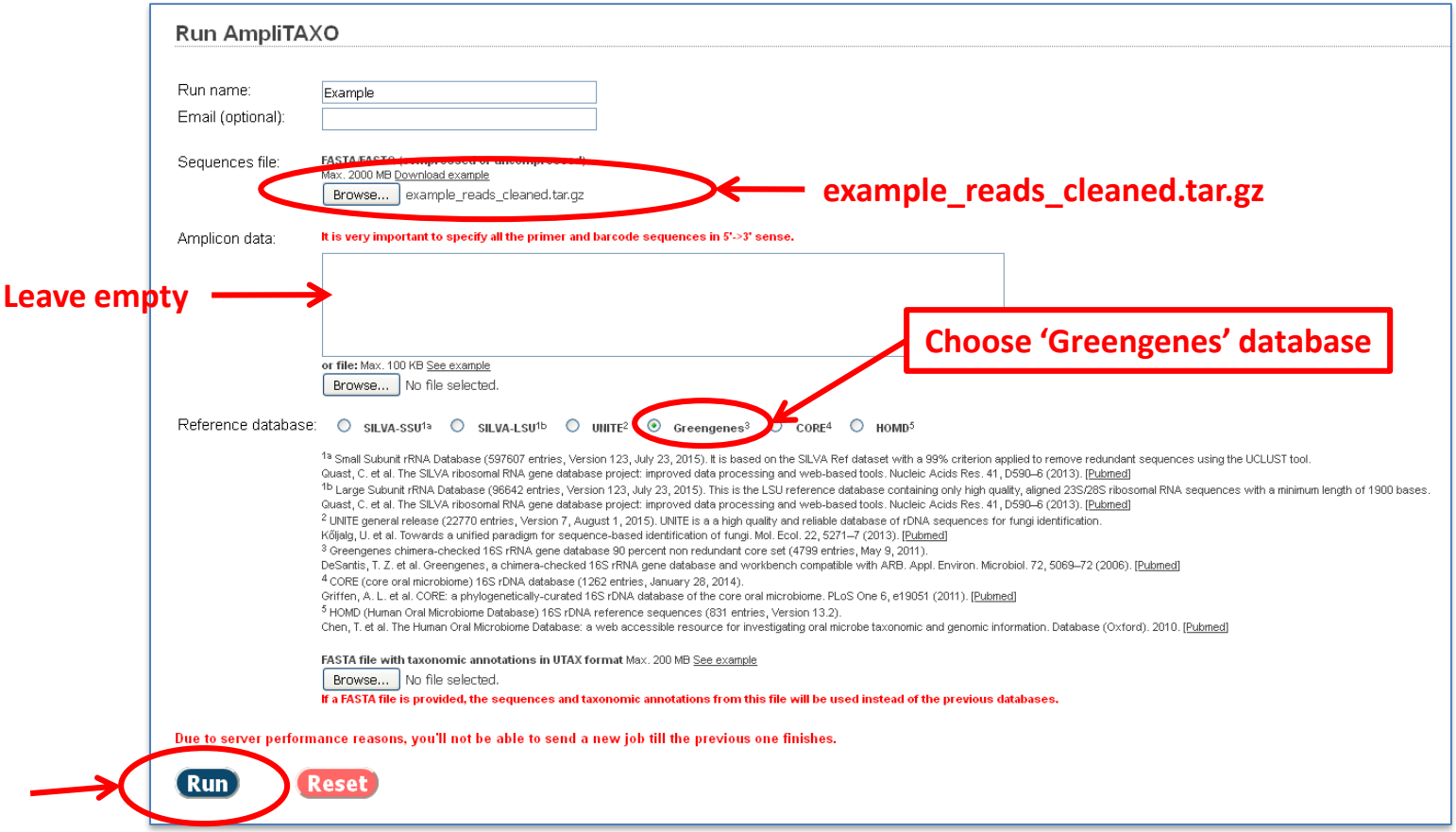


www.sixthresearcher.com
@SixthResearcher

AmpliTAXO analysis

- ✓ We will use the file with the cleaned reads as input for **AmpliTAXO**:

  http://evobiolab.biol.amu.edu.pl/amplisat/index.php?amplitaxo



AmpliTAXO analysis

✓ We will download AmpliTAXO results into the example folder.

AmpliTAXO results for Example

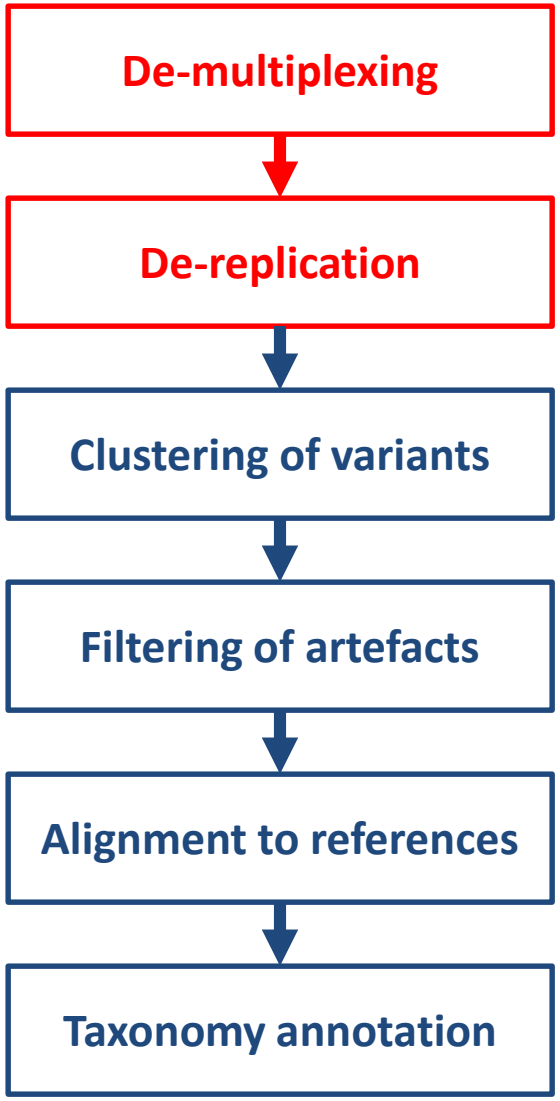
Download AmpliTAXO analysis results ← **Example.zip**

Analysis details:

```
Running 'bin/ampliTAXO.pl -thr 20 -v -i results/51dfad8161bb2841a666a06a781c20a2/16S_V4.tar.gz

Database 'Greengenes' will be used in the analysis.

Reading sequences and taxonomic information from 'Greengenes 20110509' database.
        4799 sequences read

Running AmpliSAS for sequence de-multiplexing, clustering and filtering.

Running 'bin/ampliSAS.pl -i results/51dfad8161bb2841a666a06a781c20a2/16S_V4.tar.gz

'AUTO MODE' activated.
        The program will automatically detect the sequencing technology used
        and adjust the sequencing error rates and other analysis parameters.

Reading File 'results/51dfad8161bb2841a666a06a781c20a2/16S_V4.tar.gz'.
        16S_V4-F3D0_S188_L001 de-multiplexing
        16S_V4-F3D0_S188_L001 de-multiplexed (7040 sequences, 1796 unique)
        16S_V4-F3D1_S189_L001 de-multiplexing
        16S_V4-F3D1_S189_L001 de-multiplexed (5381 sequences, 1443 unique)
        16S_V4-F3D141_S207_L001 de-multiplexing
        16S_V4-F3D141_S207_L001 de-multiplexed (5308 sequences, 1328 unique)
        16S_V4-F3D142_S208_L001 de-multiplexing
        16S_V4-F3D142_S208_L001 de-multiplexed (2794 sequences, 828 unique)
        16S_V4-F3D143_S209_L001 de-multiplexing
        16S_V4-F3D143_S209_L001 de-multiplexed (2798 sequences, 822 unique)
        16S_V4-F3D144_S210_L001 de-multiplexing
        16S_V4-F3D144_S210_L001 de-multiplexed (4073 sequences, 1124 unique)
        16S_V4-F3D145_S211_L001 de-multiplexing
        16S_V4-F3D145_S211_L001 de-multiplexed (6307 sequences, 1553 unique)
        16S_V4-F3D146_S212_L001 de-multiplexing
        16S_V4-F3D146_S212_L001 de-multiplexed (4482 sequences, 1278 unique)
        16S_V4-F3D147_S213_L001 de-multiplexing
        16S_V4-F3D147_S213_L001 de-multiplexed (14078 sequences, 3055 unique)
```

Analysis details file by file

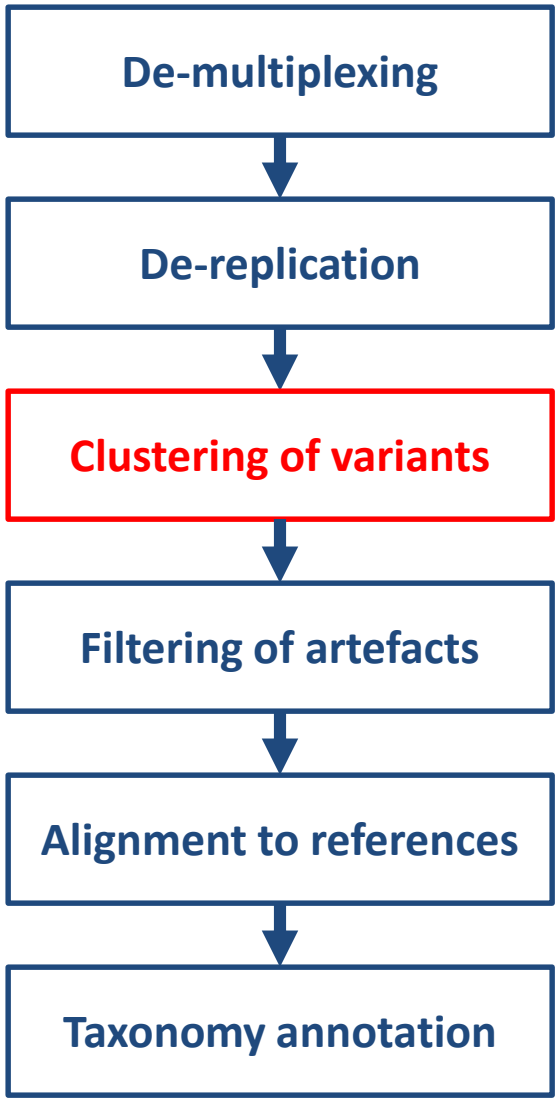
AmpliTAXO analysis




```
Reading File 'results/51dfad8161bb2841a666a06a781c20a2/16S_V4.tar.gz'.

16S_V4-F3D0_S188_L001 de-multiplexing
16S_V4-F3D0_S188_L001 de-multiplexed (7040 sequences, 1796 unique)
16S_V4-F3D1_S189_L001 de-multiplexing
16S_V4-F3D1_S189_L001 de-multiplexed (5381 sequences, 1443 unique)
16S_V4-F3D141_S207_L001 de-multiplexing
16S_V4-F3D141_S207_L001 de-multiplexed (5308 sequences, 1328 unique)
16S_V4-F3D142_S208_L001 de-multiplexing
16S_V4-F3D142_S208_L001 de-multiplexed (2794 sequences, 828 unique)
16S_V4-F3D143_S209_L001 de-multiplexing
16S_V4-F3D143_S209_L001 de-multiplexed (2798 sequences, 822 unique)
16S_V4-F3D144_S210_L001 de-multiplexing
16S_V4-F3D144_S210_L001 de-multiplexed (4073 sequences, 1124 unique)
16S_V4-F3D145_S211_L001 de-multiplexing
16S_V4-F3D145_S211_L001 de-multiplexed (6307 sequences, 1553 unique)
16S_V4-F3D146_S212_L001 de-multiplexing
16S_V4-F3D146_S212_L001 de-multiplexed (4482 sequences, 1278 unique)
16S_V4-F3D147_S213_L001 de-multiplexing
16S_V4-F3D147_S213_L001 de-multiplexed (14078 sequences, 3055 unique)
16S_V4-F3D148_S214_L001 de-multiplexing
16S_V4-F3D148_S214_L001 de-multiplexed (10764 sequences, 2398 unique)
16S_V4-F3D149_S215_L001 de-multiplexing
16S_V4-F3D149_S215_L001 de-multiplexed (11553 sequences, 2630 unique)
16S_V4-F3D150_S216_L001 de-multiplexing
16S_V4-F3D150_S216_L001 de-multiplexed (4807 sequences, 1419 unique)
16S_V4-F3D2_S190_L001 de-multiplexing
16S_V4-F3D2_S190_L001 de-multiplexed (17850 sequences, 3366 unique)
16S_V4-F3D3_S191_L001 de-multiplexing
16S_V4-F3D3_S191_L001 de-multiplexed (5979 sequences, 1294 unique)
16S_V4-F3D5_S193_L001 de-multiplexing
16S_V4-F3D5_S193_L001 de-multiplexed (3919 sequences, 1080 unique)
16S_V4-F3D6_S194_L001 de-multiplexing
16S_V4-F3D6_S194_L001 de-multiplexed (7280 sequences, 1655 unique)
16S_V4-F3D7_S195_L001 de-multiplexing
16S_V4-F3D7_S195_L001 de-multiplexed (4631 sequences, 1045 unique)
16S_V4-F3D8_S196_L001 de-multiplexing
16S_V4-F3D8_S196_L001 de-multiplexed (4868 sequences, 1241 unique)
16S_V4-F3D9_S197_L001 de-multiplexing
16S_V4-F3D9_S197_L001 de-multiplexed (6563 sequences, 1589 unique)
16S_V4-Mock_S280_L001 de-multiplexing
16S_V4-Mock_S280_L001 de-multiplexed (4551 sequences, 767 unique)
```

www.sixthresearcher.com
@SixthResearcher

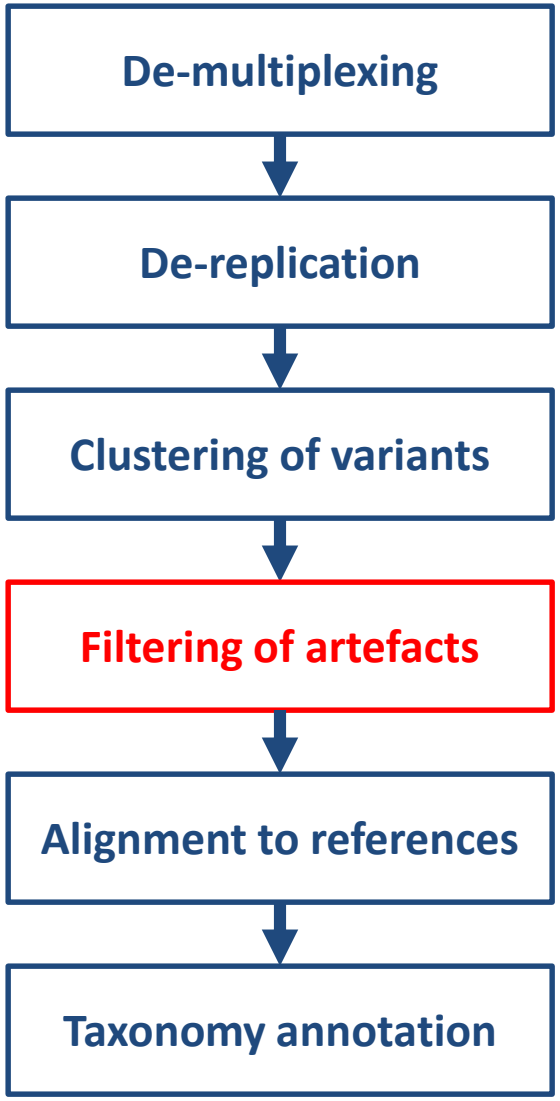
AmpliTAXO analysis




```
Checking data and setting automatically clustering parameters.
Technology: Illumina
Error rates: 0.87% substitutions, 0.001% indels
Minimum amplicon depth: 100

Checking data and setting marker lengths.
Marker '16S_V4' length: 252 (auto)

Clustering amplicon sequences with the following parameters ('threshold' 'marker'
'values'):
*substitution_threshold       all            0.87
*indel_threshold              all            0.001
*min_dominant_frequency_threshold            16S_V4         36.79

*WARNING: These parameters have been automatically adjusted and may require manual
verification.

16S_V4-F3D0_S188_L001 clustering
16S_V4-F3D1_S189_L001 clustering
16S_V4-F3D141_S207_L001 clustering
16S_V4-F3D142_S208_L001 clustering
16S_V4-F3D143_S209_L001 clustering
16S_V4-F3D144_S210_L001 clustering
16S_V4-F3D2_S190_L001 clustering
16S_V4-F3D3_S191_L001 clustering
16S_V4-F3D5_S193_L001 clustering
16S_V4-F3D6_S194_L001 clustering
16S_V4-F3D7_S195_L001 clustering
16S_V4-F3D8_S196_L001 clustering
16S_V4-F3D9_S197_L001 clustering
16S_V4-Mock_S280_L001 clustering
16S_V4-Mock_S280_L001 clustered (4381 sequences, 29 unique)
16S_V4-F3D142_S208_L001 clustered (2460 sequences, 112 unique)
16S_V4-F3D143_S209_L001 clustered (2473 sequences, 132 unique)
16S_V4-F3D7_S195_L001 clustered (4311 sequences, 145 unique)
16S_V4-F3D144_S210_L001 clustered (3654 sequences, 151 unique)
16S_V4-F3D5_S193_L001 clustered (3554 sequences, 161 unique)
16S_V4-F3D3_S191_L001 clustered (5583 sequences, 155 unique)
16S_V4-F3D8_S196_L001 clustered (4482 sequences, 180 unique)
16S_V4-F3D141_S207_L001 clustered (4880 sequences, 175 unique)
16S_V4-F3D1_S189_L001 clustered (4918 sequences, 165 unique)
16S_V4-F3D6_S194_L001 clustered (6764 sequences, 186 unique)
16S_V4-F3D9_S197_L001 clustered (6079 sequences, 201 unique)
16S_V4-F3D0_S188_L001 clustered (6498 sequences, 211 unique)
16S_V4-F3D2_S190_L001 clustered (16966 sequences, 291 unique)
```

www.sixthresearcher.com
@SixthResearcher

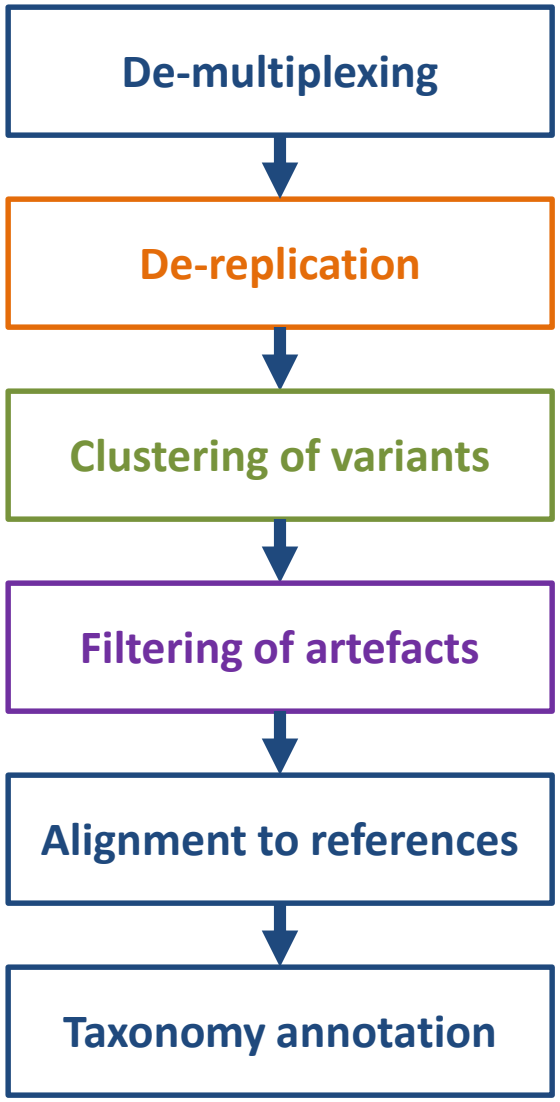
AmpliTAXO analysis




```
Filtering sequences with the following criteria ('filter' 'marker' 'values'):
                *min_amplicon_seq_frequency    all            0
                *min_amplicon_depth            all            100
                *min_chimera_length            all            10

                *WARNING: These parameters have been automatically adjusted and may
require manual verification.

16S_V4-F3D0_S188_L001 filtering
16S_V4-F3D1_S189_L001 filtering
16S_V4-F3D141_S207_L001 filtering
16S_V4-F3D142_S208_L001 filtering
16S_V4-F3D143_S209_L001 filtering
16S_V4-F3D144_S210_L001 filtering
16S_V4-F3D145_S211_L001 filtering
16S_V4-F3D146_S212_L001 filtering
16S_V4-F3D147_S213_L001 filtering
16S_V4-F3D2_S190_L001 filtering
16S_V4-F3D3_S191_L001 filtering
16S_V4-F3D5_S193_L001 filtering
16S_V4-F3D6_S194_L001 filtering
16S_V4-F3D7_S195_L001 filtering
16S_V4-F3D8_S196_L001 filtering
16S_V4-F3D9_S197_L001 filtering
16S_V4-Mock_S280_L001 filtering
16S_V4-Mock_S280_L001 filtered (4362 sequences, 20 unique)
16S_V4-F3D142_S208_L001 filtered (2357 sequences, 82 unique)
16S_V4-F3D143_S209_L001 filtered (2351 sequences, 91 unique)
16S_V4-F3D144_S210_L001 filtered (3385 sequences, 93 unique)
16S_V4-F3D3_S191_L001 filtered (5253 sequences, 102 unique)
16S_V4-F3D7_S195_L001 filtered (4119 sequences, 106 unique)
16S_V4-F3D141_S207_L001 filtered (4666 sequences, 119 unique)
16S_V4-F3D145_S211_L001 filtered (5399 sequences, 122 unique)
16S_V4-F3D1_S189_L001 filtered (4862 sequences, 146 unique)
16S_V4-F3D5_S193_L001 filtered (3474 sequences, 133 unique)
16S_V4-F3D146_S212_L001 filtered (3791 sequences, 143 unique)
16S_V4-F3D6_S194_L001 filtered (6587 sequences, 152 unique)
16S_V4-F3D0_S188_L001 filtered (6367 sequences, 170 unique)
16S_V4-F3D8_S196_L001 filtered (4421 sequences, 161 unique)
16S_V4-F3D9_S197_L001 filtered (5980 sequences, 175 unique)
16S_V4-F3D2_S190_L001 filtered (16342 sequences, 188 unique)
16S_V4-F3D147_S213_L001 filtered (11924 sequences, 183 unique)
```

www.sixthresearcher.com
@SixthResearcher

AmpliTAXO analysis

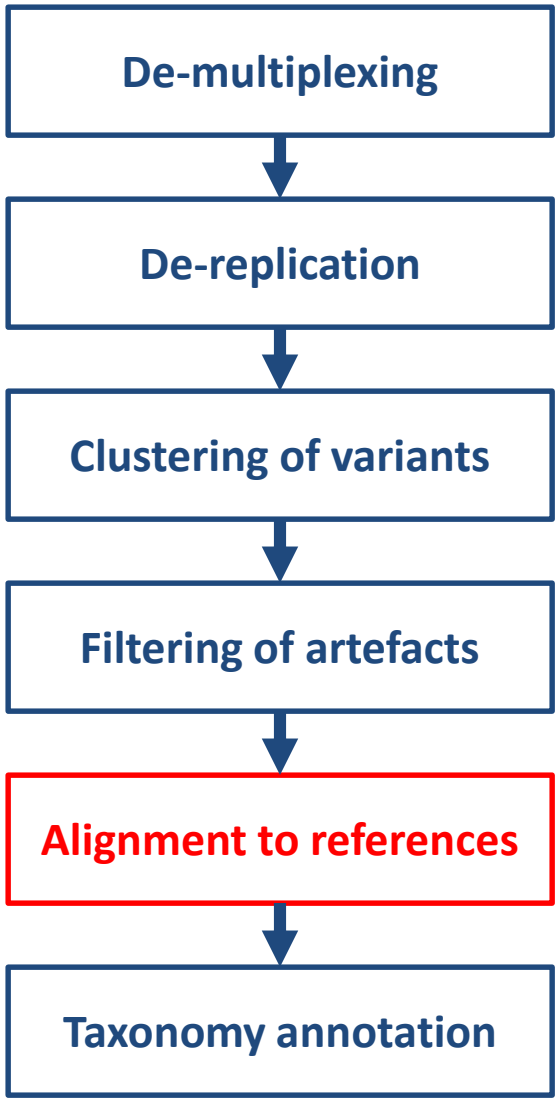


Sequences per amplicon:

Amplicon	Total	Unique	Total	Unique	Total	Unique
16S_V4-F3D0_S188_L001	7040	1796	6498	211	6367	170
16S_V4-F3D1_S189_L001	5381	1443	4918	165	4862	146
16S_V4-F3D141_S207_L001	5308	1328	4880	175	4666	119
16S_V4-F3D142_S208_L001	2794	828	2460	112	2357	82
16S_V4-F3D143_S209_L001	2798	822	2473	132	2351	91
16S_V4-F3D144_S210_L001	4073	1124	3654	151	3385	93
16S_V4-F3D145_S211_L001	6307	1553	5795	196	5399	122
16S_V4-F3D146_S212_L001	4482	1278	4001	196	3791	143
16S_V4-F3D147_S213_L001	14078	3055	13184	338	11924	183
16S_V4-F3D148_S214_L001	10764	2398	10096	283	9416	170
16S_V4-F3D149_S215_L001	11553	2630	10761	275	10144	175
16S_V4-F3D150_S216_L001	4807	1419	4295	197	4087	139
16S_V4-F3D2_S190_L001	17850	3366	16966	291	16342	188
16S_V4-F3D3_S191_L001	5979	1294	5583	155	5253	102
16S_V4-F3D5_S193_L001	3919	1080	3554	161	3474	133
16S_V4-F3D6_S194_L001	7280	1655	6764	186	6587	152
16S_V4-F3D7_S195_L001	4631	1045	4311	145	4119	106
16S_V4-F3D8_S196_L001	4868	1241	4482	180	4421	161
16S_V4-F3D9_S197_L001	6563	1589	6079	201	5980	175
16S_V4-Mock_S280_L001	4551	767	4381	29	4362	20

www.sixthresearcher.com
@SixthResearcher

AmpliTAXO analysis



Assigning OTUs by matching reference sequences from 'Greengenes'.

Sequences assigned to OTUs:

Sample	Total depth	OTUs depth	Unique OTUs	Taxonomies
F3D0_S188_L001	7040	6367	170	30
F3D1_S189_L001	5381	4862	146	24
F3D2_S190_L001	17850	16342	188	31
F3D3_S191_L001	5979	5253	102	22
F3D5_S193_L001	3919	3474	133	23
F3D6_S194_L001	7280	6587	152	25
F3D7_S195_L001	4631	4119	106	23
F3D8_S196_L001	4868	4421	161	28
F3D9_S197_L001	6563	5980	175	23
F3D141_S207_L001	5308	4666	119	23
F3D142_S208_L001	2794	2357	82	18
F3D143_S209_L001	2798	2351	91	22
F3D144_S210_L001	4073	3385	93	21
F3D145_S211_L001	6307	5399	122	28
F3D146_S212_L001	4482	3791	143	21
F3D147_S213_L001	14078	11924	183	30
F3D148_S214_L001	10764	9416	170	26
F3D149_S215_L001	11553	10144	175	24
F3D150_S216_L001	4807	4087	139	24
Mock_S280_L001	4551	4362	20	18

www.sixthresearcher.com
@SixthResearcher

AmpliTAXO analysis

We are going to inspect the final results from AmpliTAXO analysis.

- ✓ First, rename the downloaded file to 'amplitaxo_results.zip':

```
> mv workshop.zip amplitaxo_results.zip
```

- ✓ Create a folder to save the results:

```
> mkdir amplitaxo_results
```

- ✓ Enter into the folder :

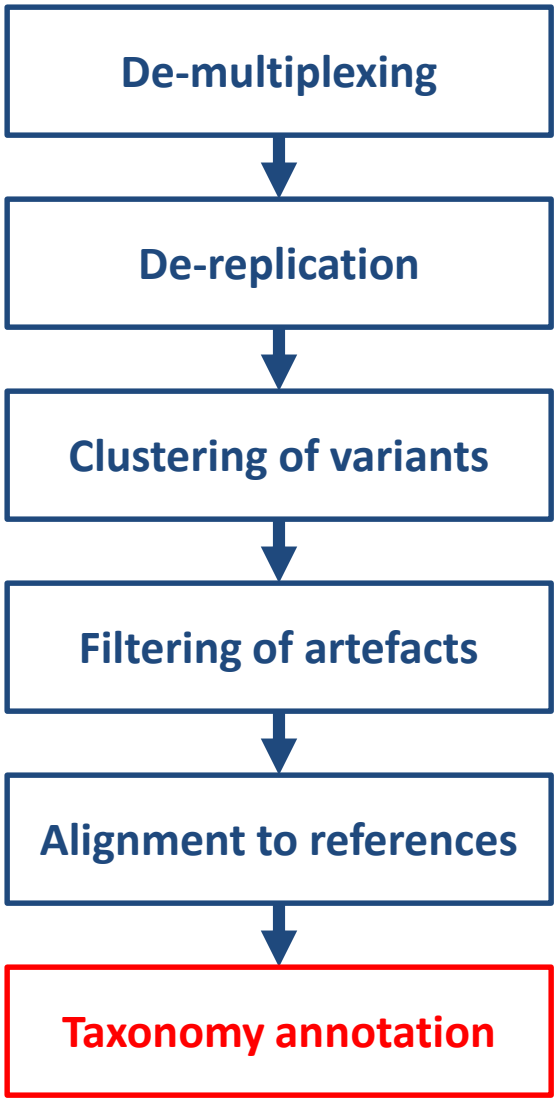
```
> cd amplitaxo_results
```

- ✓ Extract the results:

```
> unzip ../amplitaxo_results.zip
```

- ✓ Open with the 'results.xlsx' Excel file in Windows:
  `C:\workshop\MiSeq_SOP\amplitaxo_results\results.xlsx`

AmpliTAXO results



OTU representative sequences

OTU Taxonomy assignments

OTU sequence depths per sample

OTUs

RICHNESS

EVENNESS

OTU_ID	SEQUENCE	TAXONOMY	Mock_S280	F3D9_S197
16S-00005	TACGGAGGATGC	tax=k:Bacteria,p:Bacteroidetes,c:Bacteroidia,o:Bacteroidales,f:Bacteroidaceae,g:Bacteroides		650
16S-00028	TACGTAGGTGGC	tax=k:Bacteria,p:Firmicutes,c:Bacilli,o:Bacillales,f:Staphylococcaceae,g:Staphylococcus	588	
16S-00001	TACGGAGGATGC	tax=k:Bacteria,p:Bacteroidetes,c:Bacteroidia,o:Bacteroidales		465
16S-00007	TACGGAGGATTC	tax=k:Bacteria,p:Bacteroidetes,c:Bacteroidia,o:Bacteroidales,f:Rikenellaceae,g:Alistipes,s:putredi		449
16S-00003	TACGGAGGATGC	tax=k:Bacteria,p:Bacteroidetes,c:Bacteroidia,o:Bacteroidales		447
16S-00041	TACAGAGGGTGC	tax=k:Bacteria,p:Proteobacteria,c:Gammaproteobacteria,o:Pseudomonadales,f:Moraxellaceae,g:A	395	
16S-00002	TACGGAGGATGC	tax=k:Bacteria,p:Bacteroidetes,c:Bacteroidia,o:Bacteroidales		394
16S-00045	TACGTAGGTGGC	tax=k:Bacteria,p:Firmicutes,c:Bacilli,o:Bacillales	360	
16S-00050	TACGGAGGGTGC	tax=k:Bacteria,p:Proteobacteria,c:Epsilonproteobacteria,o:Campylobacterales,f:Helicobacteraceae	347	
16S-00053	TACGTAGGGCGC	tax=k:Bacteria,p:Actinobacteria,c:Actinobacteria (class),o:Actinomycetales,f:Actinomycetaceae,g:A	326	
16S-00064	TACGTAGGTGGC	tax=k:Bacteria,p:Firmicutes,c:Clostridia,o: s,f:Clostridiaceae,g:Clostridium	2	
16S-00059	TACGTAGGGTGC	tax=k:Bacteria,p:Proteobacteria,c:Betapr ,o:Neisseriales,f:Neisseriaceae,g:Neisseria	2	
16S-00008	TACGGAGGATGC	tax=k:Bacteria,p:Bacteroidetes,c:Bactero oidales		278
16S-00085	TACGGAGGATGC	tax=k:Bacteria,p:Bacteroidetes,c:Bactero oidales,f:Bacteroidaceae,g:Bacteroides,s:dor	2	
16S-00073	TACGTAGGTCCC	tax=k:Bacteria,p:Firmicutes,c:Bacilli,o:La f:Streptococcaceae,g:Streptococcus	2	
16S-00009	TACGGAGGATGC	tax=k:Bacteria,p:Bacteroidetes,c:Bactero oidales		213
16S-00004	TACGGAGGATGC	tax=k:Bacteria,p:Bacteroidetes,c:Bactero oidales		209
16S-00087	TACGTAGGTCCC	tax=k:Bacteria,p:Firmicutes,c:Bacilli,o:La f:Streptococcaceae,g:Streptococcus	2	
16S-00078	TACGGAGGGTGC	tax=k:Bacteria,p:Proteobacteria,c:Gamm eria,o:Enterobacteriales,f:Enterobacteriaceae	1	
16S-00088	TACGTAGGTGGC	tax=k:Bacteria,p:Firmicutes,c:Bacilli,o:La f:Enterococcaceae,g:Enterococcus	1	
16S-00012	TACGTAGGTGGC	tax=k:Bacteria,p:Firmicutes,c:Bacilli,o:La f:Lactobacillaceae,g:Lactobacillus		174
16S-00010	TACGTAGGTGGC	tax=k:Bacteria,p:Firmicutes,c:Bacilli,o:La f:Lactobacillaceae,g:Lactobacillus	1	34
16S-00101	TACGTAGGTGGC	tax=k:Bacteria,p:Firmicutes,c:Bacilli,o:La f:Carnobacteriaceae,g:Carnobacterium,s:virid	1	
16S-00110	TACGAAGGGTGC	tax=k:Bacteria,p:Proteobacteria,c:Gamm eria,o:Pseudomonadales,f:Pseudomonadace	1	
16S-00136	TACGGAGGGTGC	tax=k:Bacteria,p:Thermi,c:Deinococci,o:D ,f:Deinococcaceae,g:Deinococcus,s:radiodur	1	
16S-00015	TACGTAGGGGGC	tax=k:Bacteria,p:Firmicutes,c:Clostridia,o: s,f:Lachnospiraceae		98
16S-00057	TACGTAGGGGGC	tax=k:Bacteria,p:Firmicutes,c:Clostridia,o: s,f:Lachnospiraceae		98
16S-00039	TACGTAGGGGGC	tax=k:Bacteria,p:Firmicutes,c:Clostridia,o: s,f:Lachnospiraceae		97
16S-00144	TACGGAGGATGC	tax=k:Bacteria,p:Bacteroidetes,c:Bactero oidales,f:Porphyromonadaceae,g:Porphyrom	9	
16S-00032	TACGTAGGGGGC	tax=k:Bacteria,p:Firmicutes,c:Clostridia,o: s,f:Lachnospiraceae		89
16S-00034	TACGTAGGGGGC	tax=k:Bacteria,p:Firmicutes,c:Clostridia,o: s,f:Lachnospiraceae,g:Coprococcus		89
16S-00029	TACGTAGGTGGC	tax=k:Bacteria,p:Firmicutes,c:Clostridia,o:Clostridiales,f:Ruminococcaceae		64
16S-00026	TACGTAGGGGGC	tax=k:Bacteria,p:Firmicutes,c:Clostridia,o:Clostridiales,f:Lachnospiraceae		57
16S-00074	TACGTAGGTGGC	tax=k:Bacteria,p:Firmicutes,c:Clostridia,o:Clostridiales,f:Ruminococcaceae		56
16S-00063	TACGTATGGAGC	tax=k:Bacteria,p:Firmicutes,c:Clostridia,o:Clostridiales,f:Lachnospiraceae		54
16S-00033	TACGTAGGGGGC	tax=k:Bacteria,p:Firmicutes,c:Clostridia,o:Clostridiales		52
16S-00044	TACGTAGGGGGC	tax=k:Bacteria,p:Firmicutes,c:Clostridia,o:Clostridiales | tax=k:Bacteria,p:Firmicutes,c:Clostridia,o:		52
16S-00248	TACGGAGGGGGC	tax=k:Bacteria,p:Proteobacteria,c:Alphaproteobacteria,o:Rhodobacterales,f:Rhodobacteraceae,g:T	50	
16S-00023	TACGTAGGGGGC	tax=k:Bacteria,p:Firmicutes,c:Clostridia,o:Clostridiales,f:Lachnospiraceae		48
16S-00024	TACGTAGGTGGC	tax=k:Bacteria,p:Firmicutes,c:Clostridia,o:Clostridiales,f:Ruminococcaceae | tax=k:Bacteria,p:Firmi		47
16S-00077	TACGTAGGGGGC	tax=k:Bacteria,p:Firmicutes,c:Clostridia,o:Clostridiales,f:Lachnospiraceae		46
16S-00035	TACGTAGGGGGC	tax=k:Bacteria,p:Firmicutes,c:Clostridia,o:Clostridiales,f:Lachnospiraceae		43
16S-00018	TACGTAGGGGGC	tax=k:Bacteria,p:Firmicutes,c:Clostridia,o:Clostridiales,f:Lachnospiraceae		42

www.sixthresearcher.com

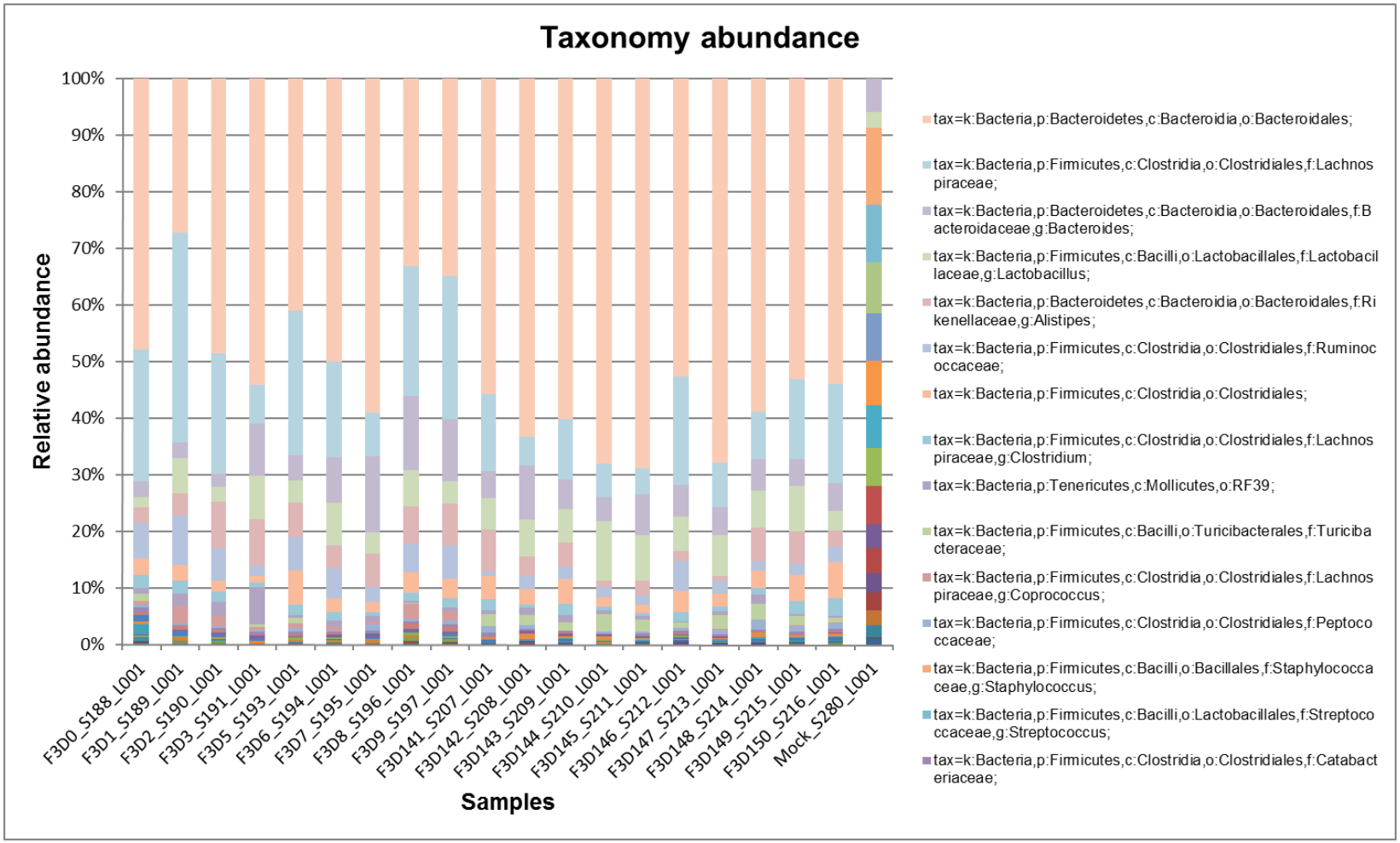
@SixthResearcher

- Retrieve experimental data
- Checking of reads
- Read pre-processing
- AmpliTAXO analysis
- Results interpretation

www.sixthresearcher.com
@SixthResearcher

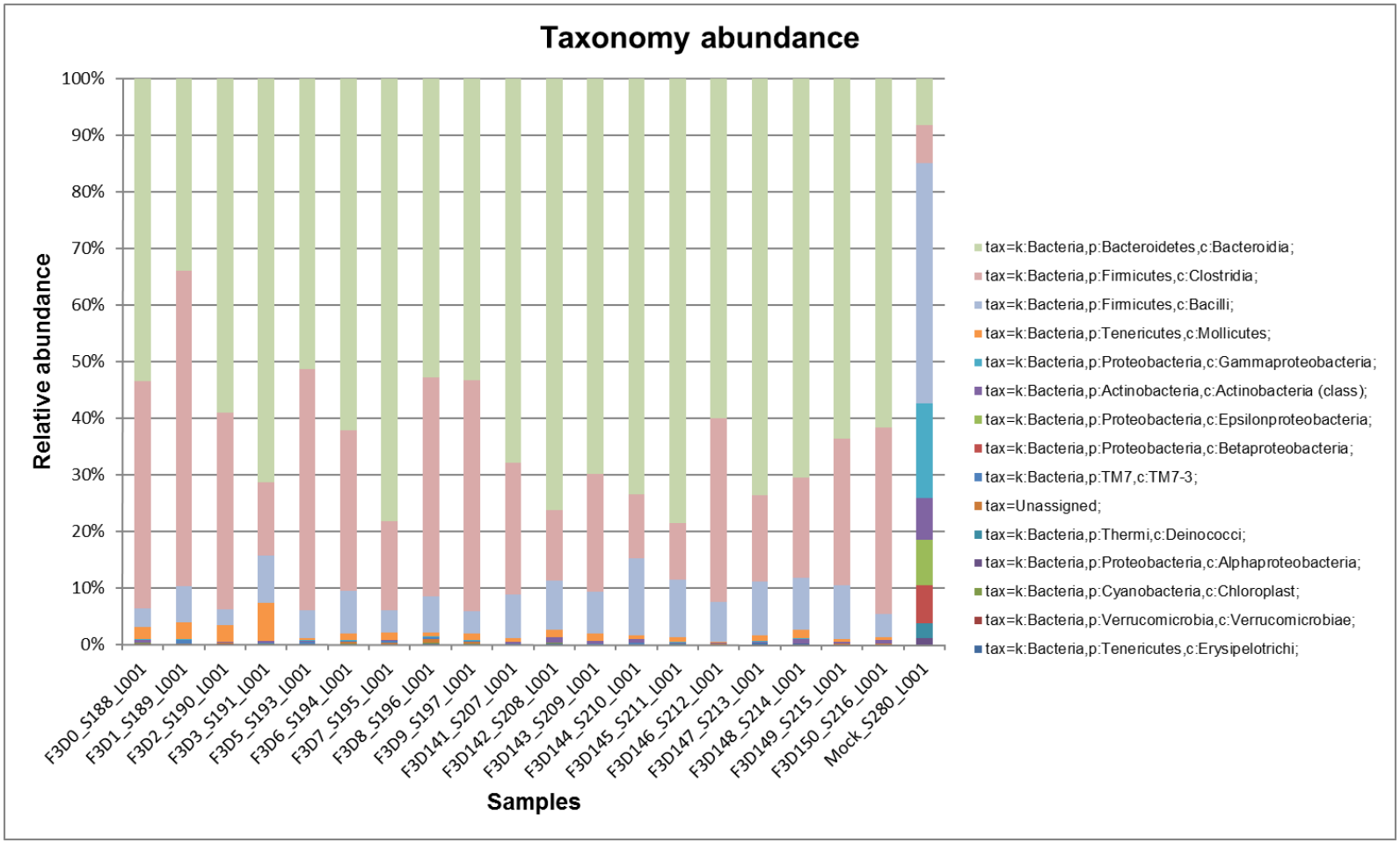
AmpliTAXO results

Taxonomy abundance by genus:



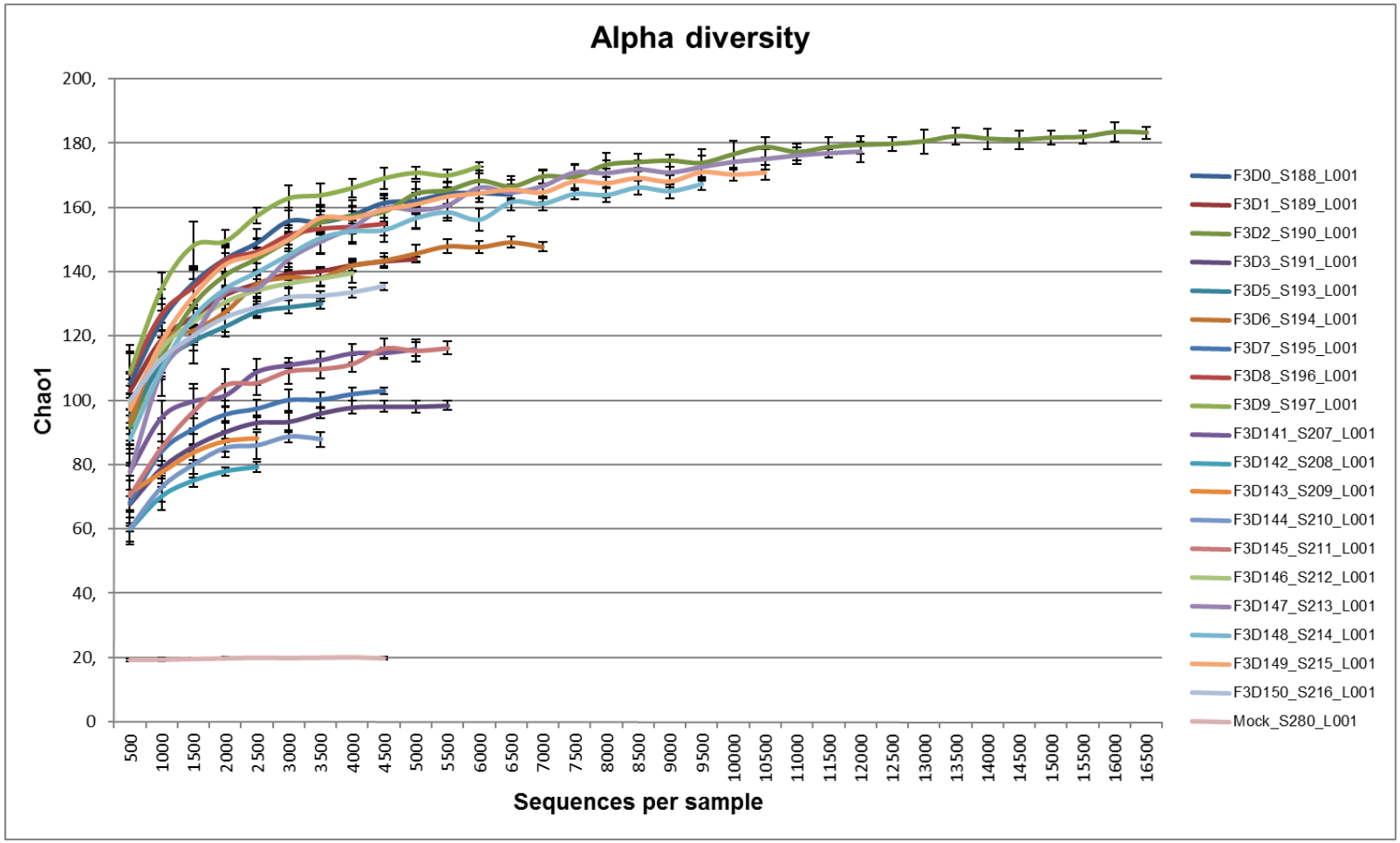
AmpliTAXO results

Taxonomy abundance by class:



AmpliTAXO results

Alpha diversity rarefraction plot:



AmpliTAXO vs. MOTHUR

- MOTHUR tutorial:

  http://www.mothur.org/wiki/MiSeq_SOP

"**You'll see that for 4061 sequences, we'd have 35 OTUs from the Mock community.** This number of course includes some stealthy chimeras that escaped our detection methods. If we used 3000 sequences, we would have about 31 OTUs. **In a perfect world with no chimeras and no sequencing errors, we'd have 20 OTUs.** This is not a perfect world. But this is pretty darn good! "

AmpliTAXO vs. MOTHUR

- MOTHUR tutorial:

  http://www.mothur.org/wiki/MiSeq_SOP

- AmpliTAXO results:

		DEPTH_AMPLICON	4551
		DEPTH_OTUS	4262
		COUNT_OTUS	20
OTU_ID	SEQUENCE	TAXONOMY	Mo[illegible]280
16S-00028	TACGTAGGTGGC	tax=k:Bacteria,p:Firmicutes,c:Bacilli,o:Bacillales,f:Staphylococcaceae,g:Staphylococcus	588
16S-00041	TACAGAGGGTGC	tax=k:Bacteria,p:Proteobacteria,c:Gammaproteobacteria,o:Pseudomonadales,f:Moraxellaceae,g:Acinet	395
16S-00045	TACGTAGGTGGC	tax=k:Bacteria,p:Firmicutes,c:Bacilli,o:Bacillales	360
16S-00050	TACGGAGGGTGC	tax=k:Bacteria,p:Proteobacteria,c:Epsilonproteobacteria,o:Campylobacterales,f:Helicobacteraceae,g:He	347
16S-00053	TACGTAGGGCGC	tax=k:Bacteria,p:Actinobacteria,c:Actinobacteria (class),o:Actinomycetales,f:Actinomycetaceae,g:Actinc	326
16S-00064	TACGTAGGTGGC	tax=k:Bacteria,p:Firmicutes,c:Clostridia,o:Clostridiales,f:Clostridiaceae,g:Clostridium	295
16S-00059	TACGTAGGGTGC	tax=k:Bacteria,p:Proteobacteria,c:Betaproteobacteria,o:Neisseriales,f:Neisseriaceae,g:Neisseria	294
16S-00085	TACGGAGGATCC	tax=k:Bacteria,p:Bacteroidetes,c:Bacteroidia,o:Bacteroidales,f:Bacteroidaceae,g:Bacteroides,s:dorei	257
16S-00073	TACGTAGGTCCC	tax=k:Bacteria,p:Firmicutes,c:Bacilli,o:Lactobacillales,f:Streptococcaceae,g:Streptococcus	217
16S-00087	TACGTAGGTCCC	tax=k:Bacteria,p:Firmicutes,c:Bacilli,o:Lactobacillales,f:Streptococcaceae,g:Streptococcus	208
16S-00078	TACGGAGGGTGC	tax=k:Bacteria,p:Proteobacteria,c:Gammaproteobacteria,o:Enterobacteriales,f:Enterobacteriaceae	191
16S-00088	TACGTAGGTGGC	tax=k:Bacteria,p:Firmicutes,c:Bacilli,o:Lactobacillales,f:Enterococcaceae,g:Enterococcus	188
16S-00101	TACGTAGGTGGC	tax=k:Bacteria,p:Firmicutes,c:Bacilli,o:Lactobacillales,f:Carnobacteriaceae,g:Carnobacterium,s:viridans	151
16S-00110	TACGAAGGGTGC	tax=k:Bacteria,p:Proteobacteria,c:Gammaproteobacteria,o:Pseudomonadales,f:Pseudomonadaceae,g:P	138
16S-00010	TACGTAGGTGGC	tax=k:Bacteria,p:Firmicutes,c:Bacilli,o:Lactobacillales,f:Lactobacillaceae,g:Lactobacillus	124
16S-00136	TACGGAGGGTGC	tax=k:Bacteria,p:Thermi,c:Deinococci,o:Deinococcales,f:Deinococcaceae,g:Deinococcus,s:radiodurans	112
16S-00144	TACGGAGGATGC	tax=k:Bacteria,p:Bacteroidetes,c:Bacteroidia,o:Bacteroidales,f:Porphyromonadaceae,g:Porphyromonas,	97
16S-00248	TACGGAGGGGGC	tax=k:Bacteria,p:Proteobacteria,c:Alphaproteobacteria,o:Rhodobacterales,f:Rhodobacteraceae,g:Thiocl	50
16S-00367	TACGTAGGTCCC	tax=k:Bacteria,p:Firmicutes,c:Bacilli,o:Lactobacillales,f:Streptococcaceae,g:Streptococcus,s:pseudopneu	22
16S-00212	TACAGAGGGTGC	tax=k:Bacteria,p:Proteobacteria,c:Gammaproteobacteria,o:Pseudomonadales,f:Pseudomonadaceae,g:P	2

Interesting materials

- Materials from Strategies and Techniques for Analyzing Microbial Population Structure Course
  https://stamps.mbl.edu/index.php/Sue_Huse

- SSU Metagenomics (UPARSE)
  http://drive5.com/ssu.html

- MOTHUR manual
  http://www.mothur.org/wiki/Mothur_manual

- QIIME overview tutorial
  http://www.wernerlab.org/teaching/qiime/overview

www.sixthresearcher.com
@SixthResearcher