

# Microbiome Analysis

Israel Barrantes

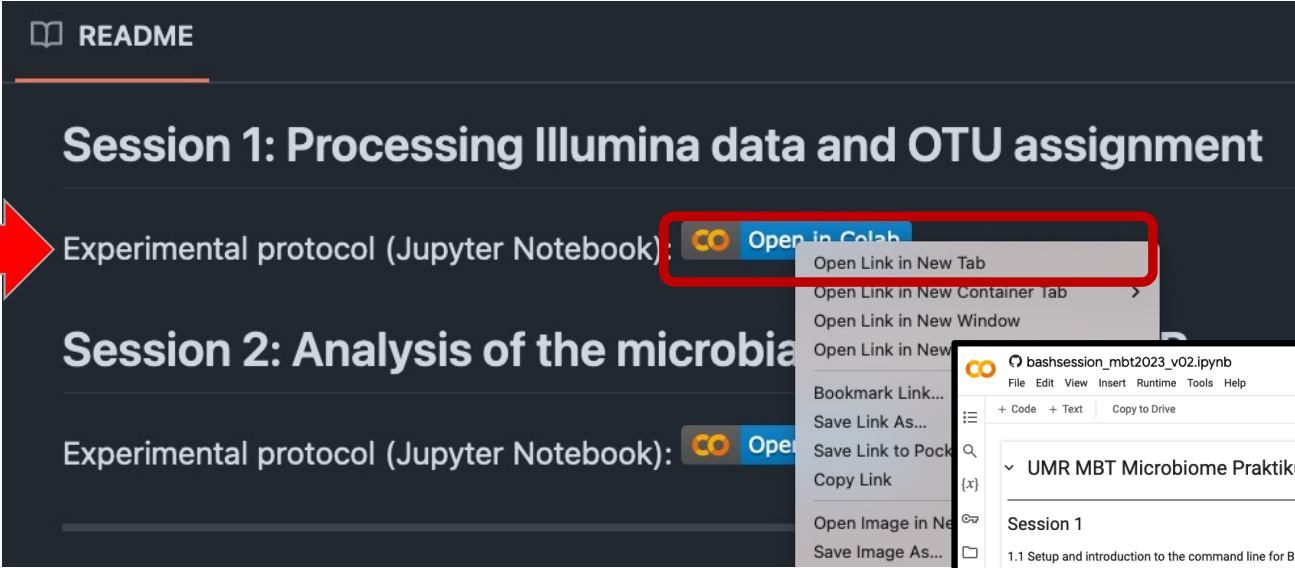
[israel.barrantes@uni-rostock.de](mailto:israel.barrantes@uni-rostock.de)

Inst. Biostatistics and Informatics in Medicine and Ageing Research  
Rostock University Medical Center

2024-10-23

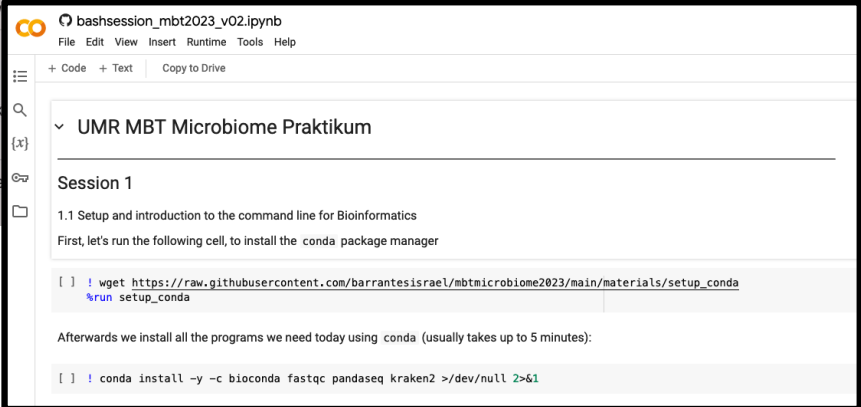
# Data and Links for Hands-on Practice

<https://github.com/barrantesisrael/mbtmicrobiome2023>



The screenshot shows a GitHub README page with a dark theme. It features two main sections: "Session 1: Processing Illumina data and OTU assignment" and "Session 2: Analysis of the microbiome". Each section has a link to an "Experimental protocol (Jupyter Notebook)" with a Colab icon and the text "Open in Colab". A red arrow points to the first "Open in Colab" link. A context menu is open over this link, showing options like "Open Link in New Tab", "Open Link in New Container Tab", "Open Link in New Window", "Open Link in New Window", "Bookmark Link...", "Save Link As...", "Save Link to Pocket", "Copy Link", "Open Image in New Window", and "Save Image As...".

right click "Open in Colab" to open on a new tab



The screenshot shows a Jupyter Notebook titled "bashsession\_mbt2023\_v02.ipynb". The notebook content includes a section titled "UMR MBT Microbiome Praktikum" and a subsection "Session 1". Under "Session 1", there is a sub-section "1.1 Setup and introduction to the command line for Bioinformatics". The text says: "First, let's run the following cell, to install the conda package manager". Below this, there is a code cell with the following commands:

```
[ ] ! wget https://raw.githubusercontent.com/barrantesisrael/mbtmicrobiome2023/main/materials/setup_conda
    %run setup_conda
```

Afterwards we install all the programs we need today using conda (usually takes up to 5 minutes):

```
[ ] ! conda install -y -c bioconda fastqc pandaseq kraken2 >/dev/null 2>&1
```

# Why the command line for Bioinformatics?

- **Reproducibility:** Command-line tools and programming enable the creation of reproducible workflows. By writing scripts or workflows, researchers can document and automate their analyses, ensuring that others can replicate their results.
- **Pipelines:** programs talking to each other (pipes)
- **Redirection:** programs write and read to files
- **Text Streams:** Allow us to both couple programs together and process data without storing huge amounts of data in our computers' memory
- **Modularity**
  - Modular workflows allow us to experiment with alternate methods and approaches, since independent components can be easily swapped out
  - In a modular workflow each component is independent, which makes it easier to inspect intermediate results for inconsistencies and isolate problematic steps
  - Modular components allow us to choose tools and languages that are appropriate for specific tasks
  - Modular programs are reusable and applicable to many types of data

Buffalo, V. (2015) *Bioinformatics Data Skills*. O'Reilly Media, Sebastopol CA.

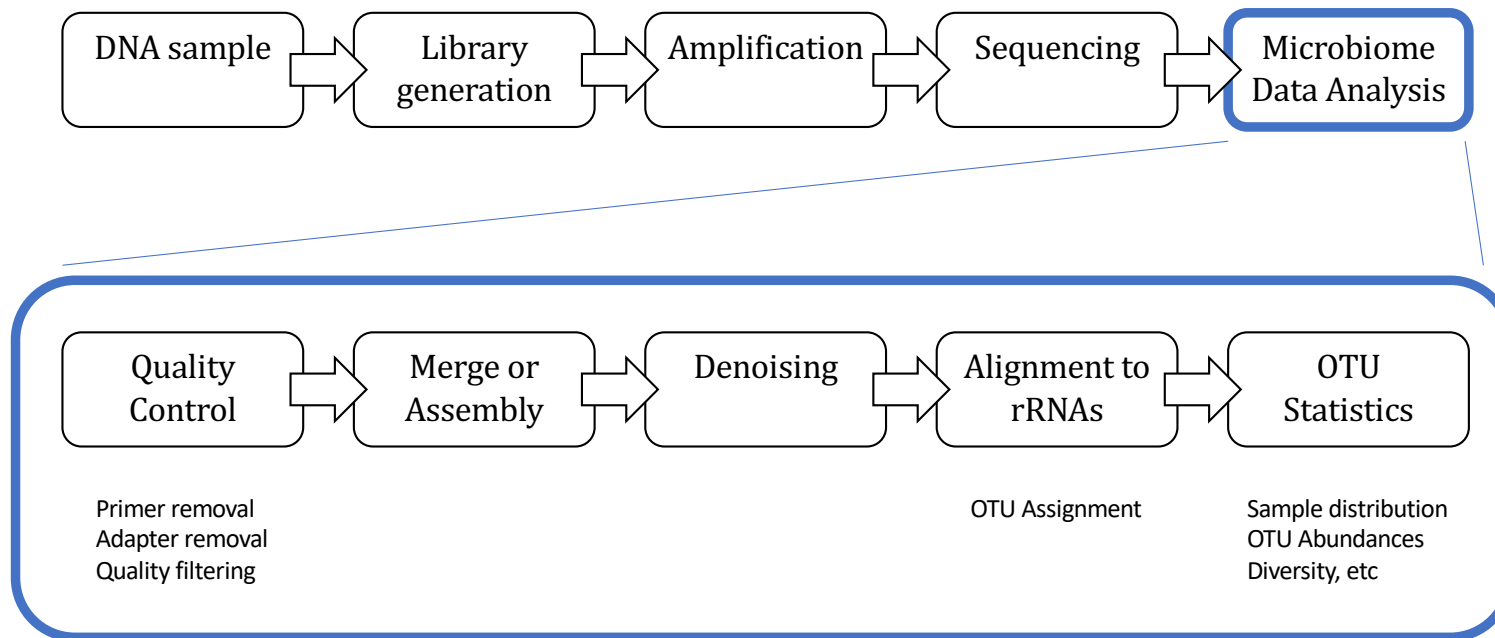
# Reproducibility

- **Literate programming:** Chunks of programming (analytical code) with human-readable text (comments)
- **Version control:** Tracking changes made to sets of files (of a project), typically program source code, scripts and documentation
- **Environment control:** Versions of all programs (plus libraries, packages, OS) used; archival copies for future reference. Example: `sessionInfo()`
- **Persistent data sharing:** collaborative, transparent, accessible science
- Documentation, e.g. README file
- Project (data + code + ...) validation
- **Command-line tools** and programming enable the creation of reproducible workflows (aka pipelines). By writing scripts or workflows, researchers can document and automate their analyses, ensuring that others can replicate their results.
- **Modularity of the command-line:** Possibility of running long pipelines of programs, one after another + *piping*

# The 16S rRNA gene for community profiling

- Operational taxonomic unit (OTU): Operational concept of classification
- Community profiling: Identifying OTUs in samples
- 16S rRNA part of the 30S small subunit (SSU) of the prokaryotic ribosome
  - All prokaryotes have one (or more) copy of this gene
  - NOT true for any protein coding genes
- Different parts of the gene exhibit different levels of conservation
  - More conserved regions can be used to analyse distantly related species
  - More variable regions can be used to analyse more closely related species
- Gene ~1540 nt
  - Not too short so as to be uninformative; not too long so as to be unmanageable
- Most 16S rRNA gene is highly conserved between different species
  - Hypervariable regions: Nine much less conserved (V1 – V9)

# Microbiome analysis



# Sequence Data Formats

## FASTA

```
head -10 egfr_flank.fasta
>ENSMUSG00000020122|ENSMUST00000138518
CCCTCCTATCATGCTGTCAGTGTATCTCTAAATAGCACTCTCAACCCCGTGAACCTGGT
TATTA AAAACATGCCCAAAGTCTGGGAGCCAGGGCTGCAGGGAATACCACAGCCTCAGT
TCATCAAAACAGTTCATTGCCCAAATGTTCTCAGCTGCAGCTTTCATGAGGTAACCTCA
GGGCCACCTGTTCTCTGGT
>ENSMUSG00000020122|ENSMUST00000125984
GAGTCAGGTTGAAGCTGCCCTGAACACTACAGAGAAGAGAGGCCTTGGTGTCTGTTGTC
TCCAGAACCCCAATATGTCTTGTGAAGGCACACAACCCCTCAAAGGGGTGCACTTCTT
CTGATCACTTTTGTACTGTTTACTAACTGATCCTATGAATCACTGTGTCTTCTCAGAGG
CCGTGAACCACGTCTGCAAT
```

>id, description  
sequence....

## FASTQ

```
@DJB775P1:248:D0MDGACXX:7:1202:12362:49613
TGCTTACTCTGCGTTGATACCACTGCTTAGATCGGAAGAGCACACGTCTGAA
+
JJJJJJJJJJJJHHHHGHFFFFFCEEEEEEDBD?DDDDDBDDDABDDCA
@DJB775P1:248:D0MDGACXX:7:1202:12782:49716
CTCTGCGTTGATACCACTGCTTACTCTGCGTTGATACCACTGCTTAGATCGG
+
IIIIIIIIIIIIHHHHHHFFFFFEECCCBCECCCCCCCCCCCCCCCC
```

@id, description  
sequence  
+indicates end of sequence  
base quality (Phred format)

# FASTQ Quality

A quality value  $Q$  is an integer mapping of  $p$  (i.e., the probability that the corresponding base call is incorrect).

$$Q_{\text{sanger}} = -10 \log_{10} p$$

$$Q_{\text{sanger}} = -10 \log_{10} p$$

$$30 = -10 \log_{10} p$$

$$\frac{30}{-10} = \log_{10} p$$

$$p = 10^{-3}$$



# Quality Control

Unnamed history

3 shown, 8 deleted

24.84 MB

11: FastQC on data 9: Raw Data

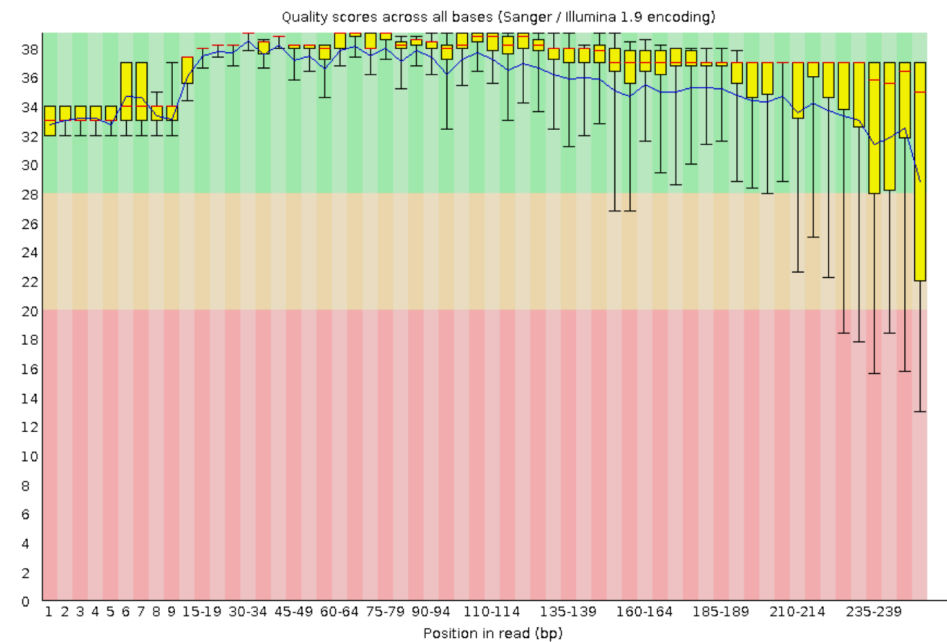
10: FastQC on data 9: We bpage

9: paired1.fastq

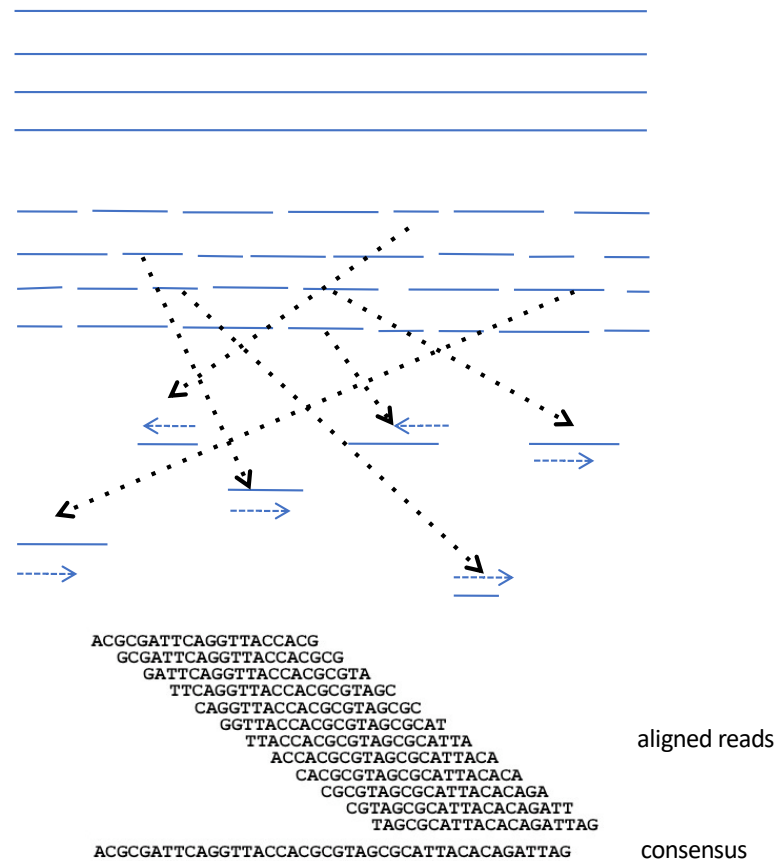
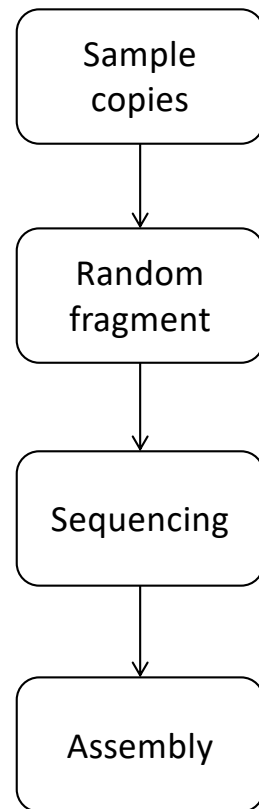
## Basic Statistics

| Measure                           | Value                   |
|-----------------------------------|-------------------------|
| Filename                          | paired1.fastq           |
| File type                         | Conventional base calls |
| Encoding                          | Sanger / Illumina 1.9   |
| Total Sequences                   | 7793                    |
| Sequences flagged as poor quality | 0                       |
| Sequence length                   | 249-251                 |
| %GC                               | 54                      |

## Per base sequence quality

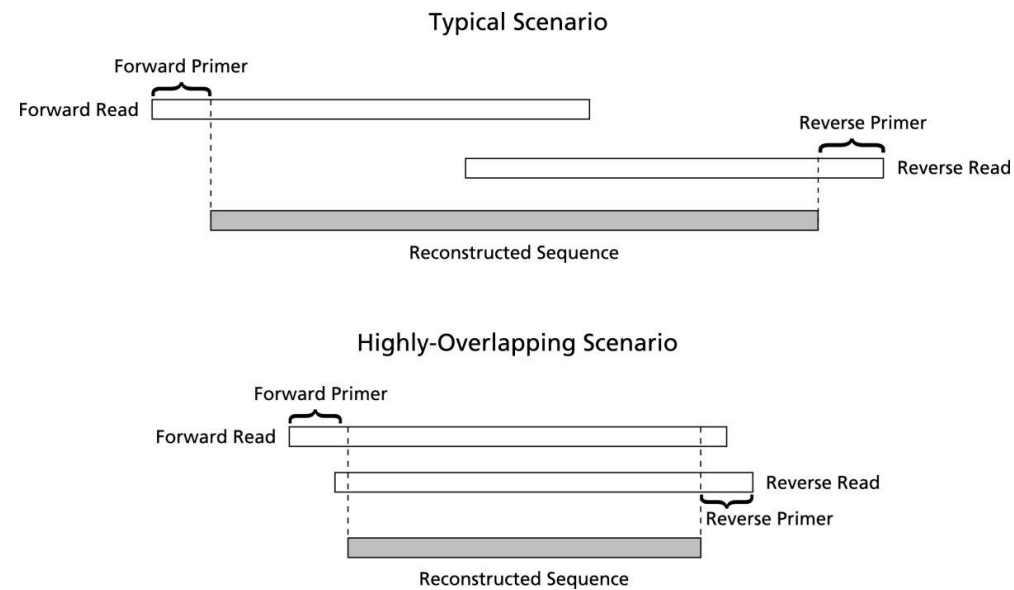


# Sequencing and Assembly

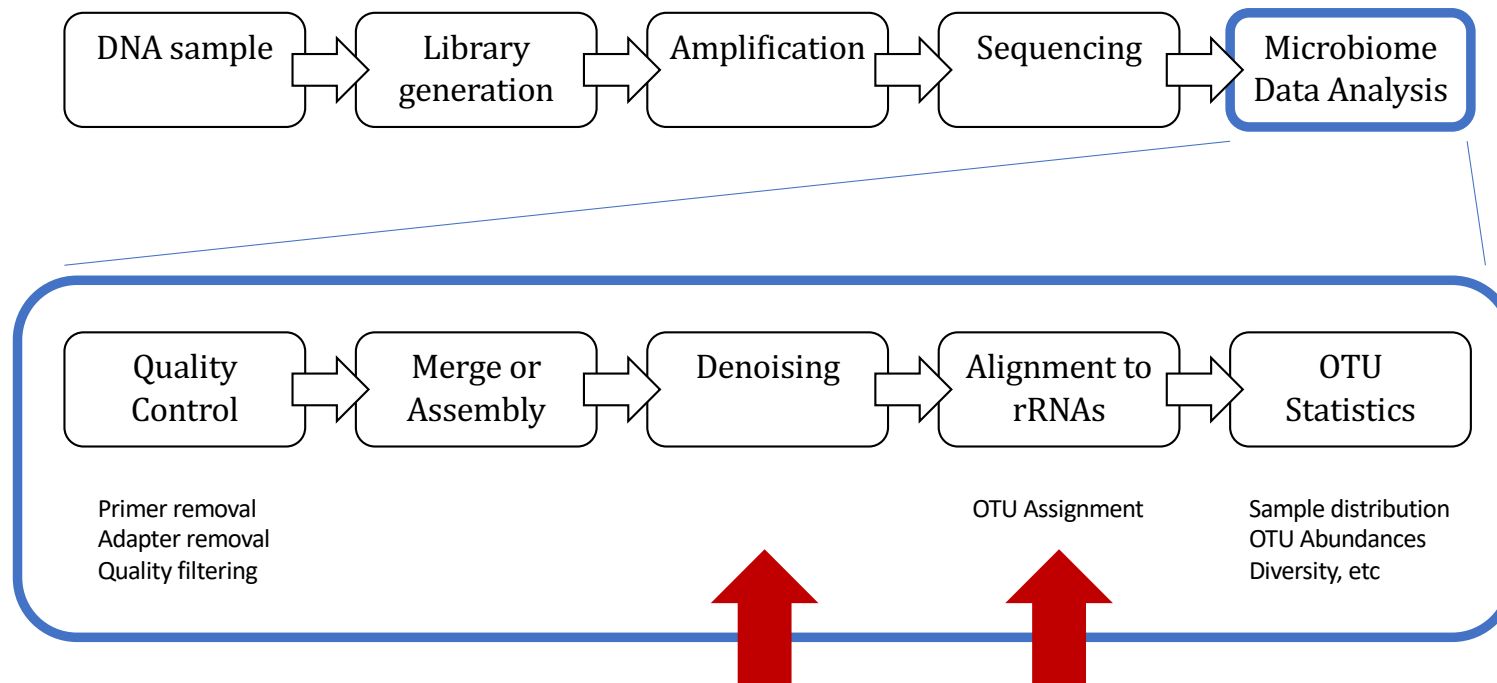


# Assembly: PANDASeq

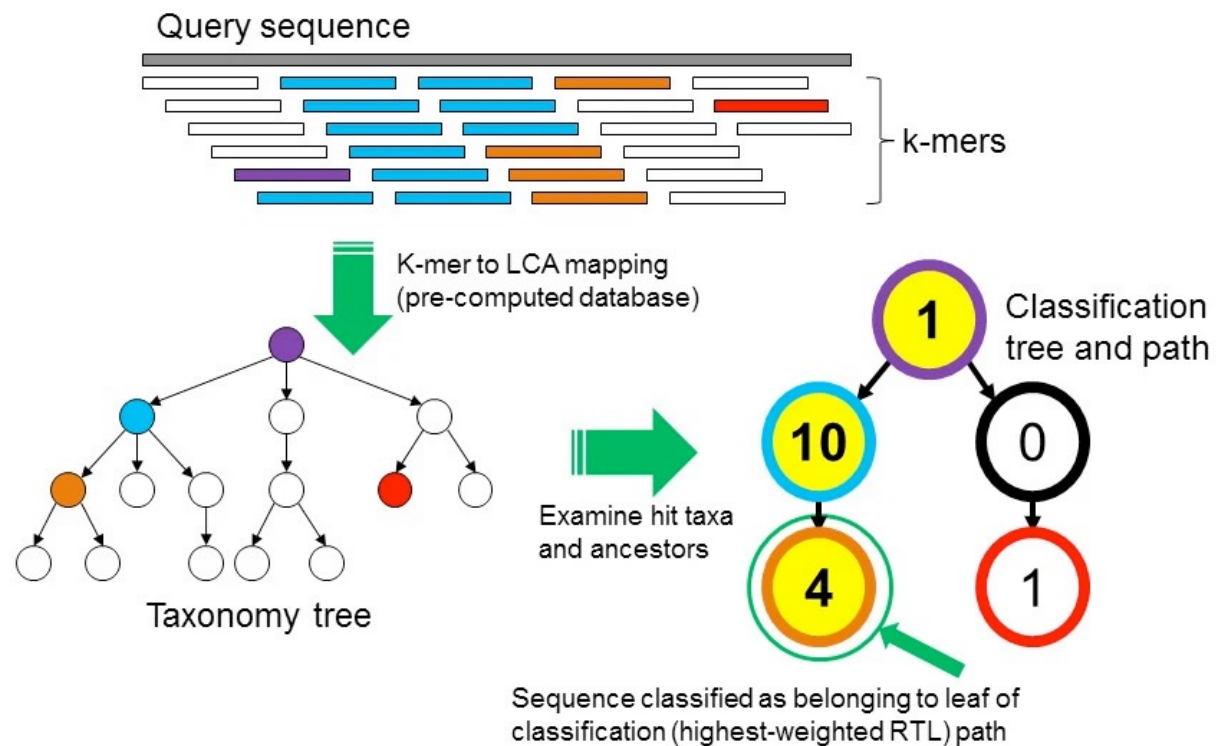
Aligns Illumina reads, and reconstruct an overlapping sequence.



# Microbiome analysis



# Kraken: Taxonomic assignment



# OTU table

- Operational Taxonomic Unit (OTU)
- Represents the taxonomic composition of a microbial community obtained from DNA sequencing data.
- A typical OTU table contains sample data in the columns, taxa information in the rows, and the amount of amplicons on each cell.

## samples

|    | A                                  | B       | C       | D       | E      | F       | G       | H      | I      | J       | K      | L      | M      | N       | O      | P      | Q        | R     | S          | T             | U                   |                    |                      |                       |                     |        |
|----|------------------------------------|---------|---------|---------|--------|---------|---------|--------|--------|---------|--------|--------|--------|---------|--------|--------|----------|-------|------------|---------------|---------------------|--------------------|----------------------|-----------------------|---------------------|--------|
| 1  | Constructed from biom file         |         |         |         |        |         |         |        |        |         |        |        |        |         |        |        |          |       |            |               |                     |                    |                      |                       |                     |        |
| 2  | OTU ID                             | Plate18 | Plate10 | Plate14 | Plate2 | Plate16 | Plate13 | Plate1 | Plate3 | Plate17 | Plate5 | Plate7 | Plate6 | Plate12 | Plate8 | Plate9 | taxonomy |       |            |               |                     |                    |                      |                       |                     |        |
| 3  | 18413ace715a299727980e78a2ed1c     | 5293    | 1874    | 1168    | 2884   | 2079    | 617     | 2077   | 815    | 1843    | 4519   | 3310   | 0      | 4913    | 3782   | 7958   | 4478     | 0     | d__Bacteri | p__Firmicutes | c__Clostridia       | o__Oscillospirales | f__Ruminococcaceae   | g__Faecalibacterium   |                     |        |
| 4  | 40a610a019d5e4e4211511d1c79a4d7520 | 3066    | 867     | 486     | 1603   | 2480    | 0       | 1570   | 523    | 845     | 2248   | 8864   | 2608   | 0       | 2821   | 2093   | 3935     | 1777  | 0          | d__Bacteri    | p__Firmicutes       | c__Clostridia      | o__Oscillospirales   | f__Ruminococcaceae    | g__Faecalibacterium |        |
| 5  | 437f0d4ec290da24184894821a57       | 2930    | 0       | 0       | 859    | 0       | 0       | 0      | 0      | 0       | 0      | 0      | 0      | 0       | 0      | 0      | 0        | 0     | 0          | d__Bacteri    | p__Bacteroidetes    | c__Bacteroidia     | o__Bacteroidales     | f__Bacteroidaceae     | g__Bacteroides      |        |
| 6  | 8713aaef804cc58a6d4421a4945442a    | 2592    | 0       | 6748    | 0      | 0       | 185     | 4      | 0      | 0       | 0      | 0      | 0      | 0       | 8      | 1961   | 0        | 0     | 0          | d__Bacteri    | p__Firmicutes       | c__Clostridia      | o__Oscillospirales   | f__Ruminococcaceae    |                     |        |
| 7  | 61227f9c3a3807c5286c78965959588    | 2227    | 434     | 1208    | 0      | 3566    | 0       | 0      | 0      | 0       | 0      | 0      | 1975   | 0       | 173    | 0      | 0        | 0     | 0          | d__Bacteri    | p__Bacteroidetes    | c__Bacteroidia     | o__Bacteroidales     | f__Bacteroidaceae     | g__Bacteroides      |        |
| 8  | 536eef8ba4535da8401214519a2e2      | 1602    | 853     | 1332    | 328    | 0       | 45      | 0      | 0      | 0       | 775    | 0      | 136    | 0       | 0      | 11151  | 0        | 0     | 0          | d__Bacteri    | p__Bacteroidetes    | c__Bacteroidia     | o__Bacteroidales     | f__Bacteroidaceae     | g__Bacteroides      |        |
| 9  | 9352a10454842932a2938a672a883      | 1529    | 0       | 0       | 5      | 0       | 14      | 4      | 3      | 0       | 0      | 0      | 14     | 10      | 103    | 32     | 254      | 3     | 20         | d__Bacteri    | p__Firmicutes       | c__Bacilli         | o__Lactobacillales   | f__Streptococcaceae   | g__Lactococcus      |        |
| 10 | 58b7f02629f919a78f155a093217       | 1471    | 0       | 0       | 446    | 0       | 0       | 0      | 0      | 0       | 0      | 0      | 0      | 0       | 0      | 0      | 0        | 0     | 0          | d__Bacteri    | p__Bacteroidetes    | c__Bacteroidia     | o__Bacteroidales     | f__Bacteroidaceae     | g__Bacteroides      |        |
| 11 | c45183a7a7a8d878a4375d4a6c         | 1400    | 231     | 0       | 793    | 295     | 0       | 2645   | 0      | 49      | 0      | 0      | 2460   | 0       | 338    | 2      | 341      | 9055  | 1          | d__Bacteri    | p__Firmicutes       | c__Clostridia      | o__Oscillospirales   | f__Ruminococcaceae    | g__Ruminococcus     | s__Rum |
| 12 | 0a945a4d21c3584c7a3275ca64c39      | 1307    | 0       | 4036    | 0      | 0       | 79      | 2      | 5      | 0       | 0      | 0      | 0      | 0       | 5      | 1052   | 0        | 0     | 0          | d__Bacteri    | p__Firmicutes       | c__Clostridia      | o__Oscillospirales   | f__Ruminococcaceae    |                     |        |
| 13 | c2a7a0010a67888f23302d83a61c11     | 1108    | 257     | 690     | 0      | 4794    | 0       | 0      | 0      | 0       | 0      | 0      | 1663   | 0       | 0      | 0      | 0        | 0     | 0          | d__Bacteri    | p__Bacteroidetes    | c__Bacteroidia     | o__Bacteroidales     | f__Bacteroidaceae     | g__Bacteroides      |        |
| 14 | 1f417a7c26cd891feeb8a8818c9573     | 1014    | 1477    | 13024   | 2041   | 810     | 1016    | 595    | 1742   | 6224    | 1356   | 6785   | 2054   | 4637    | 5730   | 20556  | 2739     | 10593 | 5278       | d__Bacteri    | p__Firmicutes       | c__Clostridia      | o__Lachnospirales    | f__Lachnospiraceae    | g__Agathobacter     |        |
| 15 | a72a0ba42a57569454951c29594c       | 917     | 0       | 0       | 384    | 0       | 0       | 4762   | 0      | 0       | 0      | 0      | 0      | 0       | 0      | 1897   | 843      | 193   | 0          | d__Bacteri    | p__Firmicutes       | c__Clostridia      | o__Lachnospirales    | f__Lachnospiraceae    |                     |        |
| 16 | 3520606a07747137aa7e3231c14f       | 866     | 787     | 1996    | 239    | 1032    | 367     | 561    | 1245   | 2719    | 371    | 747    | 955    | 7424    | 5337   | 0      | 1804     | 1907  | 1085       | d__Bacteri    | p__Firmicutes       | c__Clostridia      | o__Lachnospirales    | f__Lachnospiraceae    | g__Blautia          |        |
| 17 | 84a0a6767a2f10aef9f0446712a52a     | 852     | 264     | 439     | 0      | 0       | 7       | 3      | 0      | 2450    | 4      | 0      | 0      | 30      | 173    | 2456   | 109      | 0     | 22         | d__Bacteri    | p__Firmicutes       | c__Bacilli         | o__Lactobacillales   | f__Streptococcaceae   | g__Streptococcus    |        |
| 18 | 1a487ae09f05086e5598a48311781      | 806     | 349     | 1033    | 816    | 1282    | 3121    | 940    | 2417   | 5180    | 1305   | 4523   | 1264   | 0       | 1806   | 617    | 1887     | 2830  | 1615       | d__Bacteri    | p__Firmicutes       | c__Clostridia      | o__Lachnospirales    | f__Lachnospiraceae    |                     |        |
| 19 | 45c12726235e9f9f159a4a509f49       | 799     | 0       | 735     | 92     | 0       | 0       | 0      | 0      | 0       | 387    | 0      | 191    | 2       | 0      | 6406   | 0        | 0     | 0          | d__Bacteri    | p__Bacteroidetes    | c__Bacteroidia     | o__Bacteroidales     | f__Bacteroidaceae     | g__Bacteroides      |        |
| 20 | ca0ba0a64f4cb3575a5a15c4077        | 784     | 198     | 2823    | 285    | 706     | 395     | 496    | 1893   | 9551    | 1777   | 10190  | 3131   | 0       | 59     | 12857  | 5235     | 9953  | 15141      | d__Bacteri    | p__Bacteroidetes    | c__Bacteroidia     | o__Bacteroidales     | f__Bacteroidaceae     | g__Bacteroides      |        |
| 21 | 054af07287a272a3204cc85a11701f     | 775     | 0       | 0       | 3665   | 0       | 883     | 1194   | 0      | 788     | 2085   | 1360   | 3054   | 0       | 5429   | 2427   | 3676     | 2557  | 14028      | d__Bacteri    | p__Firmicutes       | c__Clostridia      | o__Oscillospirales   | f__Ruminococcaceae    | g__Subdoligranulum  |        |
| 22 | 9497525ee8f2183f26367864872349     | 772     | 0       | 3       | 0      | 0       | 0       | 0      | 0      | 0       | 0      | 0      | 9      | 0       | 0      | 0      | 0        | 0     | 0          | d__Bacteri    | p__Firmicutes       | c__Bacilli         | o__Lactobacillales   | f__Streptococcaceae   | g__Lactococcus      |        |
| 23 | 88c037a872107f48a40885116a4679f    | 755     | 0       | 2164    | 0      | 1353    | 0       | 0      | 6477   | 0       | 0      | 0      | 297    | 0       | 363    | 0      | 0        | 0     | 0          | d__Bacteri    | p__Bacteroidetes    | c__Bacteroidia     | o__Bacteroidales     | f__Bacteroidaceae     | g__Bacteroides      |        |
| 24 | 0521f6e285174073d201a16a6b72bb     | 545     | 4       | 0       | 702    | 122     | 0       | 11961  | 4300   | 5847    | 0      | 478    | 0      | 2456    | 0      | 3606   | 6131     | 1385  | 0          | d__Bacteri    | p__Bacteroidetes    | c__Bacteroidia     | o__Bacteroidales     | f__Bacteroidaceae     | g__Bacteroides      |        |
| 25 | 7b194ba3e274e5746e5ca0ba0c54a52    | 528     | 0       | 0       | 198    | 0       | 0       | 3206   | 0      | 0       | 0      | 0      | 0      | 0       | 0      | 0      | 98       | 0     | 0          | d__Bacteri    | p__Firmicutes       | c__Clostridia      | o__Lachnospirales    | f__Lachnospiraceae    |                     |        |
| 26 | 42215a418204049d1e30d1a0f8039ae    | 523     | 771     | 7846    | 1106   | 962     | 0       | 427    | 1108   | 3313    | 0      | 0      | 1613   | 0       | 3194   | 11635  | 0        | 5147  | 0          | d__Bacteri    | p__Firmicutes       | c__Clostridia      | o__Lachnospirales    | f__Lachnospiraceae    | g__Agathobacter     |        |
| 27 | 336244c983034a0c7946323c364d6d     | 515     | 0       | 0       | 0      | 0       | 0       | 0      | 0      | 0       | 0      | 0      | 0      | 0       | 0      | 346    | 0        | 0     | 0          | d__Bacteri    | p__Firmicutes       | c__Clostridia      | o__Oscillospirales   | f__Ruminococcaceae    | g__Faecalibacterium |        |
| 28 | c1a072256a737a23621072026a4a0b     | 497     | 188     | 648     | 415    | 1478    | 1947    | 563    | 1999   | 2680    | 0      | 0      | 958    | 0       | 913    | 0      | 0        | 1498  | 810        | d__Bacteri    | p__Firmicutes       | c__Clostridia      | o__Lachnospirales    | f__Lachnospiraceae    |                     |        |
| 29 | 154a013a03a6a4a36351a11651a07c9    | 469     | 0       | 1247    | 0      | 1347    | 0       | 374    | 852    | 1589    | 195    | 0      | 840    | 3129    | 3030   | 0      | 922      | 0     | 0          | d__Bacteri    | p__Firmicutes       | c__Clostridia      | o__Lachnospirales    | f__Lachnospiraceae    | g__Blautia          |        |
| 30 | 8633a67a2c11ee424a45974782         | 451     | 0       | 0       | 2158   | 0       | 358     | 854    | 0      | 0       | 1045   | 0      | 2476   | 0       | 2884   | 1377   | 0        | 0     | 0          | d__Bacteri    | p__Firmicutes       | c__Clostridia      | o__Oscillospirales   | f__Ruminococcaceae    | g__Subdoligranulum  |        |
| 31 | ed72041a6d5d2f6d50f6786c7f         | 446     | 98      | 1721    | 201    | 948     | 118     | 373    | 1363   | 5446    | 737    | 0      | 2853   | 0       | 25     | 7342   | 0        | 5412  | 10555      | d__Bacteri    | p__Bacteroidetes    | c__Bacteroidia     | o__Bacteroidales     | f__Bacteroidaceae     | g__Bacteroides      |        |
| 32 | f07d069972a9a2652465307b704a20     | 438     | 1562    | 1516    | 662    | 485     | 388     | 759    | 459    | 2561    | 2216   | 4959   | 711    | 8985    | 6811   | 1156   | 5581     | 1450  | 4641       | d__Bacteri    | p__Firmicutes       | c__Clostridia      | o__Lachnospirales    | f__Lachnospiraceae    | g__Blautia          |        |
| 33 | 451740a49a1e4984a0fda13133c23      | 414     | 0       | 0       | 0      | 0       | 0       | 0      | 0      | 0       | 0      | 0      | 0      | 0       | 0      | 0      | 0        | 0     | 0          | d__Bacteri    | p__Bacteroidetes    | c__Bacteroidia     | o__Bacteroidales     | f__Prevotellaceae     |                     |        |
| 34 | 1e1d519b746020d6781d0e32f9a0a4     | 406     | 185     | 270     | 0      | 0       | 2       | 2      | 0      | 1348    | 1      | 0      | 0      | 0       | 126    | 1311   | 39       | 0     | 49         | d__Bacteri    | p__Firmicutes       | c__Bacilli         | o__Lactobacillales   | f__Streptococcaceae   | g__Streptococcus    |        |
| 35 | 8343607a7736c278a298c3a213c0       | 403     | 0       | 0       | 285    | 111     | 0       | 395    | 0      | 0       | 0      | 0      | 775    | 0       | 0      | 613    | 362      | 0     | 0          | d__Bacteri    | p__Firmicutes       | c__Clostridia      | o__Oscillospirales   | f__Ruminococcaceae    | g__Faecalibacterium |        |
| 36 | c540a8c1854f19743ba4c36a29f59f     | 384     | 873     | 1284    | 135    | 949     | 112     | 843    | 227    | 11224   | 591    | 0      | 0      | 2       | 0      | 798    | 7088     | 0     | 0          | d__Bacteri    | p__Actinobacteriota | c__Actinobacteria  | o__Bifidobacteriales | f__Bifidobacteriaceae | g__Bifidobacter     |        |
| 37 | 403b3a085294a1ba0177a29553181f     | 344     | 349     | 0       | 800    | 502     | 449     | 153    | 150    | 4182    | 865    | 0      | 2266   | 0       | 0      | 0      | 694      | 1150  | 0          | d__Bacteri    | p__Bacteroidetes    | c__Bacteroidia     | o__Bacteroidales     | f__Tannerellaceae     | g__Parabacteroides  | s__Pa  |
| 38 | 443735461320379a2198a0a0a25        | 342     | 0       | 402     | 0      | 44      | 0       | 8052   | 2298   | 3621    | 0      | 335    | 0      | 1366    | 1805   | 3195   | 0        | 0     | 0          | d__Bacteri    | p__Bacteroidetes    | c__Bacteroidia     | o__Bacteroidales     | f__Bacteroidaceae     | g__Bacteroides      |        |
| 39 | 197932a3e11228a04a070f932d         | 333     | 0       | 1534    | 0      | 1842    | 0       | 0      | 4839   | 0       | 0      | 0      | 246    | 0       | 273    | 0      | 0        | 0     | 0          | d__Bacteri    | p__Bacteroidetes    | c__Bacteroidia     | o__Bacteroidales     | f__Bacteroidaceae     | g__Bacteroides      |        |
| 40 | 4867463c784587818761a7a4a4a2a      | 301     | 0       | 0       | 76     | 0       | 0       | 0      | 0      | 0       | 0      | 0      | 0      | 0       | 0      | 0      | 0        | 0     | 0          | d__Bacteri    | p__Bacteroidetes    | c__Bacteroidia     | o__Bacteroidales     | f__Bacteroidaceae     | g__Bacteroides      |        |
| 41 | 1a0a031f1f523a6a1f3501f5a00cc3c    | 191     | 53      | 0       | 301    | 450     | 318     | 360    | 51     | 647     | 7678   | 5071   | 1546   | 0       | 0      | 1736   | 7801     | 0     | 0          | d__Bacteri    | p__Bacteroidetes    | c__Bacteroidia     | o__Bacteroidales     | f__Bifidobacteriaceae | g__Bifidobacter     |        |

taxa

# Example OTU Table

samples

taxa

|    | A                                 | B       | C       | D       | E      | F       | G       | H       | I       | J      | K      | L       | M      | N      | O      | P      | Q       | R      | S      | T   | U |
|----|-----------------------------------|---------|---------|---------|--------|---------|---------|---------|---------|--------|--------|---------|--------|--------|--------|--------|---------|--------|--------|---|---|
| 1  | # Constructed from biom file      |         |         |         |        |         |         |         |         |        |        |         |        |        |        |        |         |        |        |   |   |
| 2  | #OTU ID                           | Platz18 | Platz10 | Platz14 | Platz2 | Platz11 | Platz16 | Platz15 | Platz13 | Platz1 | Platz3 | Platz17 | Platz5 | Platz7 | Platz6 | Platz4 | Platz12 | Platz8 | Platz9 | taxonomy  |   |
| 3  | 18413ece715a299f717980e782e3ed1c  | 5293    | 1874    | 1168    | 2984   | 2079    | 917     | 2077    | 815     | 1843   | 4594   | 20519   | 3310   | 0      | 4913   | 3782   | 7958    | 4478   | 0      | d__Bacteria; p__Firmicutes; c__Clostridia; o__Oscillospirales; f__Ruminococcaceae; g__Faecalibacterium            |   |
| 4  | d0a610a019d5ce4e211611d2f9ad7520  | 3066    | 867     | 486     | 1603   | 2480    | 0       | 1570    | 523     | 845    | 2248   | 8864    | 2608   | 0      | 2821   | 2093   | 3935    | 1777   | 0      | d__Bacteria; p__Firmicutes; c__Clostridia; o__Oscillospirales; f__Ruminococcaceae; g__Faecalibacterium            |   |
| 5  | 437fc004ec290da2418d69d46f821a57  | 2930    | 0       | 0       | 859    | 0       | 0       | 0       | 0       | 0      | 0      | 0       | 0      | 0      | 0      | 0      | 0       | 0      | 0      | d__Bacteria; p__Bacteroidota; c__Bacteroidia; o__Bacteroidales; f__Bacteroidaceae; g__Bacteroides                 |   |
| 6  | 8711aef804cc58e66f4b21a43d3442a   | 2592    | 0       | 6748    | 0      | 0       | 185     | 4       | 0       | 0      | 0      | 0       | 0      | 0      | 8      | 1961   | 0       | 0      | 0      | d__Bacteria; p__Firmicutes; c__Clostridia; o__Oscillospirales; f__Ruminococcaceae                                 |   |
| 7  | 61227bf9cb3b907c5286c7896659b588  | 2227    | 434     | 1208    | 0      | 3566    | 0       | 0       | 0       | 0      | 0      | 1975    | 0      | 173    | 0      | 0      | 0       | 0      | 0      | d__Bacteria; p__Bacteroidota; c__Bacteroidia; o__Bacteroidales; f__Bacteroidaceae; g__Bacteroides                 |   |
| 8  | 536eefdfbba4535eba8401214519b2e2  | 1602    | 853     | 1332    | 328    | 0       | 45      | 0       | 0       | 0      | 775    | 0       | 136    | 0      | 0      | 11151  | 0       | 0      | 0      | d__Bacteria; p__Bacteroidota; c__Bacteroidia; o__Bacteroidales; f__Bacteroidaceae; g__Bacteroides                 |   |
| 9  | 9352c04b5484253922e9338eb71ba883  | 1529    | 0       | 0       | 5      | 0       | 14      | 4       | 3       | 0      | 0      | 14      | 10     | 103    | 32     | 254    | 3       | 20     | 0      | d__Bacteria; p__Firmicutes; c__Bacilli; o__Lactobacillales; f__Streptococcaceae; g__Lactococcus                   |   |
| 10 | 588e76106293f918e78f335a099321f7  | 1471    | 0       | 0       | 446    | 0       | 0       | 0       | 0       | 0      | 0      | 0       | 0      | 0      | 0      | 0      | 0       | 0      | 0      | d__Bacteria; p__Bacteroidota; c__Bacteroidia; o__Bacteroidales; f__Bacteroidaceae; g__Bacteroides                 |   |
| 11 | ec51863a70cd9a8ab8f28e4375e9da0c  | 1430    | 231     | 0       | 793    | 295     | 0       | 2645    | 0       | 49     | 0      | 2460    | 0      | 328    | 2      | 341    | 9055    | 1      | 0      | d__Bacteria; p__Firmicutes; c__Clostridia; o__Oscillospirales; f__Ruminococcaceae; g__Ruminococcus; s__Rum        |   |
| 12 | 0a9454a4d21fc3586e7463275ca64e39  | 1307    | 0       | 4036    | 0      | 0       | 79      | 2       | 5       | 0      | 0      | 0       | 0      | 0      | 5      | 1052   | 0       | 0      | 0      | d__Bacteria; p__Firmicutes; c__Clostridia; o__Oscillospirales; f__Ruminococcaceae                                 |   |
| 13 | c2c37a901bbb7868fc23502d83a61c11  | 1108    | 257     | 690     | 0      | 4794    | 0       | 0       | 0       | 0      | 0      | 1663    | 0      | 0      | 0      | 0      | 0       | 0      | 0      | d__Bacteria; p__Bacteroidota; c__Bacteroidia; o__Bacteroidales; f__Bacteroidaceae; g__Bacteroides                 |   |
| 14 | fbf417e4726cbd91feebbb8818c69573  | 1014    | 1477    | 13014   | 2041   | 810     | 1016    | 595     | 1742    | 6224   | 1356   | 6785    | 2054   | 4637   | 5730   | 20656  | 2739    | 10593  | 5278   | d__Bacteria; p__Firmicutes; c__Clostridia; o__Lachnospirales; f__Lachnospiraceae; g__Agathobacter                 |   |
| 15 | af2ca0bbc42a577694f3d91fc29594c   | 917     | 0       | 0       | 384    | 0       | 0       | 4762    | 0       | 0      | 0      | 0       | 0      | 0      | 0      | 1897   | 843     | 193    | 0      | d__Bacteria; p__Firmicutes; c__Clostridia; o__Lachnospirales; f__Lachnospiraceae                                  |   |
| 16 | 3520606da0f7747137aee7e2331cc14f  | 866     | 787     | 1996    | 239    | 1032    | 367     | 561     | 1245    | 2719   | 371    | 747     | 955    | 7424   | 5337   | 0      | 1804    | 1907   | 1085   | d__Bacteria; p__Firmicutes; c__Clostridia; o__Lachnospirales; f__Lachnospiraceae; g__Blautia                      |   |
| 17 | 84a0b6767a2fc10ad3fbf446712e5b2d  | 852     | 264     | 439     | 0      | 0       | 7       | 3       | 0       | 2450   | 4      | 0       | 30     | 173    | 2456   | 109    | 0       | 22     | 0      | d__Bacteria; p__Firmicutes; c__Bacilli; o__Lactobacillales; f__Streptococcaceae; g__Streptococcus                 |   |
| 18 | fa4d87aed60165b86e55598bd3631781  | 806     | 349     | 1033    | 816    | 1282    | 3131    | 940     | 2417    | 5160   | 1325   | 4523    | 1264   | 0      | 1806   | 617    | 1887    | 2830   | 1615   | d__Bacteria; p__Firmicutes; c__Clostridia; o__Lachnospirales; f__Lachnospiraceae                                  |   |
| 19 | 4c6c1327d3295eaf9ff519be24059fa9  | 799     | 0       | 735     | 92     | 0       | 0       | 0       | 0       | 0      | 387    | 0       | 191    | 2      | 0      | 6406   | 0       | 0      | 0      | d__Bacteria; p__Bacteroidota; c__Bacteroidia; o__Bacteroidales; f__Bacteroidaceae; g__Bacteroides                 |   |
| 20 | cec04be4ab4f4cbe3375ae5aa15d2ff7  | 784     | 168     | 2623    | 285    | 706     | 395     | 496     | 1853    | 9551   | 1777   | 10190   | 3131   | 0      | 58     | 12957  | 5225    | 9953   | 19141  | d__Bacteria; p__Bacteroidota; c__Bacteroidia; o__Bacteroidales; f__Bacteroidaceae; g__Bacteroides                 |   |
| 21 | 054bfff28f2ed72af30d4cc85a31701f  | 775     | 0       | 0       | 3665   | 0       | 883     | 1194    | 0       | 788    | 2085   | 1360    | 3054   | 0      | 5429   | 2427   | 3576    | 2557   | 1408   | d__Bacteria; p__Firmicutes; c__Clostridia; o__Oscillospirales; f__Ruminococcaceae; g__Subdoligranulum             |   |
| 22 | b497525ee3fcf2183f2667f8648723d9  | 772     | 0       | 3       | 0      | 0       | 0       | 0       | 0       | 0      | 0      | 0       | 9      | 0      | 0      | 0      | 0       | 0      | 0      | d__Bacteria; p__Firmicutes; c__Bacilli; o__Lactobacillales; f__Streptococcaceae; g__Lactococcus                   |   |
| 23 | 88c037a8f27f07c8b4088511eba6679f  | 755     | 0       | 2164    | 0      | 1353    | 0       | 0       | 6477    | 0      | 0      | 0       | 297    | 0      | 363    | 0      | 0       | 0      | 0      | d__Bacteria; p__Bacteroidota; c__Bacteroidia; o__Bacteroidales; f__Bacteroidaceae; g__Bacteroides                 |   |
| 24 | 0521fe8c285174073d2e20a166b972bb  | 545     | 4       | 0       | 702    | 0       | 122     | 0       | 11961   | 4300   | 5847   | 0       | 478    | 0      | 2456   | 0      | 3606    | 6131   | 1385   | d__Bacteria; p__Bacteroidota; c__Bacteroidia; o__Bacteroidales; f__Bacteroidaceae; g__Bacteroides                 |   |
| 25 | 7b194de3e2f4a5746e5cad8adc54bc92  | 528     | 0       | 0       | 198    | 0       | 0       | 3206    | 0       | 0      | 0      | 0       | 0      | 0      | 0      | 0      | 0       | 98     | 0      | d__Bacteria; p__Firmicutes; c__Clostridia; o__Lachnospirales; f__Lachnospiraceae                                  |   |
| 26 | 4221b5e418260469d1e30d1bbf803fae  | 523     | 771     | 7846    | 1106   | 962     | 0       | 427     | 1108    | 3313   | 0      | 1613    | 0      | 3194   | 11635  | 0      | 5147    | 0      | 0      | d__Bacteria; p__Firmicutes; c__Clostridia; o__Lachnospirales; f__Lachnospiraceae; g__Agathobacter                 |   |
| 27 | b58244c9e38034ba80c794b32c36d6d   | 515     | 0       | 0       | 0      | 0       | 0       | 0       | 0       | 0      | 0      | 0       | 0      | 0      | 0      | 346    | 0       | 0      | 0      | d__Bacteria; p__Firmicutes; c__Clostridia; o__Oscillospirales; f__Ruminococcaceae; g__Faecalibacterium            |   |
| 28 | c114a0220545f73a2362c1072926a48b  | 497     | 188     | 648     | 415    | 1478    | 1347    | 563     | 1399    | 2660   | 0      | 958     | 0      | 913    | 0      | 0      | 1498    | 810    | 0      | d__Bacteria; p__Firmicutes; c__Clostridia; o__Lachnospirales; f__Lachnospiraceae                                  |   |
| 29 | af5f013da1d4de3b9261a41b051a07c9  | 469     | 0       | 1247    | 0      | 1347    | 0       | 374     | 852     | 1509   | 195    | 0       | 840    | 3123   | 3030   | 0      | 922     | 0      | 0      | d__Bacteria; p__Firmicutes; c__Clostridia; o__Lachnospirales; f__Lachnospiraceae; g__Blautia                      |   |
| 30 | 8633d67a4c3f1cecf4244fc45f47f67   | 451     | 0       | 0       | 2158   | 0       | 358     | 854     | 0       | 1045   | 0      | 2476    | 0      | 2884   | 1377   | 0      | 0       | 0      | 0      | d__Bacteria; p__Firmicutes; c__Clostridia; o__Oscillospirales; f__Ruminococcaceae; g__Subdoligranulum             |   |
| 31 | bcd72c9441a6d1d2f0ef50627f8fc7ff  | 446     | 98      | 1721    | 201    | 948     | 118     | 373     | 1363    | 5446   | 737    | 0       | 2853   | 0      | 25     | 7342   | 0       | 5412   | 10165  | d__Bacteria; p__Bacteroidota; c__Bacteroidia; o__Bacteroidales; f__Bacteroidaceae; g__Bacteroides                 |   |
| 32 | fb7b2069972a9a26524653070b77dc0d  | 438     | 1562    | 1516    | 662    | 485     | 388     | 759     | 4459    | 2561   | 2216   | 4959    | 711    | 8985   | 6811   | 1156   | 5581    | 1450   | 4641   | d__Bacteria; p__Firmicutes; c__Clostridia; o__Lachnospirales; f__Lachnospiraceae; g__Blautia                      |   |
| 33 | d51740e49a6fe498a0afdc13133c23    | 414     | 0       | 0       | 0      | 0       | 0       | 0       | 0       | 0      | 0      | 0       | 0      | 0      | 0      | 0      | 0       | 0      | 0      | d__Bacteria; p__Bacteroidota; c__Bacteroidia; o__Bacteroidales; f__Prevotellaceae                                 |   |
| 34 | 1e1df519b746020db781dbeb323fabba4 | 406     | 185     | 270     | 0      | 0       | 2       | 2       | 0       | 1348   | 1      | 0       | 0      | 0      | 126    | 1311   | 39      | 0      | 0      | d__Bacteria; p__Firmicutes; c__Bacilli; o__Lactobacillales; f__Streptococcaceae; g__Streptococcus                 |   |
| 35 | 834c3607d747f3d278b590a3e213c0    | 403     | 0       | 0       | 285    | 111     | 0       | 395     | 0       | 0      | 0      | 775     | 0      | 0      | 613    | 362    | 0       | 0      | 0      | d__Bacteria; p__Firmicutes; c__Clostridia; o__Oscillospirales; f__Ruminococcaceae; g__Faecalibacterium            |   |
| 36 | c5646be185ef619743be4e3b5cf919ff  | 384     | 873     | 1284    | 153    | 949     | 112     | 843     | 227     | 11224  | 591    | 0       | 0      | 2      | 0      | 798    | 7088    | 0      | 0      | d__Bacteria; p__Actinobacteriota; c__Actinobacteria; o__Bifidobacteriales; f__Bifidobacteriaceae; g__Bifidobacter |   |
| 37 | d60b3a685294e41bcd11f72d0555181f  | 344     | 349     | 0       | 860    | 502     | 449     | 153     | 150     | 4162   | 885    | 0       | 2266   | 0      | 0      | 634    | 1150    | 0      | 0      | d__Bacteria; p__Bacteroidota; c__Bacteroidia; o__Bacteroidales; f__Tannerellaceae; g__Parabacteroides; s__Pa      |   |
| 38 | 4c437305a6132032f9c2318be9cab25   | 342     | 0       | 0       | 402    | 0       | 44      | 0       | 862     | 2298   | 3062   | 0       | 335    | 0      | 1366   | 0      | 1885    | 3105   | 0      | d__Bacteria; p__Bacteroidota; c__Bacteroidia; o__Bacteroidales; f__Bacteroidaceae; g__Bacteroides                 |   |
| 39 | 357932f2acdb1f12289aff4b2f932df   | 333     | 0       | 1514    | 0      | 1842    | 0       | 0       | 4839    | 0      | 0      | 246     | 0      | 273    | 0      | 0      | 0       | 0      | 0      | d__Bacteria; p__Bacteroidota; c__Bacteroidia; o__Bacteroidales; f__Bacteroidaceae; g__Bacteroides                 |   |
| 40 | 4b67dcdb7b4587b187b1fa71a4a942a   | 301     | 0       | 0       | 76     | 0       | 0       | 0       | 0       | 0      | 0      | 0       | 0      | 0      | 0      | 0      | 0       | 0      | 0      | d__Bacteria; p__Bacteroidota; c__Bacteroidia; o__Bacteroidales; f__Bacteroidaceae; g__Bacteroides                 |   |
| 41 | 0b08237cf103234da11470707c95c5c35 | 273     | 22      | 0       | 203    | 438     | 318     | 260     | 31      | 647    | 7598   | 2071    | 1546   | 0      | 0      | 0      | 1736    | 2861   | 0      | d__Bacteria; p__Bacteroidota; c__Bacteroidia; o__Bacteroidales; f__Bifidobacteriaceae; g__Alistipes               |   |

e.g. ~/data2024/mbtmicrobiome2024.tsv

# Steps for the bioinformatics analysis of microbiome data

|   |   |   |
|---|---|---|
| Alignment to rRNA sequences from standard databases | T | F |
| microRNA identification                             | T | F |
| RNA structure prediction                            | T | F |
| Quality control                                     | T | F |
| Denoising or quality filtering                      | T | F |

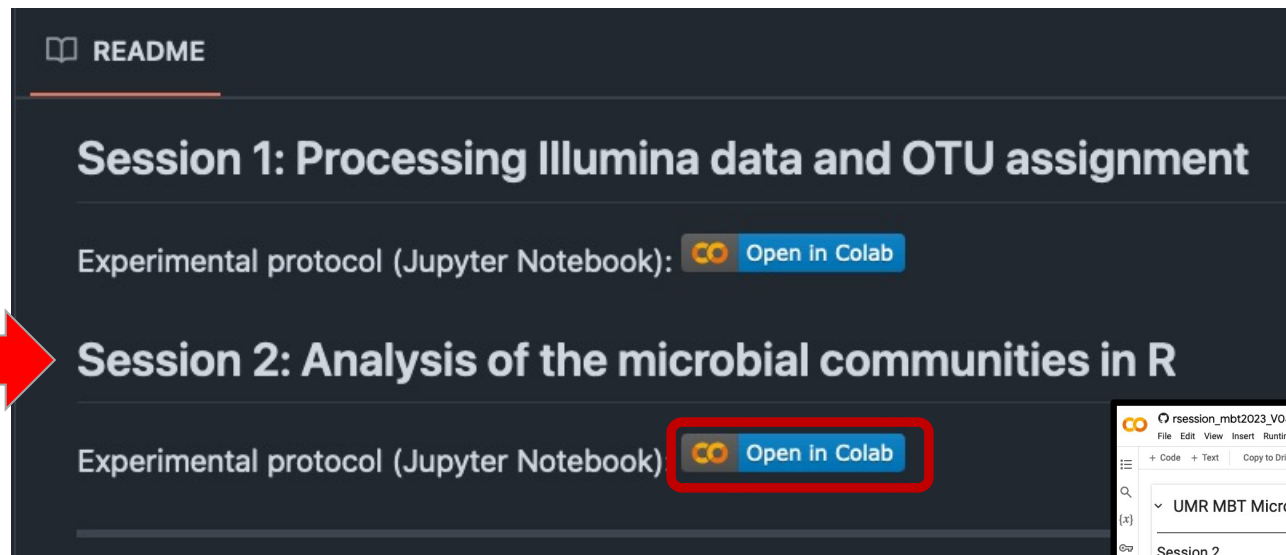


# Assembly in microbiome analysis means?

|   |   |   |
|---|---|---|
| Discard sequences with poor FASTQ quality   | T | F |
| Comparison (alignment) of the amplicon sequences versus known rRNAs from ribosomal rRNA databases | T | F |
| Preparing OTU tables by counting the number of rRNA matches present for each taxa                 | T | F |
| Merging fragments (in FASTQ format) into whole sequences (in FASTA format)                        | T | F |
| Building a small database of ribosomal RNAs   | T | F |


# Data and Links for Hands-on Practice

<https://github.com/barrantesisrael/mbtmicrobiome2023>




README

## Session 1: Processing Illumina data and OTU assignment

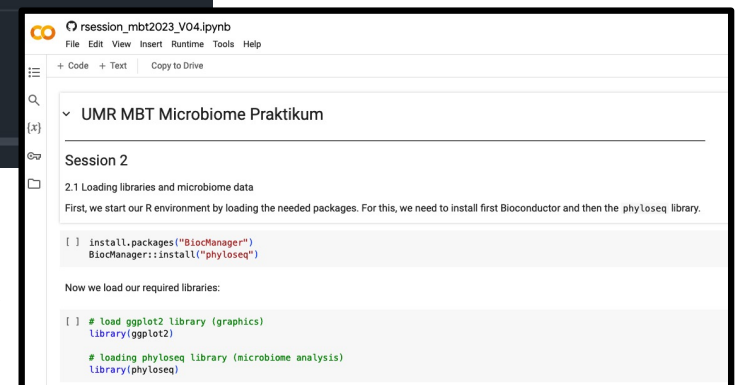
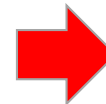
Experimental protocol (Jupyter Notebook):  [Open in Colab](#)

## Session 2: Analysis of the microbial communities in R

Experimental protocol (Jupyter Notebook)  [Open in Colab](#)

A red arrow points to the 'Session 2' header, and a red box highlights the 'Open in Colab' button for Session 2.

right click **"Open in Colab"** to open on a new tab



rsession\_mbt2023\_V04.ipynb

File Edit View Insert Runtime Tools Help

+ Code + Text Copy to Drive

UMR MBT Microbiome Praktikum

### Session 2

2.1 Loading libraries and microbiome data

First, we start our R environment by loading the needed packages. For this, we need to install first Bioconductor and then the phyloseq library.

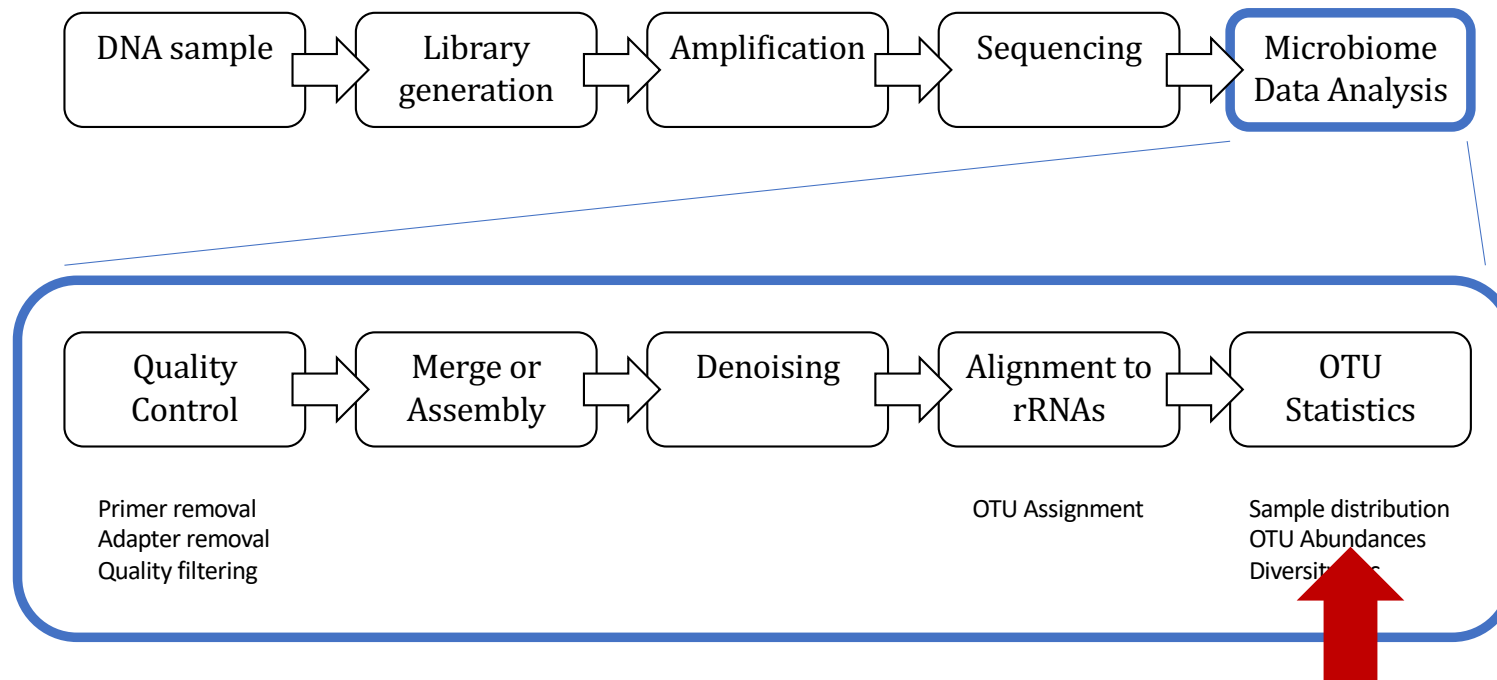
```
[ ] install.packages("BiocManager")
BiocManager::install("phyloseq")
```

Now we load our required libraries:

```
[ ] # load ggplot2 library (graphics)
library(ggplot2)

# loading phyloseq library (microbiome analysis)
library(phyloseq)
```

# Microbiome analysis



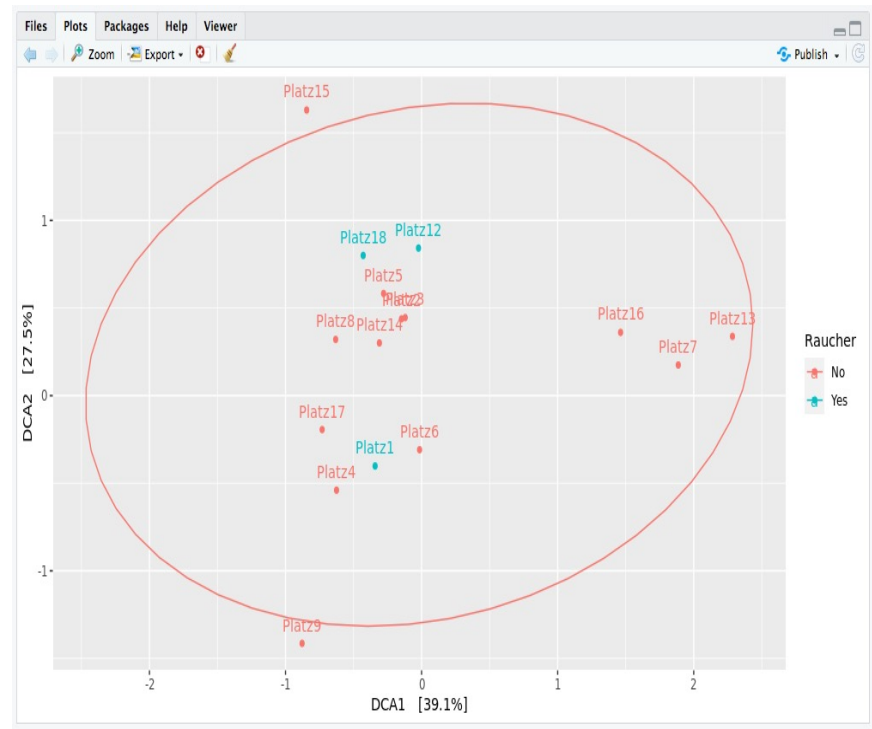
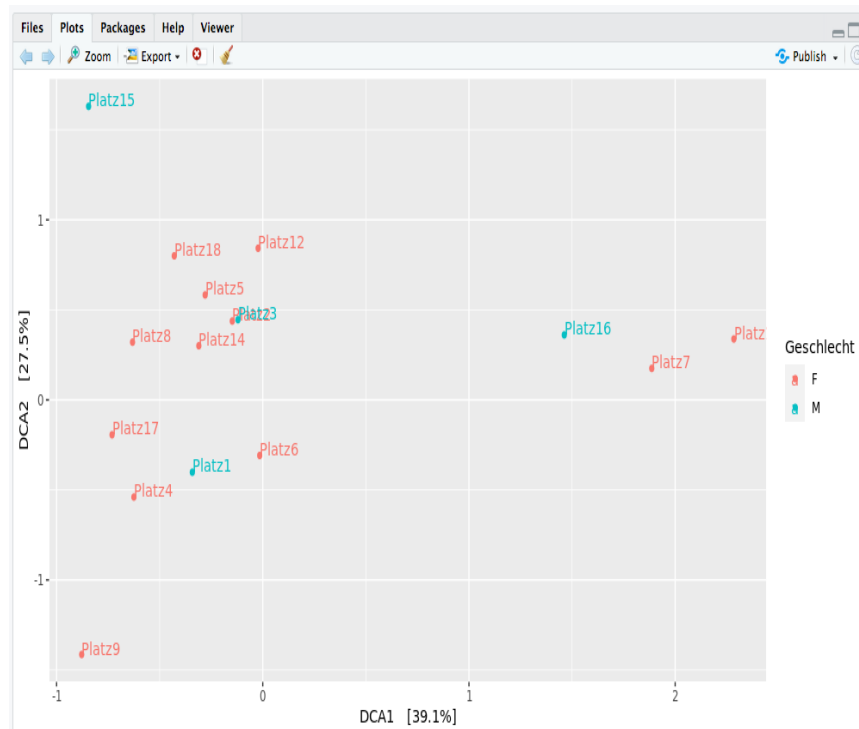
# Data Analysis

- **How do the samples group with each other?**
  - **Principal Components**
- How much diversity is there in a sample?
  - Alpha diversity
- How does the diversity differ between samples?
  - Beta diversity
- **What is the taxonomic content of the sample?**
  - **With reference only to known taxa**
  - Including unknown taxa
- Rarefaction: Adjusting for differences in library sizes across samples to aid comparisons

# Why R for Bioinformatics?

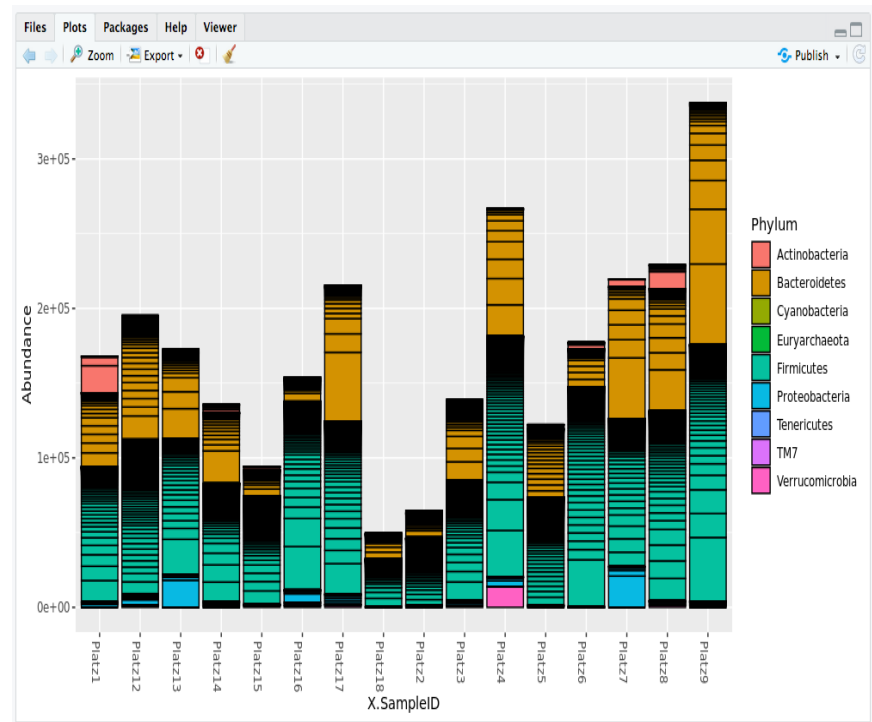
- Freely available
- User- friendly resources to learn R
- R editors and integrated development environments (IDEs)
- Active user community
- Large code repositories: CRAN, Bioconductor; github
- Ready-made packages and functions for bioinformatics
- Scientific reproducibility: Code, library (versions), Rmarkdown, Jupyter notebooks

# Sample Ordination



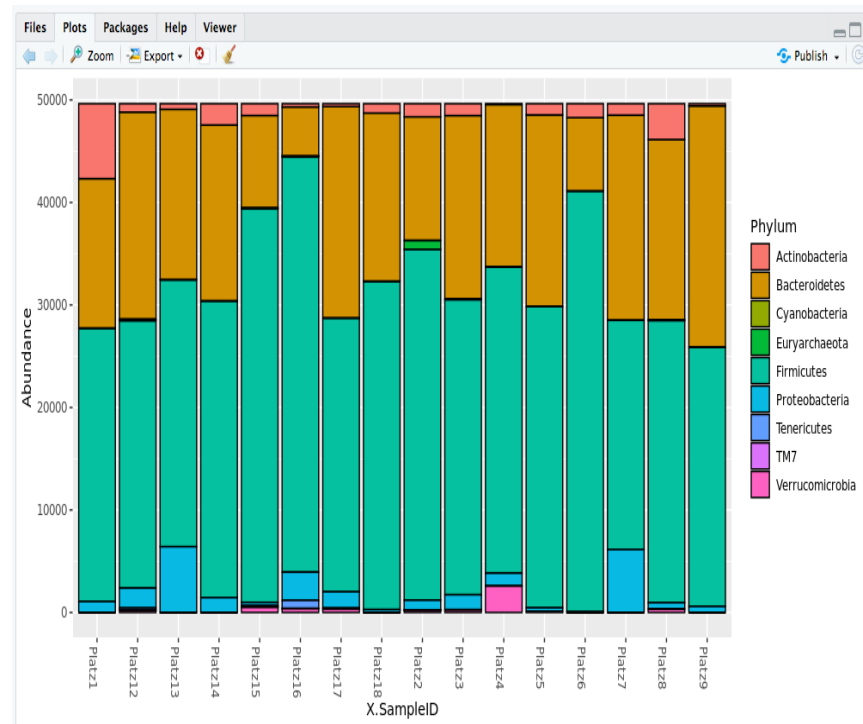
# Microbial communities

```
50
51
52 # Plot abundances
53 plot_bar(psTemp, "X.SampleID", fill="Phylum")
54
55 # Rarefaction to an even depth
56 ps.rarefied = rarefy_even_depth(psTemp)
57
58 # Rarefied abundances
59 plot_bar(ps.rarefied, "X.SampleID", fill="Phylum")
60
61 # Remove lines
62 ps.rarefied.glom <- tax_glom(ps.rarefied, "Phylum")
63 plot_bar(ps.rarefied.glom, "X.SampleID", fill="Phylum")
64
65 # Separate according to metadata
66 plot_bar(ps.rarefied.glom, "X.SampleID", fill="Phylum", facet_grid="Geschlecht")
67 plot_bar(ps.rarefied.glom, "X.SampleID", fill="Phylum", facet_grid="Geschlecht-Raucher")
68
69
70
71
```



# Microbial communities

```
51
52 # Plot abundances
53 plot_bar(psTemp, "X.SampleID", fill="Phylum")
54
55 # Rarefaction to an even depth
56 ps.rarefied = rarefy_even_depth(psTemp)
57
58 # Remove lines
59 ps.rarefied.glom <- tax_glom(ps.rarefied, "Phylum")
60
61 # Plot abundances
62 plot_bar(ps.rarefied.glom, "X.SampleID", fill="Phylum")
63
64 # Separate according to metadata
65 plot_bar(ps.rarefied.glom, "X.SampleID", fill="Phylum", facet_grid="Geschlecht")
66 plot_bar(ps.rarefied.glom, "X.SampleID", fill="Phylum", facet_grid="Geschlecht-Raucher")
67
68
69
70
71
72
63:1 (Top Level) ⚙ R Script ⚙
```

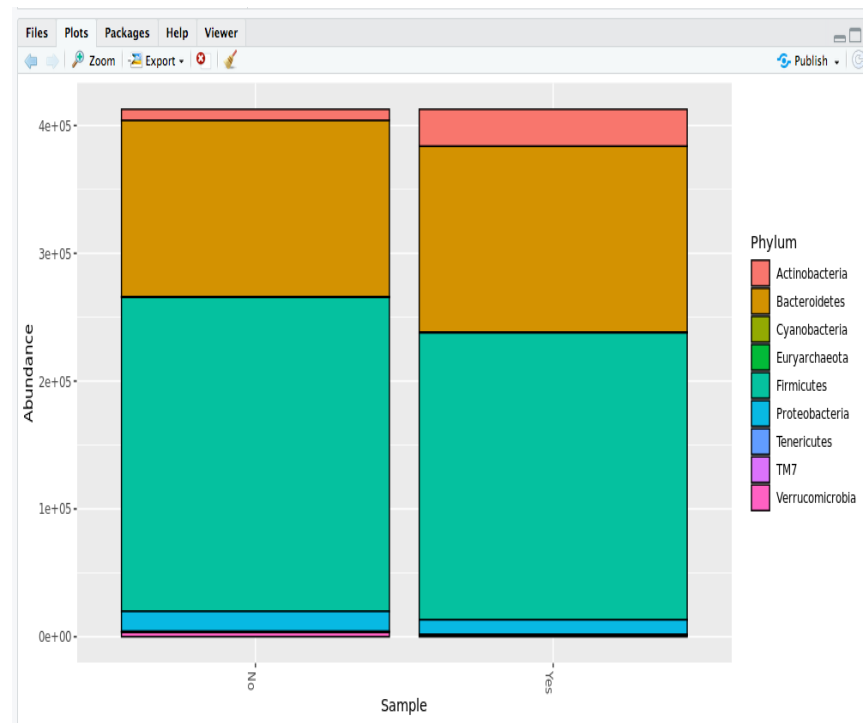


What happens when you try the same at the Genus level (**fill="Genus"**)?



# Microbial communities

```
79  
80 ### Merge samples by a category  
81 mergedGP <- merge_samples(psTemp, "Raucher")  
82  
83 # Rarefaction to an even depth  
84 ps.rarefied = rarefy_even_depth(mergedGP)  
85  
86 # Remove lines  
87 ps.rarefied.glom <- tax_glom(ps.rarefied, "Phylum")  
88  
89 # Plot abundances  
90 plot_bar(ps.rarefied.glom, fill="Phylum")  
91  
92  
93  
94  
95  
96  
97  
98  
99  
100
```



What happens when you try the same for dietary differences (`merge_samples(psTemp, "Ernaehrung")`) ?

# Abundance (composition) plots show:

|  |   |   |
|--|---|---|
| The phyla and genera compositions simultaneously for all samples in the same plot  | T | F |
| The genera (genus) and species composition simultaneously of a microbial community | T | F |
| The abundance of all taxonomic levels at the same time for every individual sample | T | F |
| The abundance of all taxonomic levels at the same time for every sample group      | T | F |
| The composition of a given single taxonomic level for a microbial community        | T | F |

# R for Bioinformatics

|  |   |   |
|--|---|---|
| Requires a commercial licence for use  | T | F |
| Large code repositories  | T | F |
| Jupyter Notebooks are limited to Python and hence cannot be used with the R programming language | T | F |
| Ready-made packages and functions for bioinformatics   | T | F |
| Reproducibility depends on R library versions  | T | F |

# Microbial communities

|  |   |   |
|--|---|---|
| Rarefaction is a method that adjusts for differences in library sizes across samples to aid comparisons      | T | F |
| The principal components method is used to visualize how samples group together                              | T | F |
| Genera composition tells the abundance of species in a sample or group of samples                            | T | F |
| The taxonomic content of a sample (or group) can only be studied using known taxa                            | T | F |
| Alpha diversity measures can be seen as a summary statistic of a single population (within-sample diversity) | T | F |