

# Microbiome Analysis

Israel Barrantes

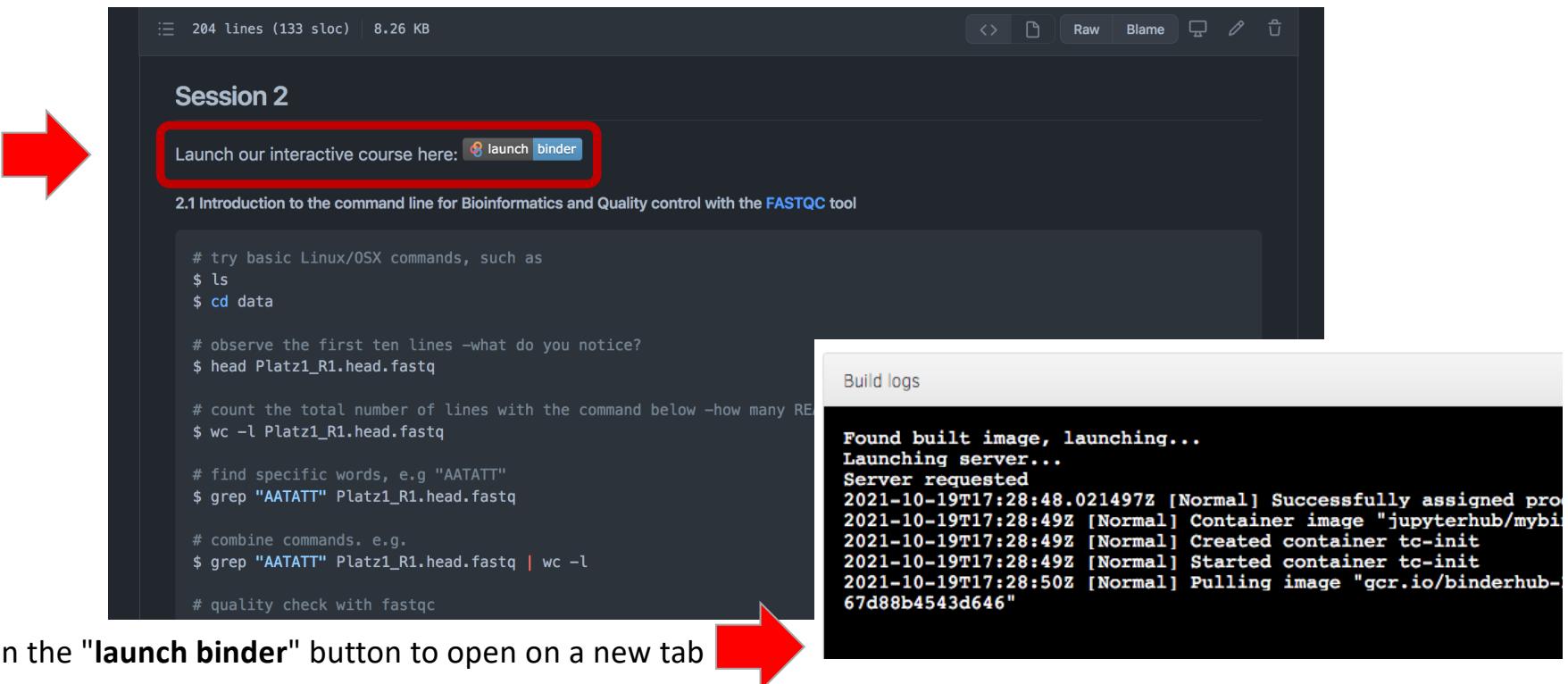
[israel.barrantes@uni-rostock.de](mailto:israel.barrantes@uni-rostock.de)

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

2022-10-12

# Data and Links for Hands-on Practice

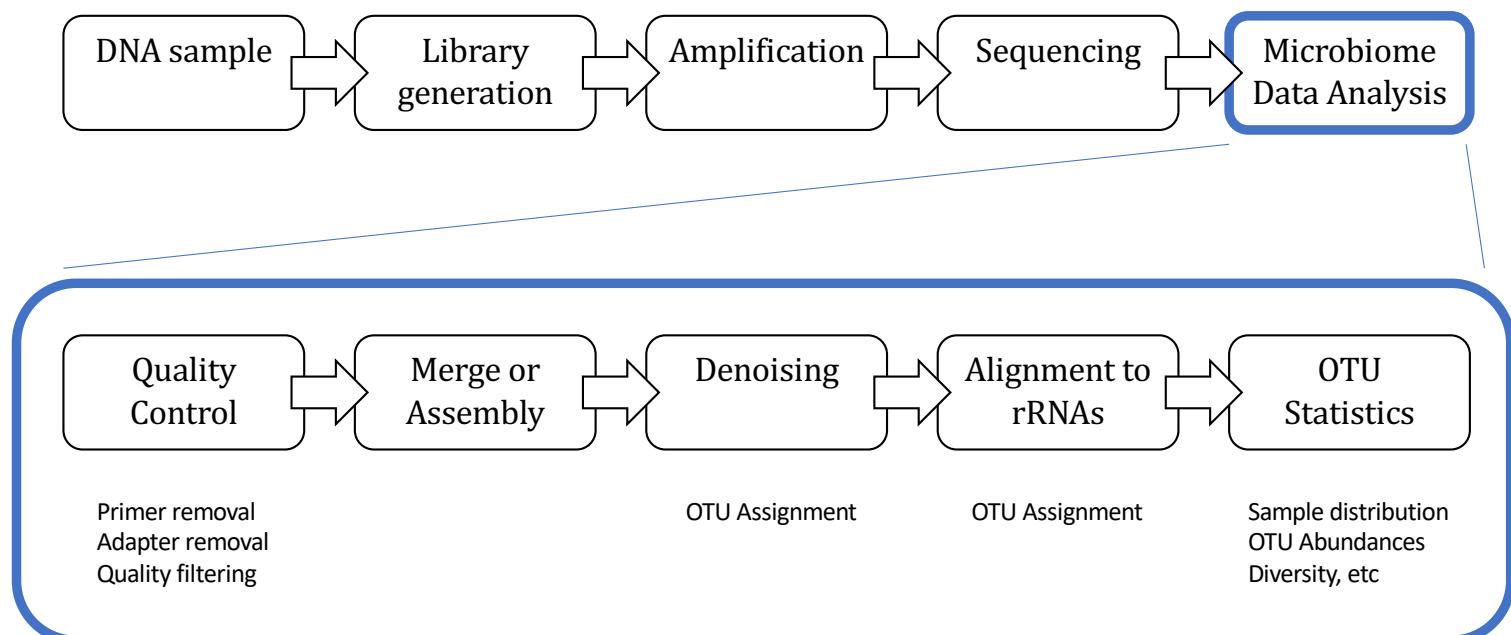
<https://github.com/barrantesisrael/mbt.microbiome.2021>



# 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
>ENSMUSG0000020122|ENSMUST00000138518
CCCTCTTATCATGCTGTCAGTGATCTCTAAATAGCACTCTAACCCCCGTGAACCTGGT
TATAAAAACATGCCAAAGCTGGGAGCCAGGGCTGCAGGAAATACCACAGCCTCAGT
TCATCAAAACAGTTATTGCCAAATGTTCTCAGCTGCAGCTTCATGAGGTAACCTCCA
GGGCCCCACCTGTTCTGGT
>ENSMUSG0000020122|ENSMUST00000125984
GAGTCAGGTTGAAGCTGCCCTGAACACTACAGAGAAGAGAGGCCCTGGTGTCCCTGTC
TCCAGAACCCCAATATGTCCTGTGAAGGGCACACAACCCCTCAAAGGGGTGTCACTTCTT
CTGATCACTTTGTTACTGTTACTAATGATCCTATGAATCACTGTGTCTTCAGAGG
CCGTGAACCACGTCTGCAAT
```

>id, description  
sequence....

## FASTQ

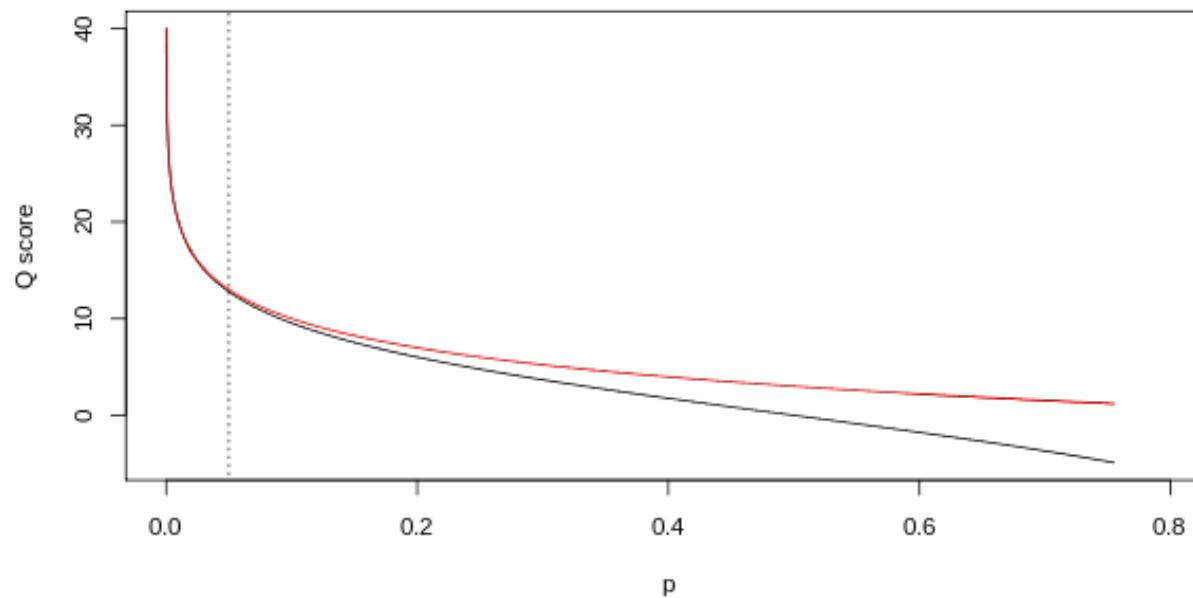
```
@DJB775P1:248:D0MDGACXX:7:1202:12362:49613
TGCTTACTCTGCGTTGATACCACTGCTTAGATCGGAAGAGCACACGTCTGAA
+
JJJJJJIIJJJJJJJIHHHHGHFFFFFFCEEEEDBD?DDDDDDDBDDABDDCA
@DJB775P1:248:D0MDGACXX:7:1202:12782:49716
CTCTGCGTTGATACCACTGCTTACTCTGCGTTGATACCACTGCTTAGATCGG
+
IIIIIIIIIIIIHHHHHHFFFFFFEECCCCBCECCCCCCCCCCCCCCC
```

@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).

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



# Quality Control

Unnamed history

3 shown, 8 deleted

24.84 MB

11: FastQC on data 9: Raw Data

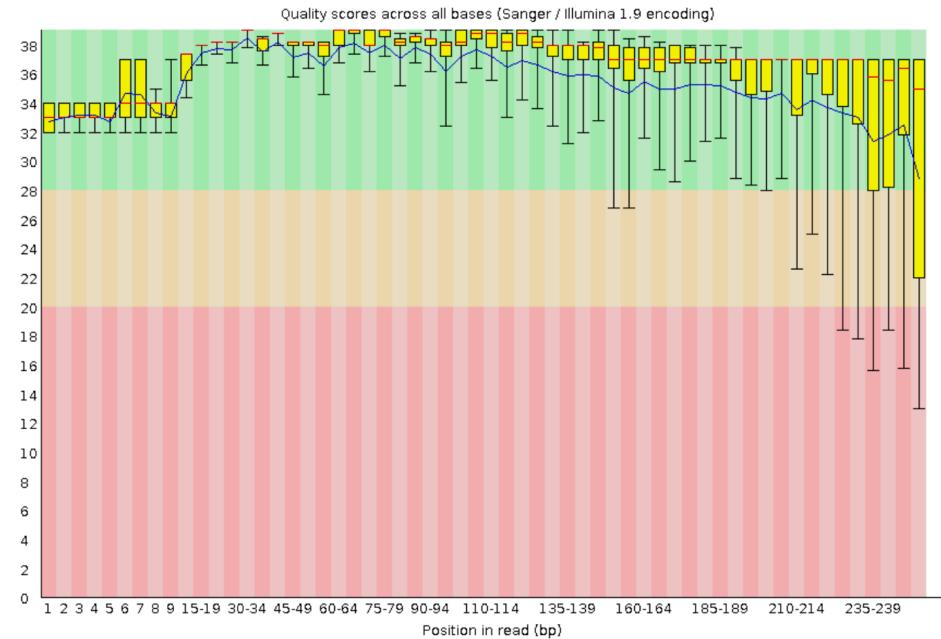
10: FastQC on data 9: Webpage

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



# FASTQ Quality

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

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

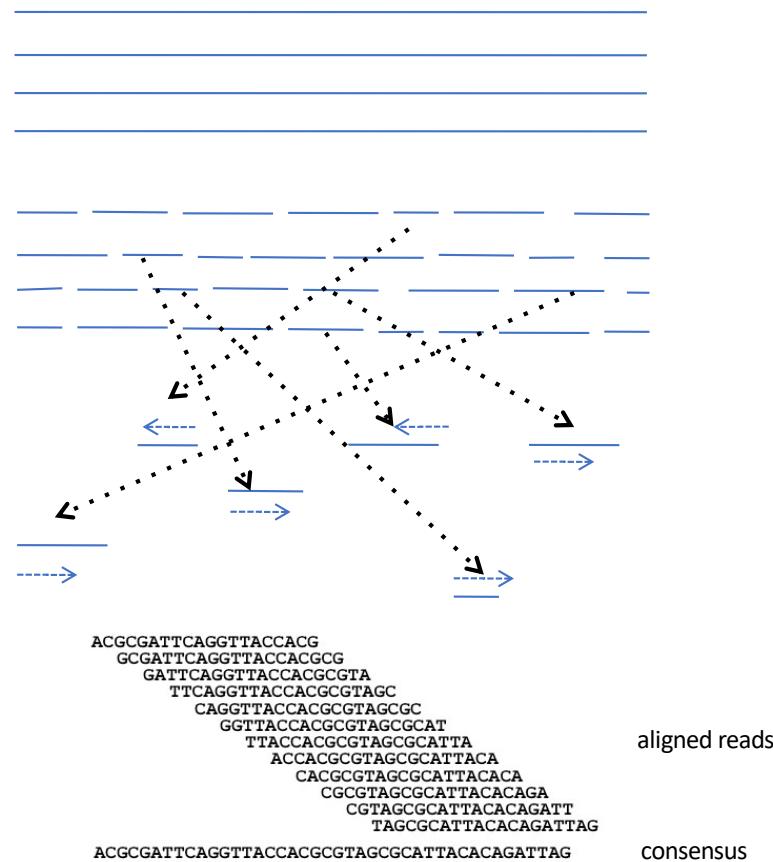
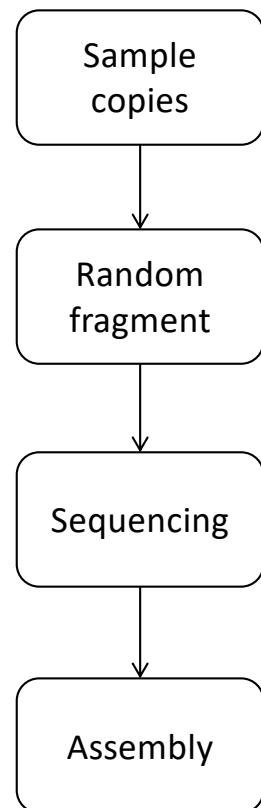
$$Q_{\text{sanger}} = -10 \log_{10} p$$

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

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

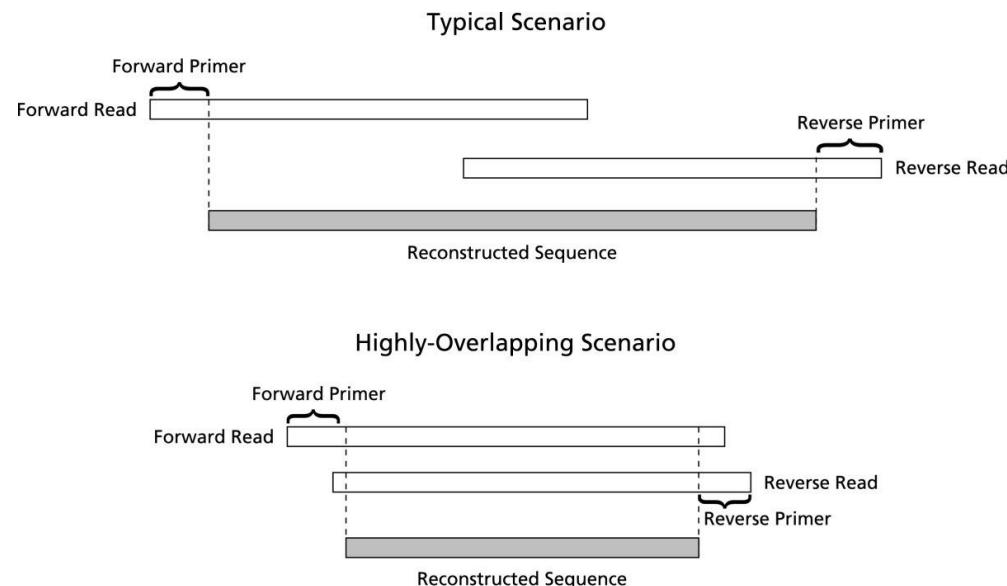
$$p = 10^{-3}$$

# Sequencing and Assembly

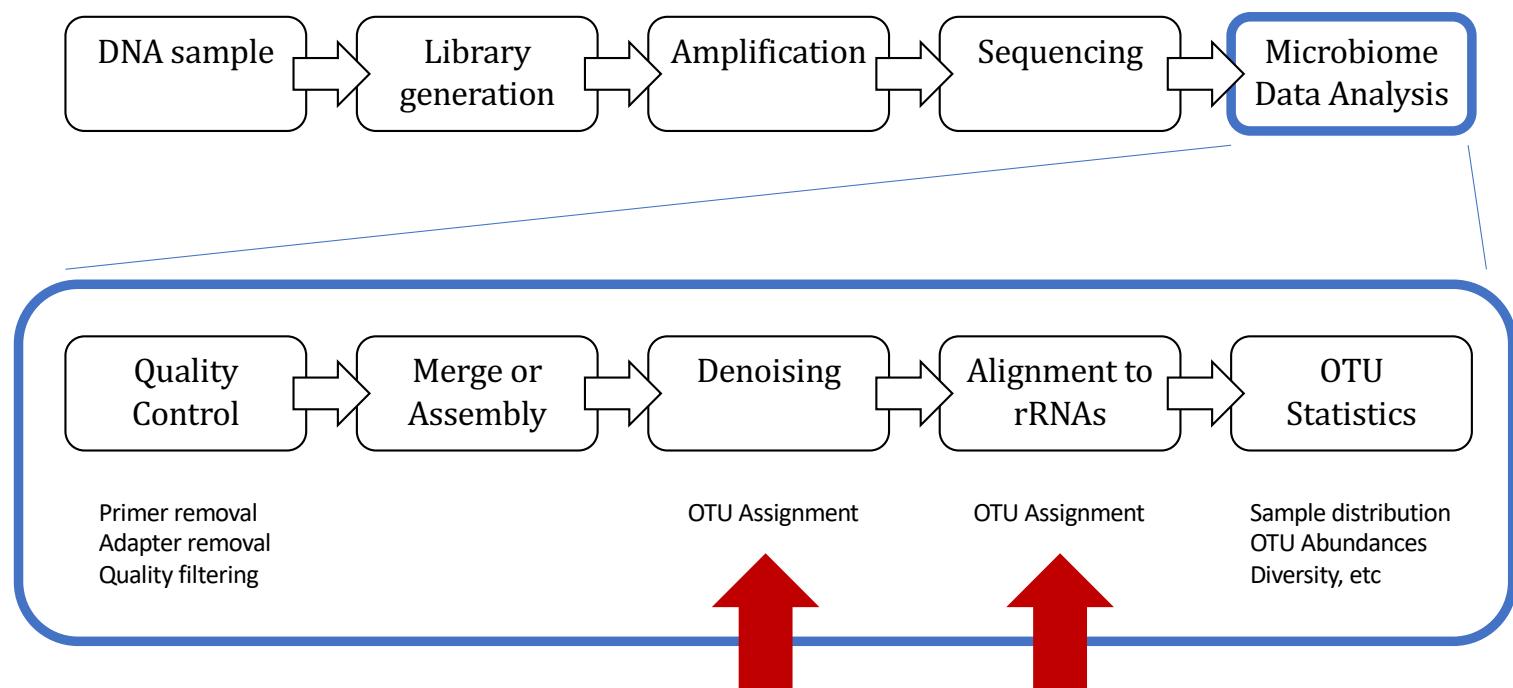


# Assembly: PANDASeq

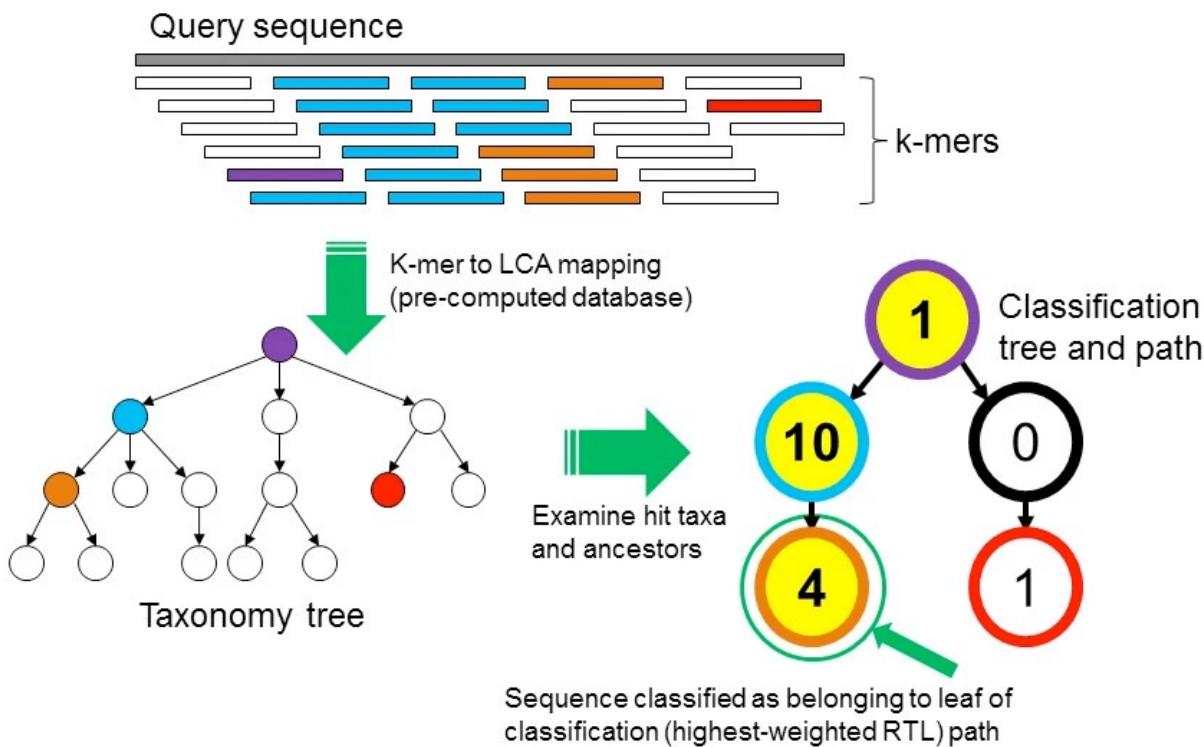
Aligns Illumina reads, and reconstruct an overlapping sequence.



# Microbiome analysis



# Kraken: Taxonomic assignment



# Example OTU Table

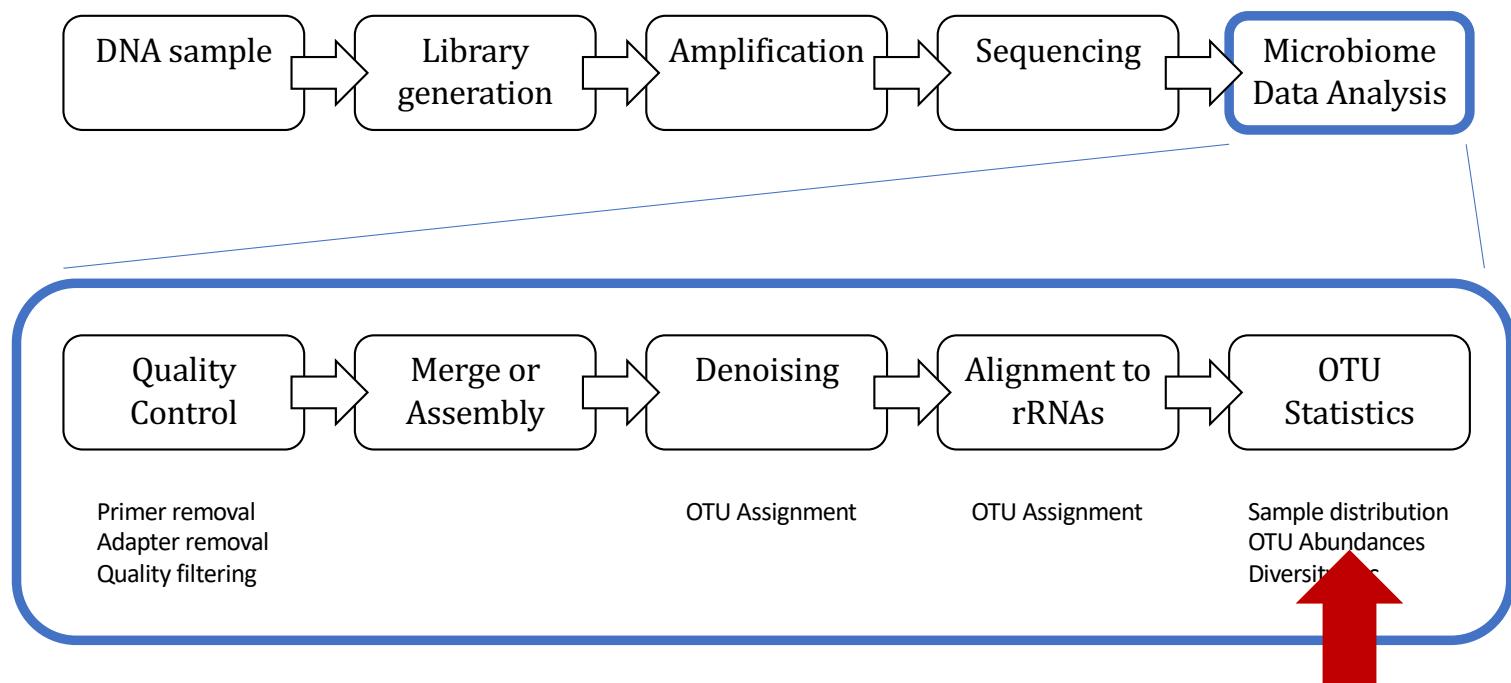
samples

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_Oscillosporales; 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_Oscillosporales; f_Ruminococcaceae; g_Faecalibacterium	
5	437f004ec299d0a2418d69a46f821a57	2930	0	859	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	d_Bacteria; p_Bacteroidata; c_Bacteroidia; o_Bacteroidales; f_Bacteroidaceae; g_Bacteroides	
6	8711aaef804c58e66f4b21a43d3442a	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_Oscillosporales; f_Ruminococcaceae	
7	61227fb9cb3b907c5286c7896659b588	2227	434	1208	0	3566	0	0	0	0	0	0	1975	0	173	0	0	0	0	d_Bacteria; p_Bacteroidata; c_Bacteroidia; o_Bacteroidales; f_Bacteroidaceae; g_Bacteroides	
8	536eeffd9ba4535eab9401214519b2e2	1602	853	1332	328	0	45	0	0	0	775	0	136	0	0	11151	0	0	0	d_Bacteria; p_Bacteroidata; c_Bacteroidia; o_Bacteroidales; f_Bacteroidaceae; g_Bacteroides	
9	9355204b5484253922e933be71ba883	1529	0	0	5	0	14	4	3	0	0	0	14	10	103	32	254	3	20 d_Bacteria; p_Firmicutes; c_Bacilli; o_Lactobacillales; f_Streptococcaceae; g_Lactococcus		
10	588e76106293f91878f35a099321f	1471	0	0	446	0	0	0	0	0	0	0	0	0	0	0	0	0	d_Bacteria; p_Bacteroidata; c_Bacteroidia; o_Bacteroidales; f_Bacteroidaceae; g_Bacteroides		
11	ec51863a70cc9a8abf28e4375e9d0c	1430	231	0	793	295	0	2645	0	49	0	0	2460	0	328	2	341	9055	1 d_Bacteria; p_Firmicutes; c_Clostridia; o_Oscillosporales; f_Ruminococcaceae; g_Ruminococcus; s_Rum		
12	0a9454a4d21f3c586e7463275ca64e39	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_Oscillosporales; f_Ruminococcaceae	
13	c2c379a01bbb7868fc23502d83a61c11	1108	257	690	0	4794	0	0	0	0	0	0	1663	0	0	0	0	0	d_Bacteria; p_Bacteroidata; c_Bacteroidia; o_Bacteroidales; f_Bacteroidaceae; g_Bacteroides		
14	fbd417e4726cbd1feebbb8818c69573	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_Oscillosporales; f_Lachnospiraceae; g_Agathobacter		
15	a72ca0bbc42a577694f3d91fc29594c	917	0	384	0	0	4762	0	0	0	0	0	0	0	1897	843	193	0	d_Bacteria; p_Firmicutes; c_Clostridia; o_Lachnosporales; f_Lachnospiraceae		
16	3520606da0f7747137aeet7e2331cc14f	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_Lachnosporales; f_Lachnospiraceae; g_Blatua		
17	84a0b6767a2fc10ad3fb446712e5b2d	852	264	439	0	0	7	3	0	2450	4	0	0	30	173	2456	109	0	22 d_Bacteria; p_Firmicutes; c_Bacilli; o_Lactobacillales; f_Streptococcaceae; g_Streptococcus		
18	f4d87aedd60165b6e55598bd3631781	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_Lachnosporales; f_Lachnospiraceae		
19	4cc61327d3295ea9ff5159be24059fa9	799	0	735	92	0	0	0	0	387	0	191	2	0	6406	0	0	0	d_Bacteria; p_Bacteroidata; c_Bacteroidia; o_Bacteroidales; f_Bacteroidaceae; g_Bacteroides		
20	ce04abe4a4b4ce23375ae5aa15d2f7	784	168	2623	285	706	395	496	1853	9551	1777	10190	3131	0	58 12957	5225	9953	19141 d_Bacteria; p_Bacteroidata; c_Bacteroidia; o_Bacteroidales; f_Bacteroidaceae; g_Bacteroides			
21	054bf1f28f2ed72af304dc85a31701f	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_Oscillosporales; f_Ruminococcaceae; g_Subdoligranulum		
22	d497525ee3fc2183f2667f648723d9	772	0	3	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	88c037a8f27f07c8b4088511e6a679f	755	0	2164	0	1353	0	0	6477	0	0	0	297	0	363	0	0	0	d_Bacteria; p_Bacteroidata; c_Bacteroidia; o_Bacteroidales; f_Bacteroidaceae; g_Bacteroides		
24	0521f8c2851740732e020a166b972b	545	4	0	702	0	122	0	11961	4300	5847	0	478	0	2456	0	3606	6131	1385 d_Bacteria; p_Bacteroidata; c_Bacteroidia; o_Bacteroidales; f_Bacteroidaceae; g_Bacteroides		
25	7b194de3e2fa5746e5cad8ad54bc92	528	0	0	198	0	0	3206	0	0	0	0	0	0	0	0	98	0	d_Bacteria; p_Firmicutes; c_Clostridia; o_Lachnosporales; f_Lachnospiraceae		
26	42215e4182604699de1301bd8f03fae	523	771	7846	1106	962	0	427	1108	3313	0	0	1613	0	3194	11635	0	5147	0	d_Bacteria; p_Firmicutes; c_Clostridia; o_Lachnosporales; f_Lachnospiraceae; g_Agathobacter	
27	b52844c9e380c80c807948323636d6	515	0	0	0	0	0	0	0	0	0	0	0	0	0	346	0	0	d_Bacteria; p_Firmicutes; c_Clostridia; o_Oscillosporales; f_Ruminococcaceae; g_Faecalibacterium		
28	c114a0220545f73a2362c1072926a48b	497	188	648	415	1478	1347	563	1399	2660	0	958	0	913	0	0	1498	810 d_Bacteria; p_Firmicutes; c_Clostridia; o_Lachnosporales; f_Lachnospiraceae			
29	5+f0d13dalddde3b9261a41b01d7c9	469	0	1247	0	1347	0	374	852	1509	195	0	840	3123	3030	0	922	0	d_Bacteria; p_Firmicutes; c_Clostridia; o_Lachnosporales; f_Lachnospiraceae; g_Blatua		
30	8633d67a4c3f1ceef4244fc05f747f7	451	0	0	2158	0	358	854	0	0	1045	0	2476	0	2884	1377	0	0	d_Bacteria; p_Firmicutes; c_Clostridia; o_Oscillosporales; f_Ruminococcaceae; g_Subdoligranulum		
31	bcd72c9441a61dd12f0e506278fc7ff	446	98	1721	201	948	118	373	1363	5446	737	0	2853	0	25	7342	0	5412	10165 d_Bacteria; p_Bacteroidata; c_Bacteroidia; o_Bacteroidales; f_Bacteroidaceae; g_Bacteroides		
32	fb7b2069972a9a265265453070b7dcd0	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_Lachnosporales; f_Lachnospiraceae; g_Blatua		
33	d51740e49a6fe984e0afda13133c23	414	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	d_Bacteria; p_Bacteroidata; c_Bacteroidia; o_Bacteroidales; f_Prevotellaceae		
34	1e1df519746020d871bd81bdeb523fab4	406	185	270	0	0	2	2	0	1348	1	0	0	0	126	1311	39	0	49 d_Bacteria; p_Firmicutes; c_Bacilli; o_Lactobacillales; f_Streptococcaceae; g_Streptococcus		
35	834c360d747f73dc278b590c3ea213c0	403	0	0	285	111	0	395	0	0	0	0	775	0	613	362	0	0	d_Bacteria; p_Firmicutes; c_Clostridia; o_Oscillosporales; f_Ruminococcaceae; g_Faecalibacterium		
36	c5646bc185f619743b4e34b5cf919ff	384	873	1284	153	949	112	843	227	11224	591	0	0	0	2	0	794	7088	0	d_Bacteria; p_Actinobacterida; c_Actinobacteria; o_Bifidobacteriales; f_Bifidobacteriaceae; g_Bifidobac	
37	d60b3a685294e41bc11f72d0555181f	344	349	0	860	502	449	153	150	4162	885	0	2266	0	0	0	634	1150	0	d_Bacteria; p_Bacteroidata; c_Bacteroidia; o_Bacteroidales; f_Tannerellaceae; g_Parabacteroides; s_Pa	
38	4c437305a61320329c21c218be9cab25	342	0	0	402	0	44	0	8062	2298	3062	0	335	0	1366	0	1885	3105	0	d_Bacteria; p_Bacteroidata; c_Bacteroidia; o_Bacteroidales; f_Bacteroidaceae; g_Bacteroides	
39	357932f2acd1f12289af4b42f932d	333	0	1514	0	1842	0	0	4839	0	0	0	246	0	273	0	0	0	d_Bacteria; p_Bacteroidata; c_Bacteroidia; o_Bacteroidales; f_Bacteroidaceae; g_Bacteroides		
40	4b67dcdbc7b4587b187b1fa71a4942a	301	0	0	76	0	0	0	0	0	0	0	0	0	0	0	0	0	d_Bacteria; p_Bacteroidata; c_Bacteroidia; o_Bacteroidales; f_Bacteroidaceae; g_Bacteroides		
A1	nb08327a7f0322a4da117701n7r05552c	777	22	0	307	438	212	360	21	847	7502	3071	1546	0	0	0	1736	2961	0	d_Bacteria; p_Bacteroidata; c_Bacteroidia; o_Bacteroidales; f_Bacteroidaceae; g_Bacteroides	

see : ~ /data2022/mbtmicrobiome2022.tsv

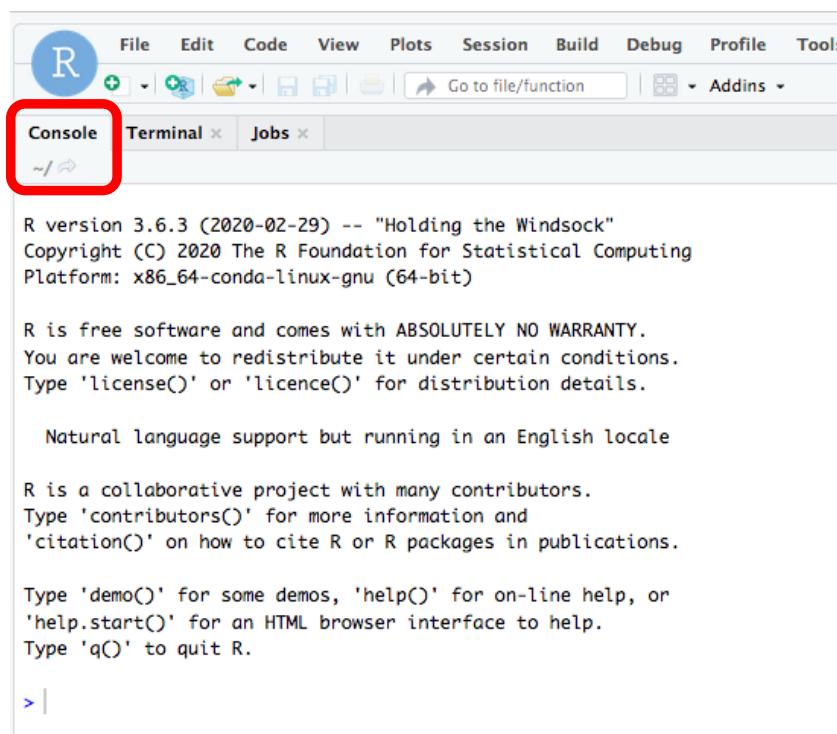
# Microbiome analysis



# Data Analysis

- How do the samples group with each other?
  - Principal Components
- How much diversity is there in the 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

# OTU analysis



The screenshot shows the RStudio interface with the 'Console' tab highlighted by a red box. The R console window displays the standard R startup message, including the version number (R version 3.6.3), copyright information, and a note about the lack of warranty. It also shows the natural language support message and the collaborative nature of the project.

```
R version 3.6.3 (2020-02-29) -- "Holding the Windsock"
Copyright (C) 2020 The R Foundation for Statistical Computing
Platform: x86_64-conda-linux-gnu (64-bit)

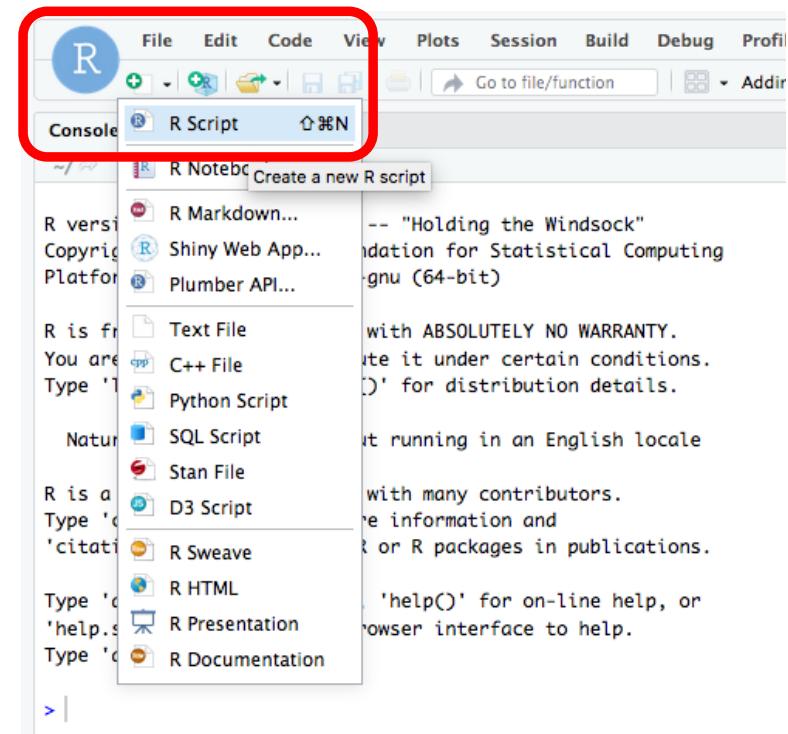
R is free software and comes with ABSOLUTELY NO WARRANTY.
You are welcome to redistribute it under certain conditions.
Type 'license()' or 'licence()' for distribution details.

Natural language support but running in an English locale

R is a collaborative project with many contributors.
Type 'contributors()' for more information and
'citation()' on how to cite R or R packages in publications.

Type 'demo()' for some demos, 'help()' for on-line help, or
'help.start()' for an HTML browser interface to help.
Type 'q()' to quit R.

> |
```



# Packages and Data

The image shows a screenshot of the RStudio interface. On the left, there is a code editor window titled "README.md" containing R code for loading libraries and microbiome data. A red box highlights the copy icon in the toolbar of this window. A green box highlights the "Copied!" message with a checkmark in the status bar. On the right, there is another code editor window titled "Untitled1" containing R code for preparing phyloseq objects. A large red arrow points from the "Copied!" message in the left window towards the "Untitled1" window.

```
# load ggplot2 library (graphics)
library(ggplot2, quietly = TRUE)

# loading phyloseq library (microbiome analysis)
library(phyloseq, quietly = TRUE)

# OTU data
InputBiomFile <- "~/data/mbtmicrobiome2021.biom"

# Samples' metadata
InputMapFile <- "~/data/sample-metadata-2021.tsv"

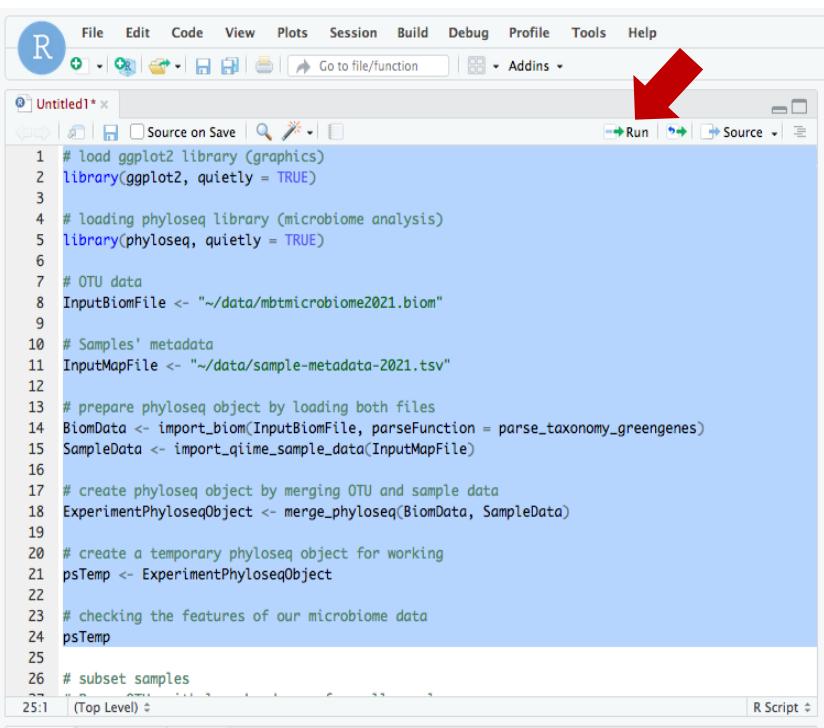
# prepare phyloseq object by loading both files
BiomData <- import_biom(InputBiomFile, parseFunction = parse_taxonomy_greengenes)
SampleData <- import_qiime_sample_data(InputMapFile)

# create phyloseq object by merging OTU and sample data
ExperimentPhyloseqObject <- merge_phyloseq(BiomData, SampleData)

# create a temporary phyloseq object for working
psTemp <- ExperimentPhyloseqObject
```

```
10 # Samples' metadata
11 InputMapFile <- "~/data/sample-metadata-2021.tsv"
12
13 # prepare phyloseq object by loading both files
14 BiomData <- import_biom(InputBiomFile, parseFunction = parse_taxonomy_greengenes)
15 SampleData <- import_qiime_sample_data(InputMapFile)
16
17 # create phyloseq object by merging OTU and sample data
18 ExperimentPhyloseqObject <- merge_phyloseq(BiomData, SampleData)
19
20 # create a temporary phyloseq object for working
21 psTemp <- ExperimentPhyloseqObject
22
23 # checking the features of our microbiome data
24 psTemp
25
26 # subset samples
27 # Prune OTUs with low abundances from all samples
28 psTemp <- prune_taxa(taxa_sums(psTemp) > 100, psTemp)
29
30 # Prune samples with no metadata
31 psTemp <- subset_samples(psTemp, Geschlecht != "ND")
32
33
```

# Load packages and data



```
# load ggplot2 library (graphics)
library(ggplot2, quietly = TRUE)

# loading phyloseq library (microbiome analysis)
library(phyloseq, quietly = TRUE)

# OTU data
InputBiomFile <- "~/data/mbtmicrobiome2021.biom"

# Samples' metadata
InputMapFile <- "~/data/sample-metadata-2021.tsv"

# prepare phyloseq object by loading both files
BiomData <- import_biom(InputBiomFile, parseFunction = parse_taxonomy_greengenes)
SampleData <- import_qiime_sample_data(InputMapFile)

# create phyloseq object by merging OTU and sample data
ExperimentPhyloseqObject <- merge_phyloseq(BiomData, SampleData)

# create a temporary phyloseq object for working
psTemp <- ExperimentPhyloseqObject

# checking the features of our microbiome data
psTemp

# subset samples
```

Console Terminal × Jobs ×

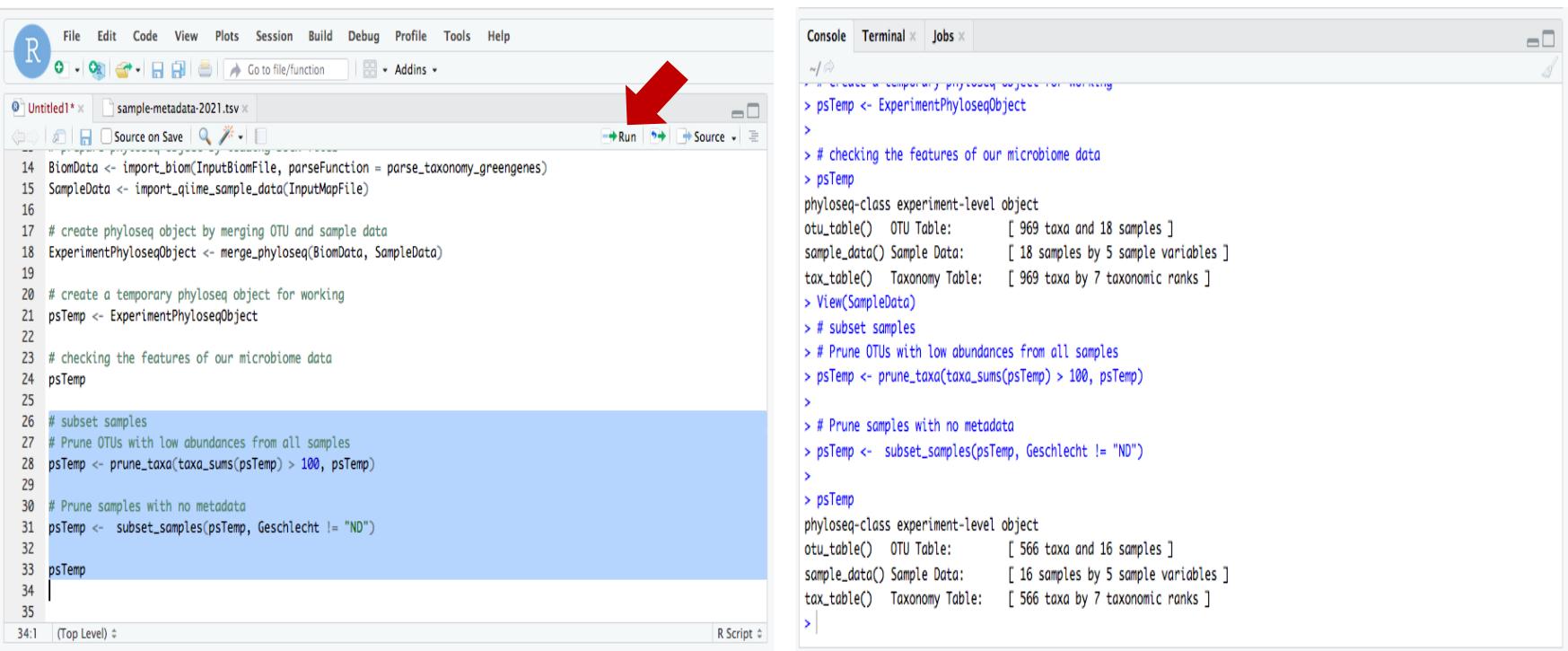
```
> # load ggplot2 library (graphics)
> library(ggplot2, quietly = TRUE)
>
> # loading phyloseq library (microbiome analysis)
> library(phyloseq, quietly = TRUE)
>
> # OTU data
> InputBiomFile <- "~/data/mbtmicrobiome2021.biom"
>
> # Samples' metadata
> InputMapFile <- "~/data/sample-metadata-2021.tsv"
>
> # prepare phyloseq object by loading both files
> BiomData <- import_biom(InputBiomFile, parseFunction = parse_taxonomy_greengenes)
> SampleData <- import_qiime_sample_data(InputMapFile)
>
> # create phyloseq object by merging OTU and sample data
> ExperimentPhyloseqObject <- merge_phyloseq(BiomData, SampleData)
>
> # create a temporary phyloseq object for working
> psTemp <- ExperimentPhyloseqObject
>
> # checking the features of our microbiome data
> psTemp
phyloseq-class experiment-level object
otu_table()  OTU Table:      [ 969 taxa and 18 samples ]
sample_data() Sample Data:    [ 18 samples by 5 sample variables ]
tax_table()   Taxonomy Table: [ 969 taxa by 7 taxonomic ranks ]
> |
```

# Packages and Data: Metadata

The image shows two side-by-side windows. The left window is a file browser with a toolbar at the top and a tree view below. The tree view shows a folder structure under 'data' containing various fastq files (Platz1\_R1.head.fastq, Platz1\_R2.head.fastq, etc.) and a 'sample-metadata-2021.tsv' file, which is highlighted with a red rectangle. The right window is an RStudio-like environment with a toolbar at the top. Below the toolbar, there are tabs for 'Untitled1\*' and 'sample-metadata-2021.tsv'. The main area displays the contents of the 'sample-metadata-2021.tsv' file as a table:

	#SampleID	Geschlecht	Ernaehrung	Raucher	Alter
1	Platz1	M	Standard	Yes	23
2	Platz2	F	Vegetarisch	No	26
3	Platz3	M	Standard	No	24
4	Platz4	F	Standard	No	56
5	Platz5	F	Standard	No	22
6	Platz6	F	Standard	No	23
7	Platz7	F	Standard	No	25
8	Platz8	F	Standard	No	23
9	Platz9	F	Standard	No	22
10	Platz10	ND	ND	ND	ND
11	Platz11	ND	ND	ND	ND
12	Platz12	F	Standard	Yes	26
13	Platz13	F	Vegetarisch	No	26
14	Platz14	F	Vegetarisch	No	27
15	Platz15	M	Standard	No	23
16	Platz16	M	Standard	No	23
17	Platz17	F	Standard	No	35
18	Platz18	F	Standard	Yes	28
19	Platz19	F	Standard	No	48
20					
21					

# Filter low abundant OTUs



The screenshot shows the RStudio interface. On the left is the script editor with an R script titled "sample-metadata-2021.tsv". The script contains R code for importing biom and sample data, creating a phyloseq object, and filtering OTUs based on abundance. A red arrow points to the "Run" button in the toolbar above the script editor. On the right is the R Console window displaying the execution of the script and the resulting phyloseq object information.

```
File Edit Code View Plots Session Build Debug Profile Tools Help
Untitled1* sample-metadata-2021.tsv
Source on Save Go to file/function Addins
Run Source
14 BiomData <- import_biom(InputBiomFile, parseFunction = parse_taxonomy_greengenes)
15 SampleData <- import_qiime_sample_data(InputMapFile)
16
17 # create phyloseq object by merging OTU and sample data
18 ExperimentPhyloseqObject <- merge_phyloseq(BiomData, SampleData)
19
20 # create a temporary phyloseq object for working
21 psTemp <- ExperimentPhyloseqObject
22
23 # checking the features of our microbiome data
24 psTemp
25
26 # subset samples
27 # Prune OTUs with low abundances from all samples
28 psTemp <- prune_taxa(taxa_sums(psTemp) > 100, psTemp)
29
30 # Prune samples with no metadata
31 psTemp <- subset_samples(psTemp, Geschlecht != "ND")
32
33 psTemp
34
35
34.1 (Top Level) R Script
```

```
Console Terminal Jobs
~/ ~ Create a temporary phyloseq object for working
> psTemp <- ExperimentPhyloseqObject
>
> # checking the features of our microbiome data
> psTemp
phyloseq-class experiment-level object
otu_table() OTU Table: [ 969 taxa and 18 samples ]
sample_data() Sample Data: [ 18 samples by 5 sample variables ]
tax_table() Taxonomy Table: [ 969 taxa by 7 taxonomic ranks ]
> View(SampleData)
> # subset samples
> # Prune OTUs with low abundances from all samples
> psTemp <- prune_taxa(taxa_sums(psTemp) > 100, psTemp)
>
> # Prune samples with no metadata
> psTemp <- subset_samples(psTemp, Geschlecht != "ND")
>
> psTemp
phyloseq-class experiment-level object
otu_table() OTU Table: [ 566 taxa and 16 samples ]
sample_data() Sample Data: [ 16 samples by 5 sample variables ]
tax_table() Taxonomy Table: [ 566 taxa by 7 taxonomic ranks ]
>
```

# Sample Ordination

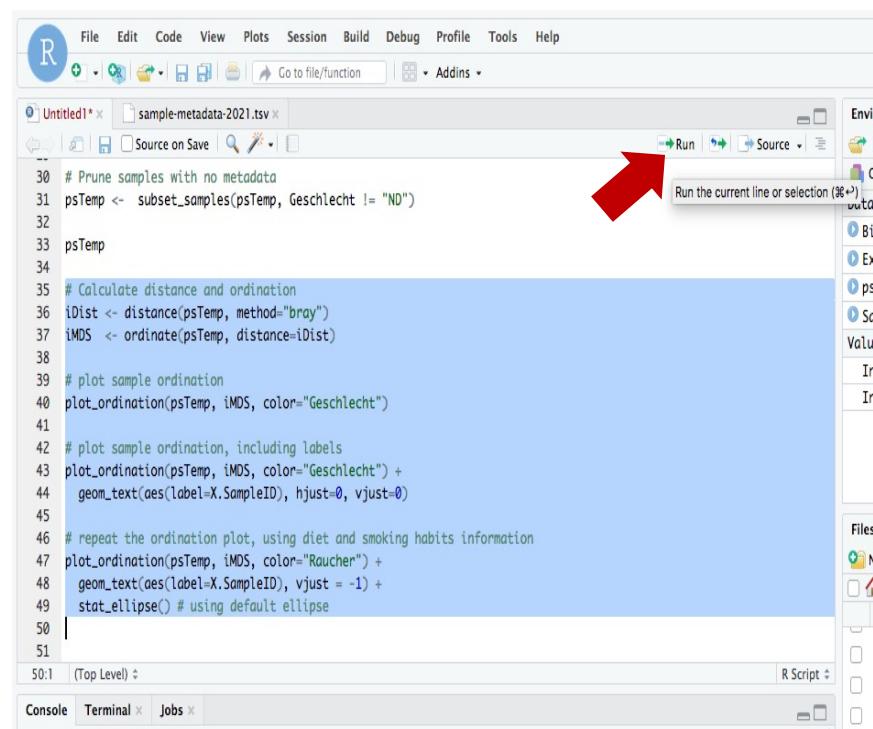
3.2 Sample ordination

```
# Calculate distance and ordination
iDist <- distance(psTemp, method="bray")
iMDS <- ordinate(psTemp, distance=iDist)

# plot sample ordination
plot_ordination(psTemp, iMDS, color="Geschlecht")

# plot sample ordination, including labels
plot_ordination(psTemp, iMDS, color="Geschlecht") +
  geom_text(aes(label=X.SampleID), hjust=0, vjust=0)

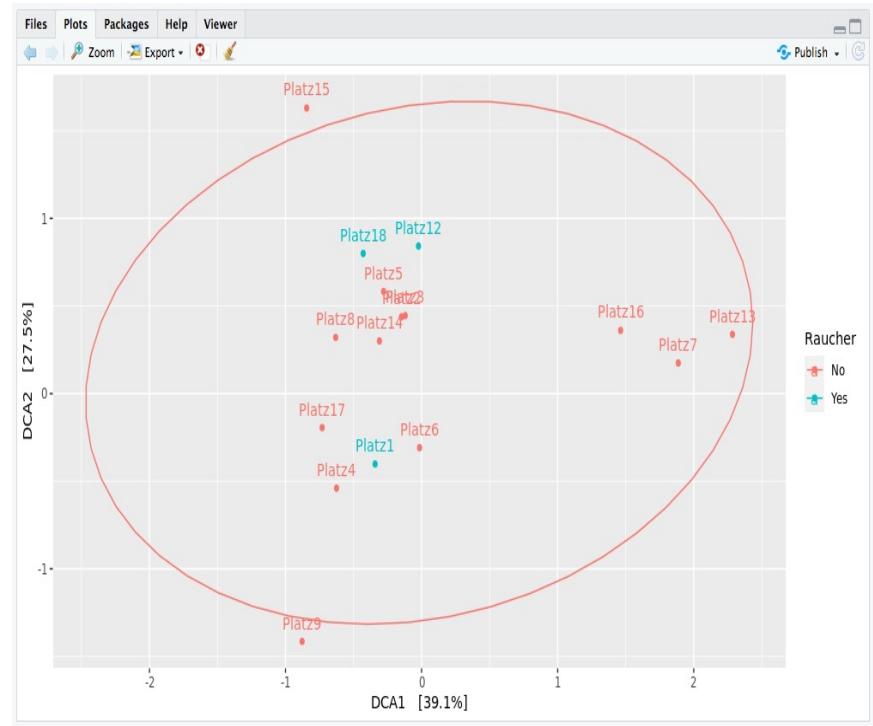
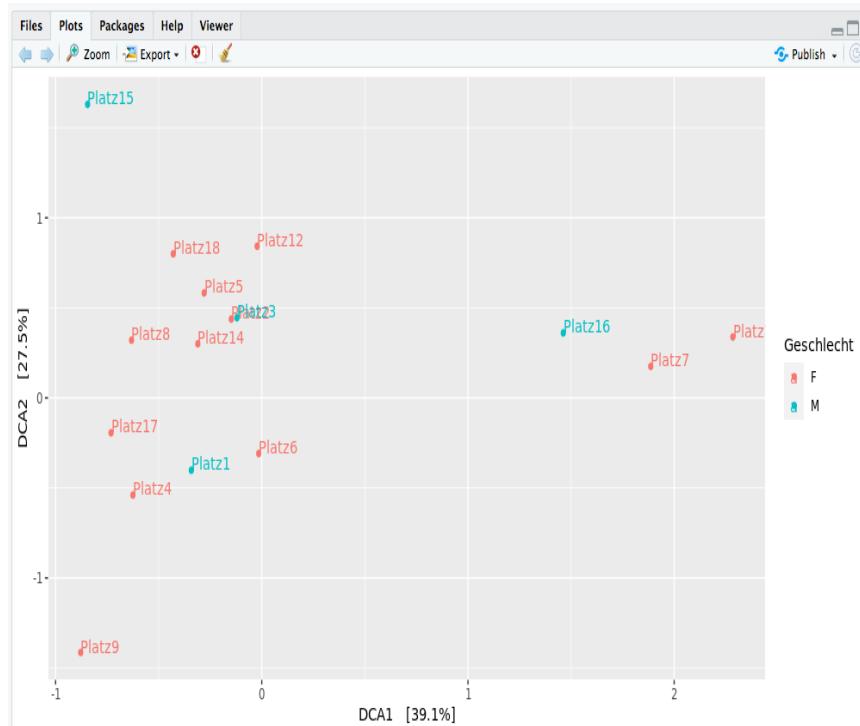
# repeat the ordination plot, using diet and smoking habits information
plot_ordination(psTemp, iMDS, color="Raucher") +
  geom_text(aes(label=X.SampleID), vjust = -1) +
  stat_ellipse() # using default ellipse
```



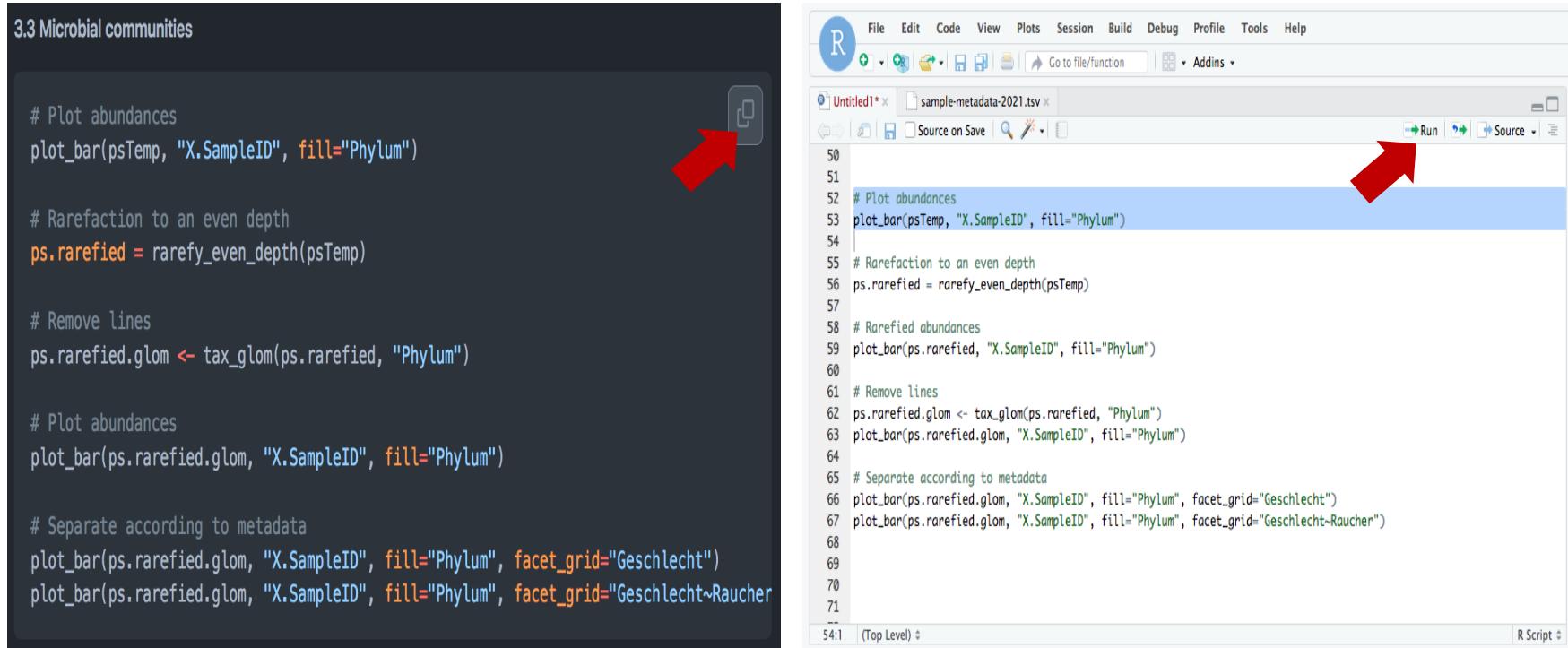
The screenshot shows the RStudio interface. The left pane displays R code for sample ordination, while the right pane shows the R console and terminal. Two red arrows highlight the 'Run' button in the toolbar above the code editor and the toolbar above the terminal, respectively.

```
30 # Prune samples with no metadata
31 psTemp <- subset_samples(psTemp, Geschlecht != "ND")
32
33 psTemp
34
35 # Calculate distance and ordination
36 iDist <- distance(psTemp, method="bray")
37 iMDS <- ordinate(psTemp, distance=iDist)
38
39 # plot sample ordination
40 plot_ordination(psTemp, iMDS, color="Geschlecht")
41
42 # plot sample ordination, including labels
43 plot_ordination(psTemp, iMDS, color="Geschlecht") +
44   geom_text(aes(label=X.SampleID), hjust=0, vjust=0)
45
46 # repeat the ordination plot, using diet and smoking habits information
47 plot_ordination(psTemp, iMDS, color="Raucher") +
48   geom_text(aes(label=X.SampleID), vjust = -1) +
49   stat_ellipse() # using default ellipse
50
51
```

# Sample Ordination



# Microbial communities



The image shows two side-by-side screenshots of RStudio. The left screenshot displays a script titled '3.3 Microbial communities' with the following R code:

```
# Plot abundances
plot_bar(psTemp, "X.SampleID", fill="Phylum")

# Rarefaction to an even depth
ps.rarefied = rarefy_even_depth(psTemp)

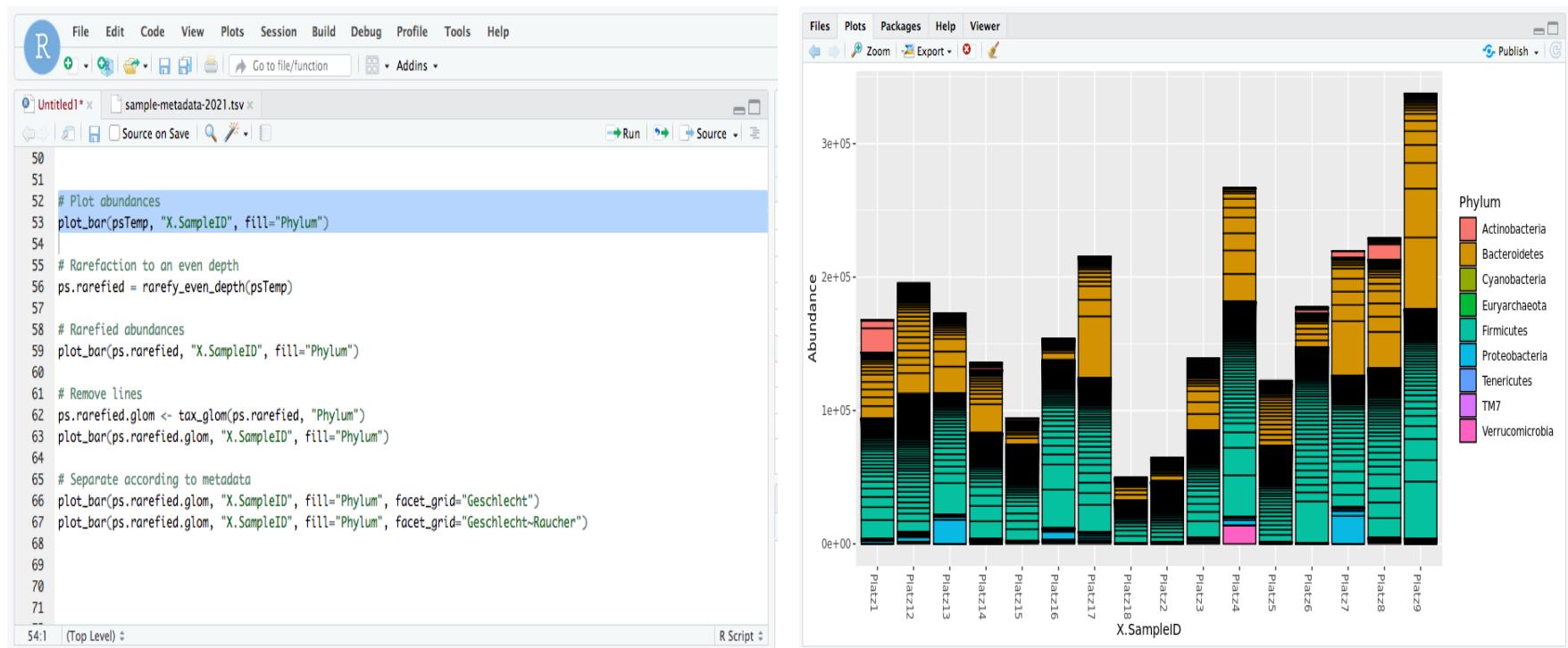
# Remove lines
ps.rarefied.glm <- tax_glm(ps.rarefied, "Phylum")

# Plot abundances
plot_bar(ps.rarefied.glm, "X.SampleID", fill="Phylum")

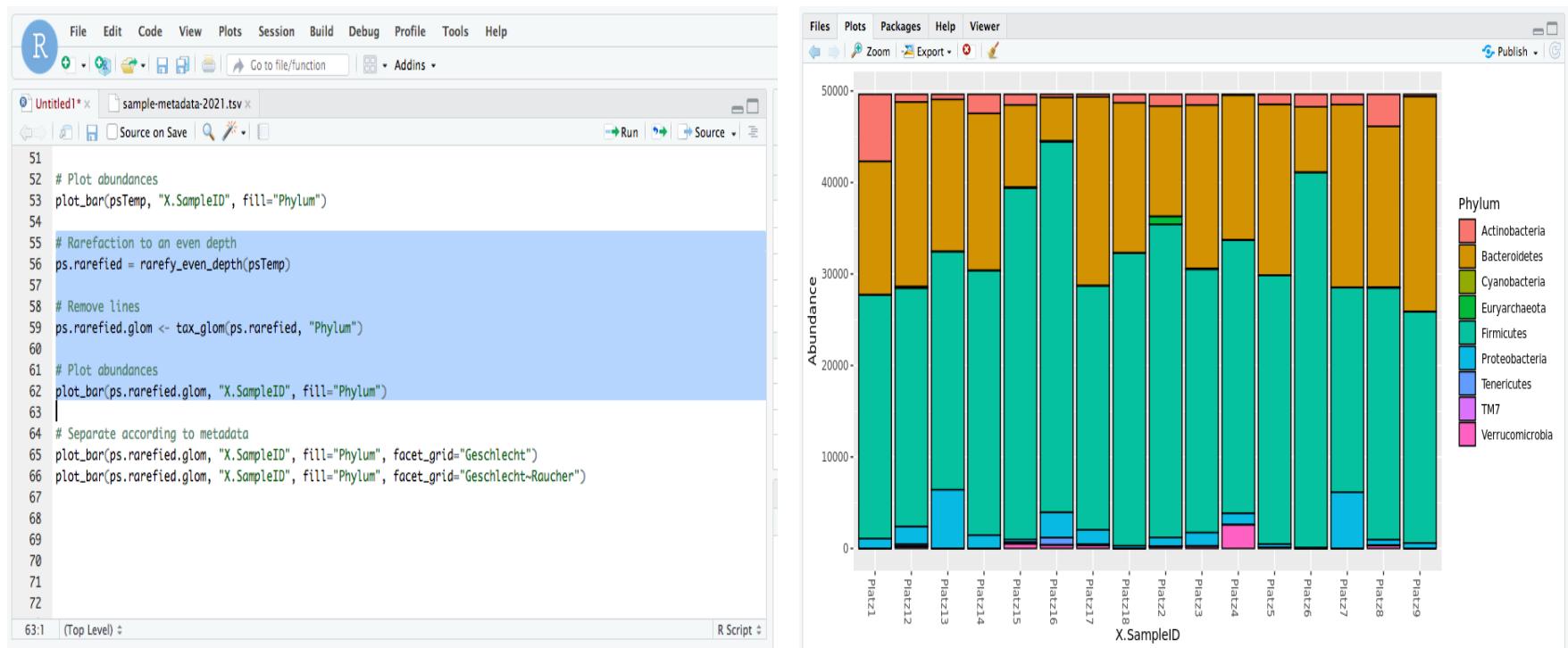
# Separate according to metadata
plot_bar(ps.rarefied.glm, "X.SampleID", fill="Phylum", facet_grid="Geschlecht")
plot_bar(ps.rarefied.glm, "X.SampleID", fill="Phylum", facet_grid="Geschlecht~Raucher")
```

The right screenshot shows the same script in the RStudio interface, with a red arrow pointing to the 'Run' button in the toolbar.

# Microbial communities

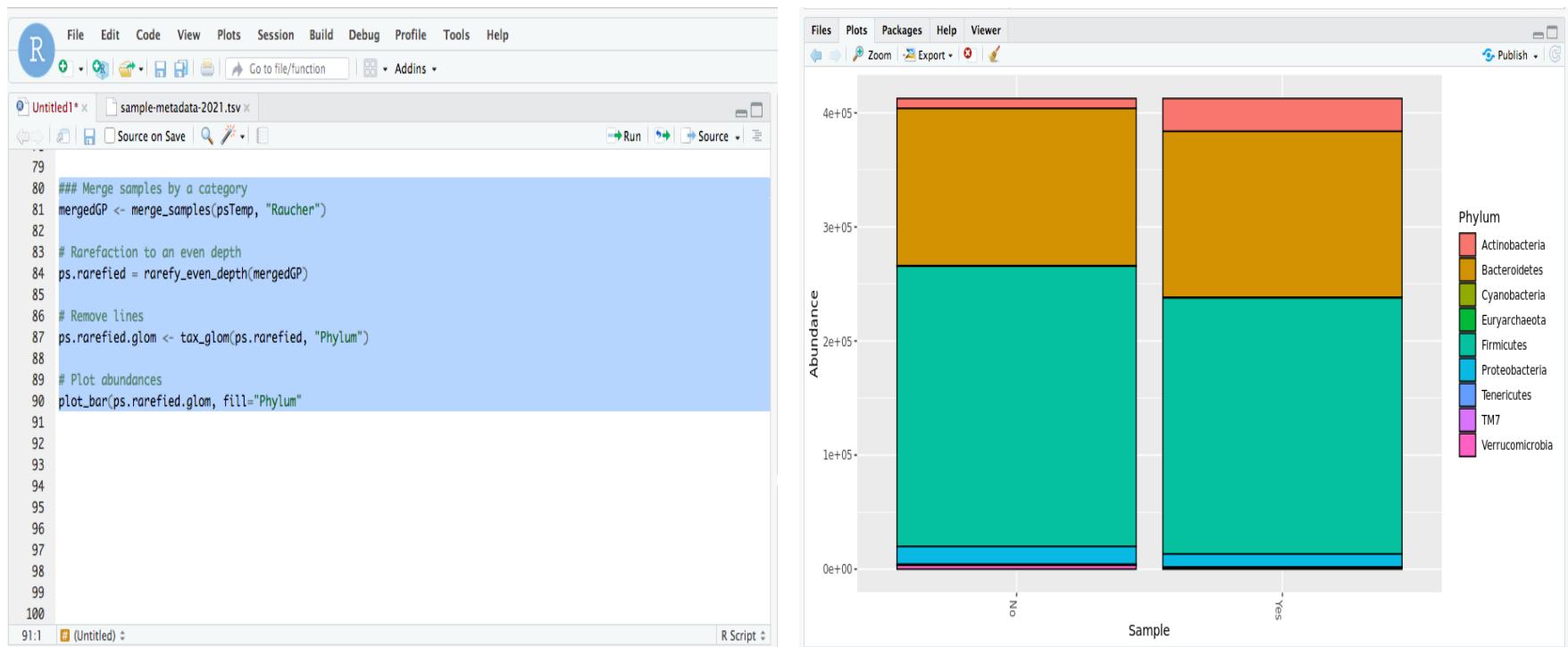


# Microbial communities



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

# Microbial communities



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