

Rhea for Beginner



Klaus Neuhaus (neuhaus@tum.de)
Ilias Lagkouvardos (ilias.lagkouvardos@tum.de)

ZIEL – Institute for Food & Health
Core Facility Microbiome/NGS
Technische Universität München
Freising, Germany

Lagkouvardos I, Fischer S, Kumar N, Clavel T (2017) *Rhea: a transparent and modular R pipeline for microbial profiling based on 16S rRNA gene amplicons*. PeerJ. 11(5):e2836.

Lagkouvardos I, Joseph D, Kapfhammer M, Giritli S, Horn M, Haller D, Clavel T (2016) *IMNGS: A comprehensive open resource of processed 16S rRNA microbial profiles for ecology and diversity studies*. Sci Rep. 23(6):33721.

①

github rhea

Alle Bilder News Videos Shopping Mehr Einstellungen Tools

Ungefähr 348.000 Ergebnisse (0,43 Sekunden)

GitHub - grs/rhea: A reactive messaging library based on the AMQP ...
https://github.com/grs/rhea ▾ Diese Seite übersetzen
var container = require('rhea'); container.on('connection_open', function (context) { context.connection.open_receiver(examples); context.connection.open_sender(examples); }); container.on('message', function (context) { console.log(context.message.body); context.connection.close(); }); container.once('sendable', ...)

GitHub - Lagkouvardos/Rhea: A set of R scripts for the analysis of ...
https://github.com/Lagkouvardos/Rhea ▾ Diese Seite übersetzen
README.md. Rhea Logo. A list of R scripts for the analysis of microbial profiles. Introduction. The importance of 16S rRNA gene amplicon profiles for understanding the influence of microbes in a variety of environments coupled with the steep reduction in sequencing costs led to a surge of microbial sequencing projects.

GitHub - Nocte-/rhea: A constraint solver based on Cassowary

Download Rhea,
install Rstudio

③

④

Rhea by Lagkouvardos

Introduction

The importance of 16S rRNA gene amplicon profiles for understanding the influence of microbes in a variety of environments coupled with the steep reduction in sequencing costs led to a surge of microbial sequencing projects. The expanding crowd of scientists and clinicians wanting to make use of datasets can choose among a range of multipurpose software platforms, the use of which can be intimidating for non-expert users. Among available pipeline options for high-throughput 16S rRNA profile analysis, the R programming language and software environment for statistical computing stands out for its power and increased flexibility, and the possibility to adhere to most recent best practices and to adjust to individual project needs. Here we present the Rhea pipeline, a set of R scripts that encode a series of well-documented choices for the downstream analysis of Operational Taxonomic Units (OTUs) tables, including normalization steps, alpha- and beta-diversity analysis, taxonomic composition, statistical comparisons, and calculation of correlations. Rhea is primarily a straightforward starting point for beginners, but can also be a framework for advanced users who can modify and expand the tool. As the community standards evolve, Rhea will adapt to always represent the current state-of-the-art in microbial profiles analysis in the clear and comprehensive way allowed by the R language.

⑤ Do not forget to download
and install Rstudio as well.

②

③

④

⑤

Rhea by Lagkouvardos

Join GitHub today

A set of R scripts for the analysis of microbial profiles https://lagkouvardos.github.io/Rhea/

124 commits 4 branches 7 releases 4 contributors

Branch: master New pull request

Lagkouvardos Add files via upload Latest commit db8ca2d 21 days ago

- Original-Data/Template-Data Add files via upload 4 months ago
- Normalization Add files via upload 21 days ago
- Alpha-Diversity Rhea Update 2 months ago
- Beta-Diversity Merge branch 'master' of https://github.com/Lagkouvardos/Rhea 2 months ago
- Taxonomic-Binning Group order Boxplot a year ago
- Serial-Group-Comparisons Increase Workspace for fisher Test 24 days ago
- Correlations Fixed output path 28 days ago
- LICENSE Change LICENCE to MIT a year ago
- Overview ReadMe.pdf Add files via upload 9 months ago
- Preparing Input Files ReadMe.pdf Add files via upload a year ago
- README.md Correct titles of README 5 months ago
- install_packages.R install packages a year ago

README.md

Rhea

Unzip Rhea

zipped folders are pink

double click

①

②

Mark all and pull to new folder (e.g., "Rhea")

Perhaps use 7zip to unzip your files. The program is free to download and use. Since sequencing files are often compressed with different compression types (gz, tar, zip), you may need 7zip anyway.

③

Windows unzips while copying

Name	Änderungsdatum	Typ	Große
0.Original-Data	09.02.2018 10:00	Dateiordner	
1.Normalization	09.02.2018 10:00	Dateiordner	
2.Alpha-Diversity	09.02.2018 10:00	Dateiordner	
3.Beta-Diversity	09.02.2018 10:00	Dateiordner	
4.Taxonomic-Binning	09.02.2018 10:00	Dateiordner	
5.Serial-Group-Comparisons	09.02.2018 10:00	Dateiordner	
6.Correlations	09.02.2018 10:00	Dateiordner	
install_packages.R	09.02.2018 10:00	R-Datei	3 KB
LICENSE	09.02.2018 10:00	Datei	2 KB
Overview ReadMe.pdf	09.02.2018 10:00	Adobe Acrobat-Dokument	207 KB
Preparing Input Files ReadMe.pdf	09.02.2018 10:00	Adobe Acrobat-Dokument	206 KB
README.md	09.02.2018 10:00	MD-Datei	9 KB

We recommend to use a new Rhea folder (e.g., Rhea-exp1, Rhea-exp2, etc.) each time you analyse a different experiment. Thus, keep zip-folder as „master back-up“.

①

Name	Änderungsdatum	Typ
0.Original-Data	09.02.2018 10:00	Dateiordner
1.Normalization	09.02.2018 10:00	Dateiordner
2.Alpha-Diversity	09.02.2018 10:00	Dateiordner
3.Beta-Diversity	09.02.2018 10:00	Dateiordner
4.Taxonomic-Binning	09.02.2018 10:00	Dateiordner
5.Serial-Group-Consensus	09.02.2018 10:00	Dateiordner
6.Correlations	09.02.2018 10:00	Dateiordner
install_packages.R	09.02.2018 10:00	R-Datei
LICENSE	09.02.2018 10:00	Datei
Overview ReadMe.pdf	09.02.2018 10:00	Adobe Acrobat-Dokument
Preparing Input Files ReadMe.pdf	09.02.2018 10:00	Adobe Acrobat-Dokument
README.md	09.02.2018 10:00	MD-Datei

Put your original data (e.g., the output of IMNGS) to the folder 0.Original-Data

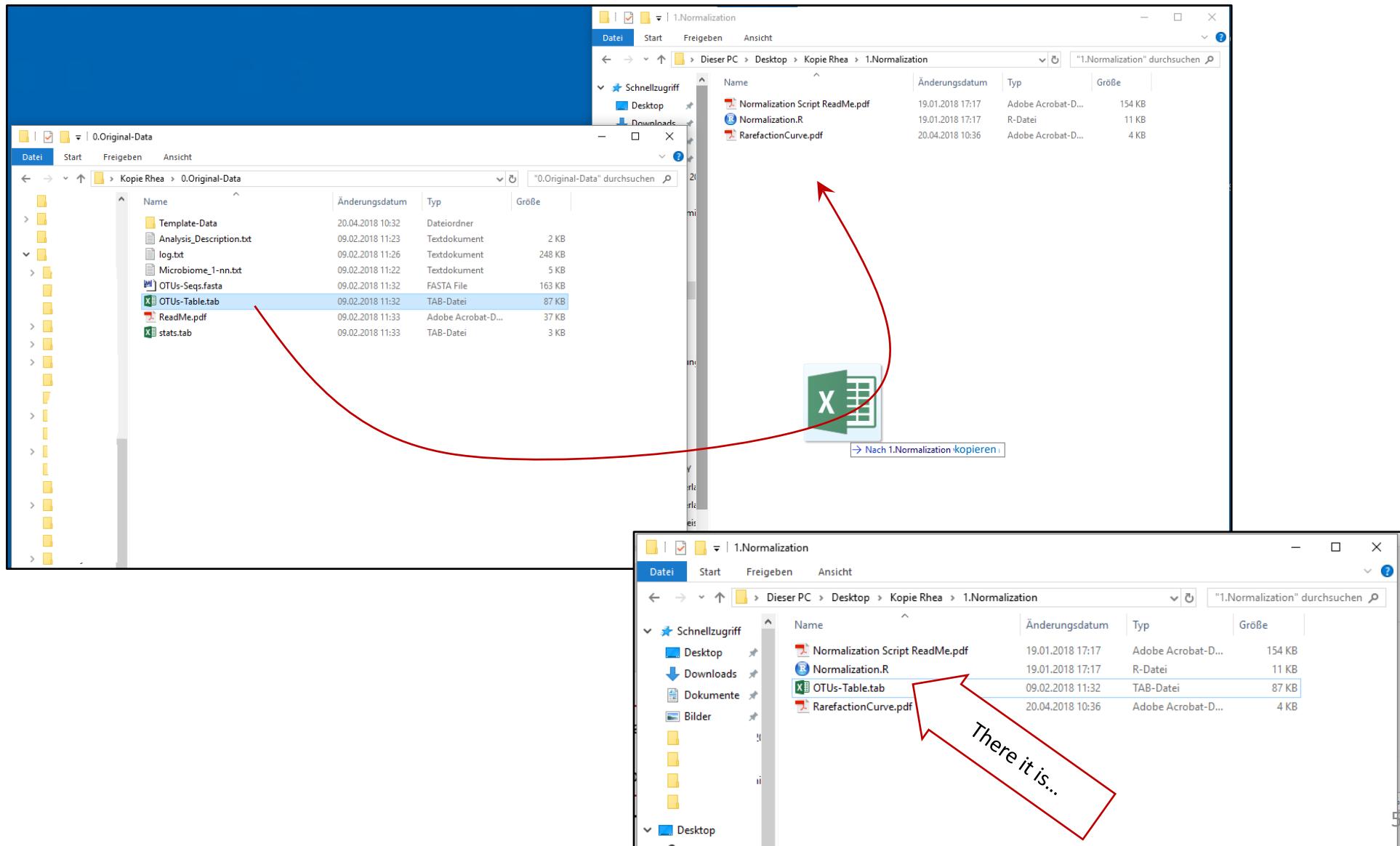
②

③

Next, copy-paste your „OTUs-Table.tab“ from 0.Original-Data to 1.Normalization.

Do not remove it, but copy-paste.

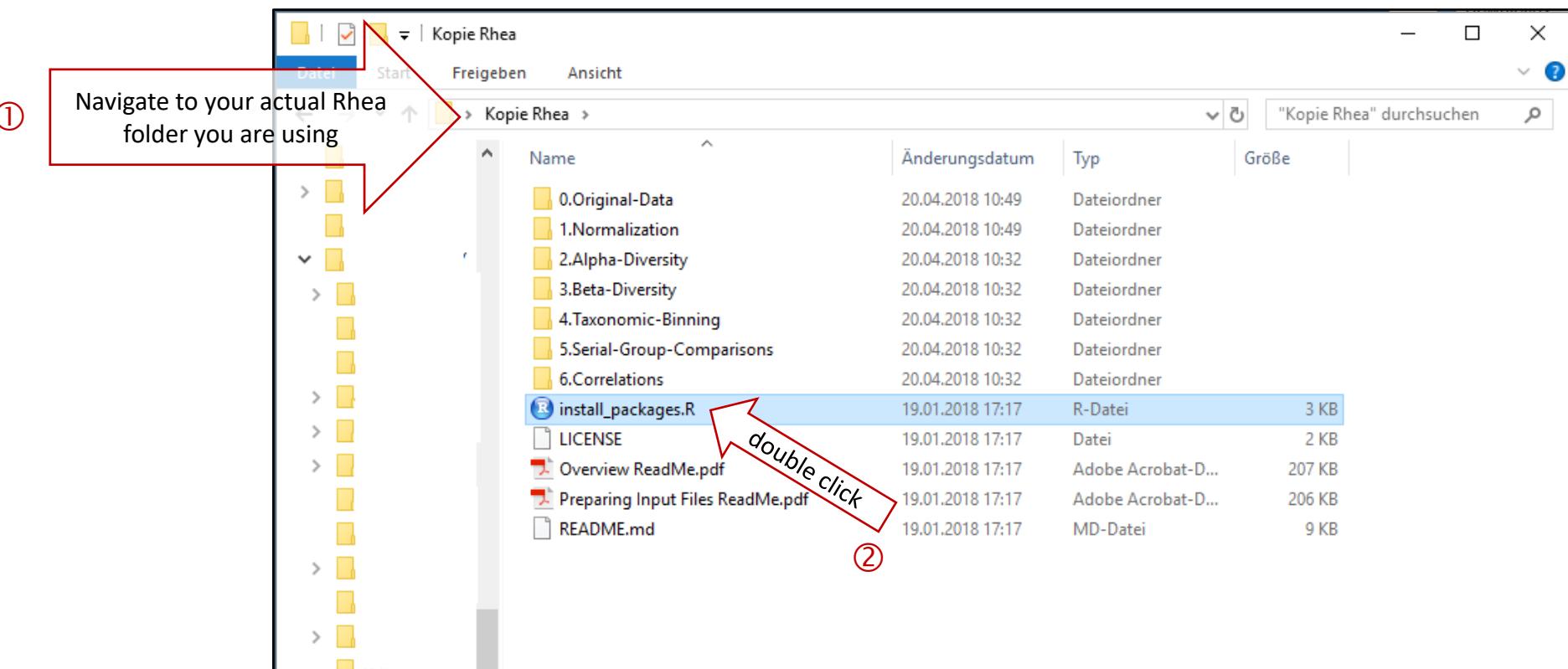
Do not change the data in the folder 0.Original-Data, but keep this as backup.



Before you start with your data analysis, load all packages and libraries necessary for the scripts to run smoothly

①

Navigate to your actual Rhea folder you are using



Some words about Rstudio

When clicking on any R-script (here „install_packages.R“ just as example), RStudio will open and look like this...

The screenshot shows the RStudio interface with several windows open:

- Script window**: Shows the actual R-script code for "install_packages.R". A red box highlights the tab bar where "install_packages.R" is selected. Another red box highlights the script content area.
- Parameter window**: Shows the Global Environment, which is currently empty. A red box highlights this window.
- Files window**: Shows a file tree with various R scripts and PDF files. A red box highlights this window.
- Console window**: Shows the R command-line interface with R version information and help text. A red box highlights this window.

tab shows the active script

Script window – shows the actual R-script code

Parameter window (possibly unimportant for you)

Files used (possibly unimportant for you)

Console window – shows the progress (blue), warnings (red) and errors (red)

How to proceed --- this is similar for all scripts. If in doubt READ THE ReadMe files!!! They are well explained!

The screenshot shows the RStudio interface with the 'install_packages.R' script open in the code editor. The script contains comments and code for installing required packages. A red callout with step 1 points to the section where parameters are set manually. Step 2 points to the pathway to the Rhea folder. Step 3 points to a file explorer window showing the 'Kopie Rhea' folder structure.

① Scroll down until you see this section

② Choose this and insert YOUR ACTUAL pathway to your Rhea folder

③ Goto to your Rhea folder and copy path

RStudio Screenshot:

```
install_packages.R
1 # This script installs all required libraries automatically
2 #' Please install libraries manually if it was not possible to install an library automatically
3 #' Missing libraries are listed in missing_packages.txt
4 #' To install an library please use following two command:
5 #' install.packages("name of the missing library")
6 #' library("name of the missing library")
7
8 #####
9 ##### Set parameters in this section manually #####
10 #####
11 #' Please set the directory of the present script as the working folder
12 #' Note: the path is denoted by forward slash
13 setwd("D:/path/to/rhea")
14
15 #####
16 # NO CHANGES NEEDED BELOW THIS LINE #####
17 #####
18 ##### Scripts #####
19 #####
20 #####
21 #####
22 #####
23 # check if required packages are already installed, and if not, install them
24 packages <- c("ade4", "unifrac", "phangorn", "cluster",
25             "rpart", "compare", "plotrix", "PerformanceAnalytics", "fossil",
26             "gtable", "Matrix", "cplm", "Hmisc", "corrr", "mustat")
27
28 # Function to check whether the package is installed
29 Inspack <- function(pack)
30 {
31   if (!pack %in% installed.packages()) == FALSE) {
32     install.packages(pack, repos = "http://cloud.r-project.org/")
33   }
34 }
35
36
37
38 # Applying the installation on the list of packages
39 lapply(packages, Inspack)
40
41 # Make the libraries
42 lib <- lapply(packages, require, character.only = TRUE)
43 not_installed <- which(lib==FALSE)
44 missing_packages <- lapply(not_installed, function(x) print(packages[x]))
45
46
47 # Adding log file analysis
48 sink(file = "missing_packages.txt")
49 cat("*****\n")
50 cat ("Please install following packages manually", "\n")
51 cat (as.character(missing_packages))
52 sink()
53
```

Global Environment Tab:

Environment History Connections

Global Environment

Environment is empty

File Explorer Screenshot:

Name	Änderungsdatum	Typ	Größe
0.Original-Data	20.04.2018 10:49	Datei	3 KB
1.Normalization	20.04.2018 11:01	Datei	2 KB
2.Alpha-Diversity	20.04.2018 10:32	Datei	207 KB
3.Beta-Diversity	20.04.2018 10:32	Datei	206 KB
4.Taxonomic-Binning	20.04.2018 10:32	Datei	9 KB
5.Serial-Group-Comparisons	20.04.2018 10:32	Datei	
6.Correlations	20.04.2018 10:32	Datei	
install_packages.R	19.01.2018 17:17	R-Datei	
LICENSE	19.01.2018 17:17	Datei	
Overview ReadMe.pdf	19.01.2018 17:17	Adobe Acrobat-Dokument	
Preparing Input Files ReadMe.pdf	19.01.2018 17:17	Adobe Acrobat-Dokument	
README.md	19.01.2018 17:17	MD-Datei	

insert correct path

```
5 ##### Set parameters in this section mandatory
10 ######
11
12 #' Please set the directory of the present script as the wo
13 #' Note: the path is denoted by forward slash "/"
14 setwd("D:/path/to/Rhea")  
mark default path ①
```

```
17 #####
18 #####
19 #####
20 ##### NO CHANGES ARE NEEDED BELOW THIS LINE
```

```
10 #####
11
12 #' Please set the directory of the present script as the wo
13 #' Note: the path is denoted by forward slash "/"
14 setwd("C:\Users\User\Desktop\Kopie Rhea")  
② paste the correct path
```

```
17 #####
18 #####
19 #####
20 ##### NO CHANGES ARE NEEDED BELOW THIS LINE
```

Main Script

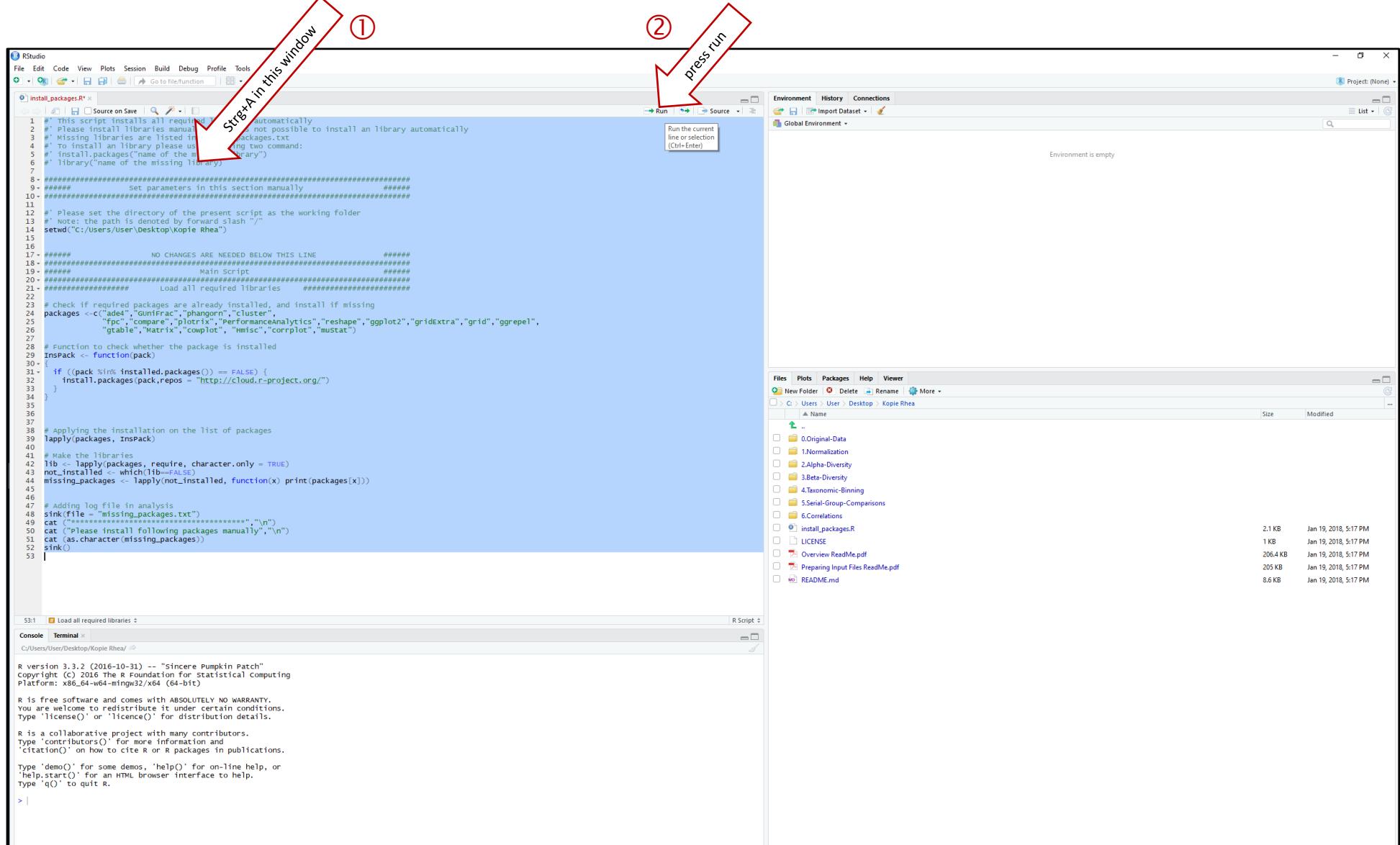
start to change \ to / ③

```
10 #####
11
12 #' Please set the directory of the present script as the wo
13 #' Note: the path is denoted by forward slash "/"
14 setwd("C:/users/User/Desktop/Kopie Rhea")  
15
16
```

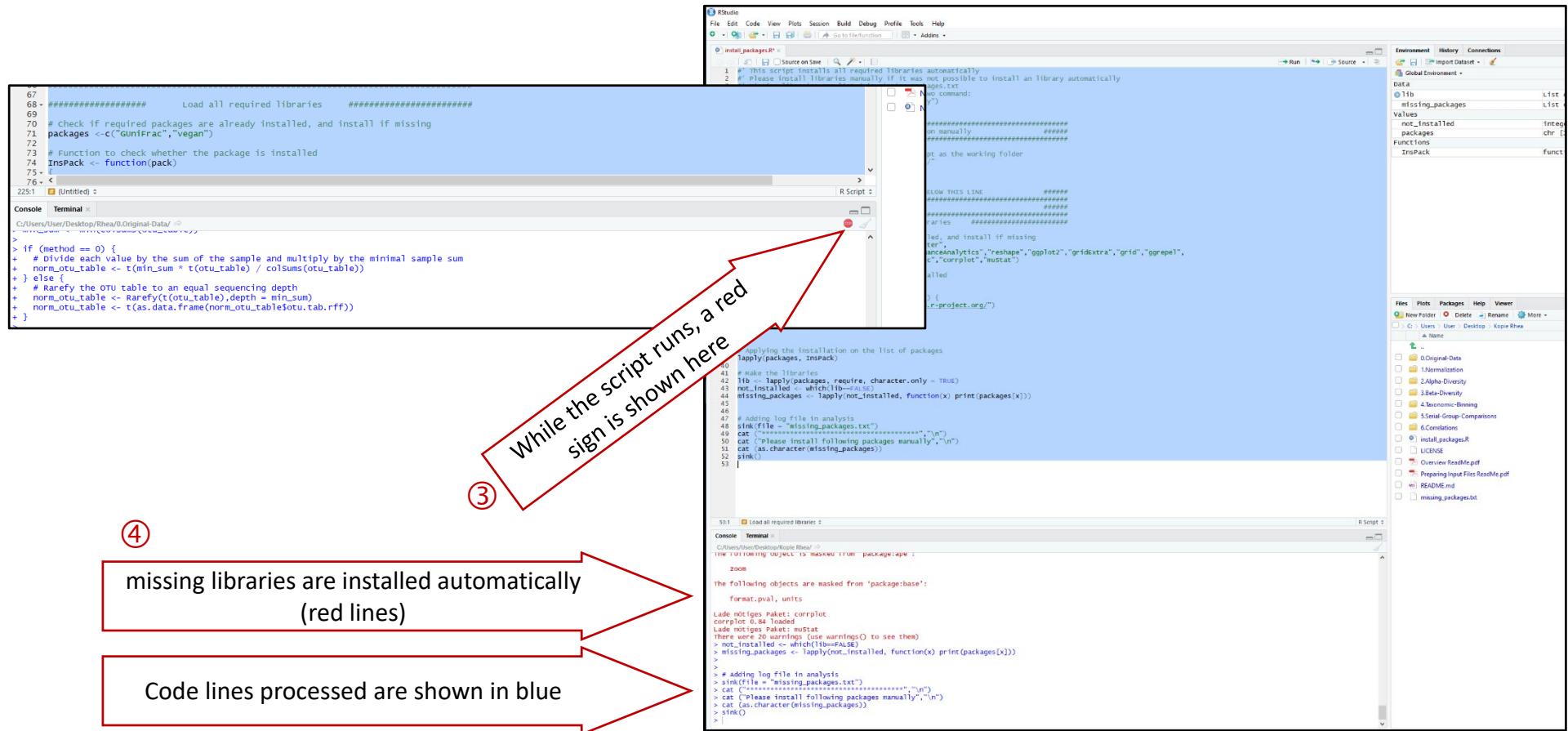
```
17 #####
18 #####
19 #####
20 ##### NO CHANGES ARE NEEDED BELOW THIS LINE
```

Main Script

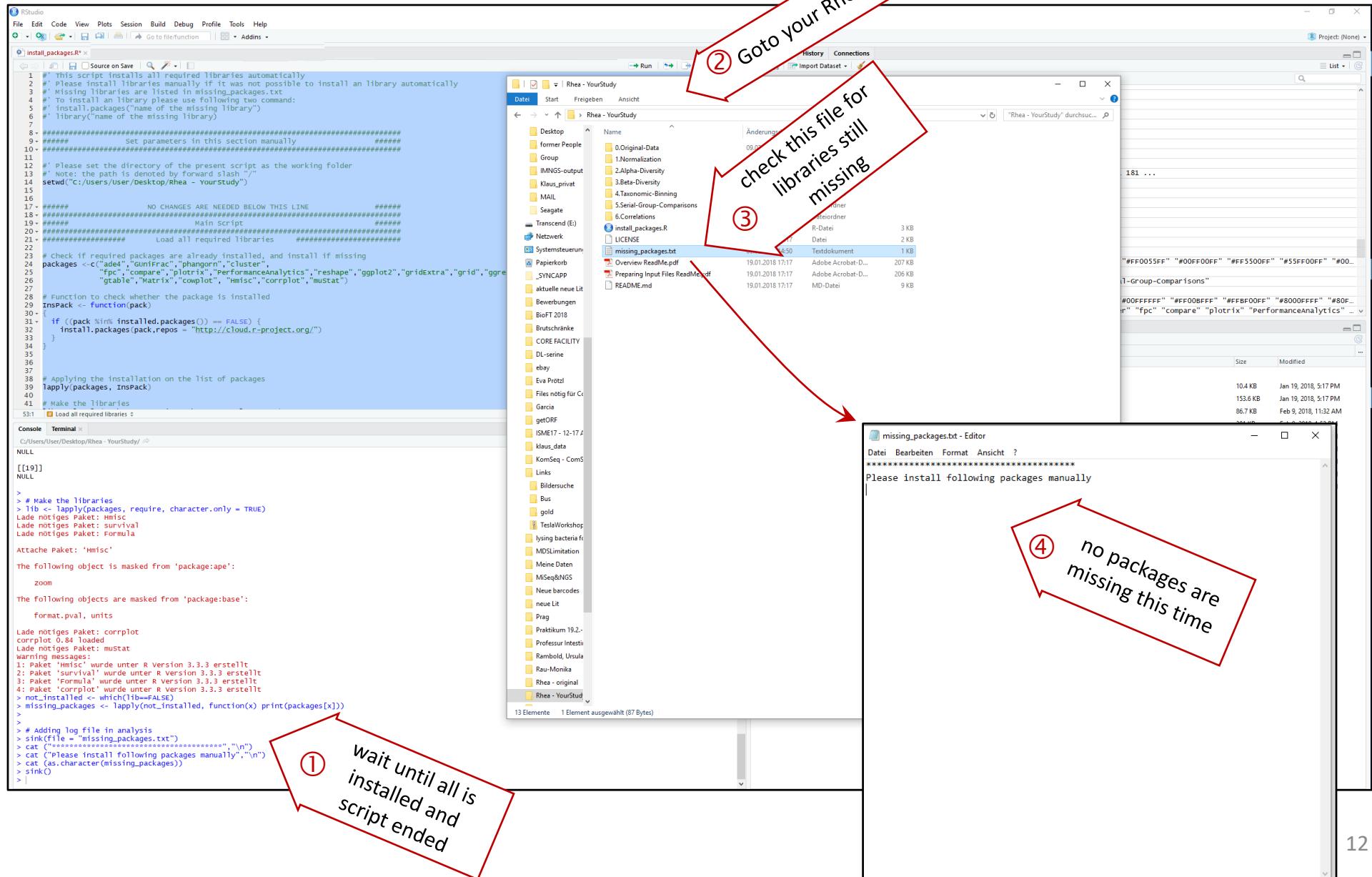
To run the script, mark all (Strg+A) code lines in the script window (the marked lines turn blue) and press „Run“



The script runs ...

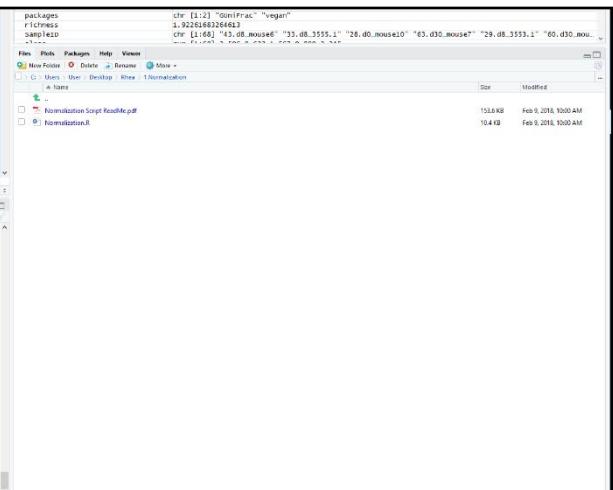


„install_packages.R“ is finished ...



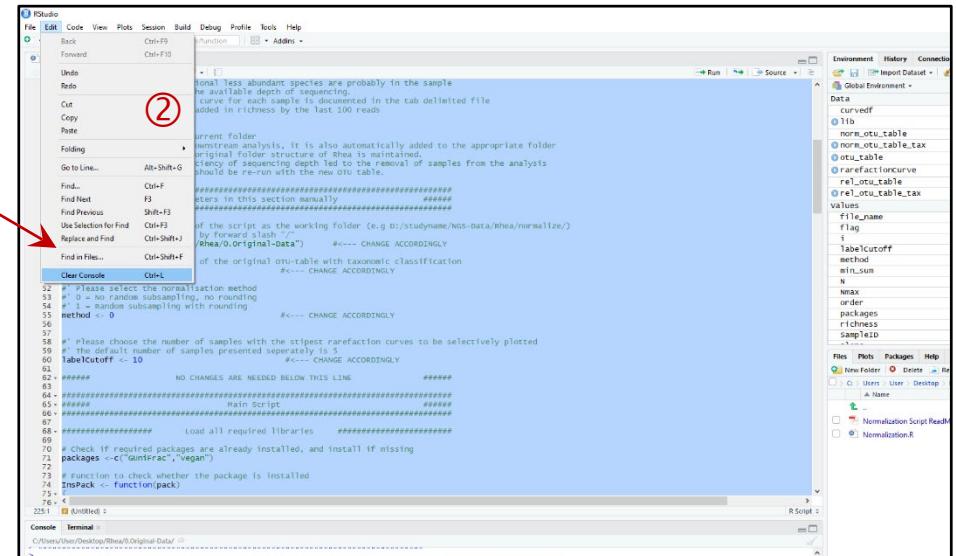
Now you have learned how to run the R-scripts.
You have run „install_packages.R“ in the Rhea-folder.
Now just a little hint about Rstudio...

① After running any script, the Console Window gets crowded. Delete all status messages by either pressing Strg + L or go to Edit > Clear console



1

```
54 #-----#
55 #-----#           Main Script          #-----#
56 #-----#
57 #
58 #
59 ##### Load all required libraries #####
60 #
61 #
62 # Check if required packages are already installed, and install if missing
63 packages <- c("tidyverse", "vegan")
64 install.packages(packages)
65 #
66 # Function to check whether the package is installed
67 InPack <- function(pack)
68 {
69   if (!requireNamespace(pack, quietly = TRUE))
70     install.packages(pack)
71 }
72 #
73 #
74 #-----#
75 #
76 #-----#
77 
```



③ The console window is empty again

We proceed with the actual data.

Call „Normalization.R“ from your Rhea folder – You have put your data „OTUs-Table.tab“ in this folder before.

a new tab is opened ②

```
install_packages.R * Normalization.R
  1 # Version 2.0
  2 # This script was last modified on 19/01/2018
  3 # Script Task: Normalize otu-tables
  4 # Author: ilias lagkouvardos
  5 #
  6 # Normalize abundance values of the input otu table
  7 # Calculate relative abundances for all OTUs based on normalized values
  8 # Calculate rarefaction curves to help estimate sufficiency of sequencing depth for each sample
  9 #
 10 # Input: Please enter following parameters
 11 # 1. Set the path to the directory where the file is stored
 12 # 2. write the name of the otu-table of interest in quotes
 13 #
 14 # The script generates five tab-delimited files and one pdf file
 15 # 1. Normalized counts with taxonomy information
 16 # 2. Normalized counts without taxonomy information
 17 # 3. Normalized relative abundances with taxonomy information
 18 # 4. Normalized relative abundances without taxonomy information
 19 # 5. Rarefaction curves for all samples and the most undersampled ones (default 5 cases) as PDF
 20 # 6. Slope of the rarefaction curve as species per 100 reads
 21 #
 22 # Concept:
 23 # The default method followed is normalization via division by the sum of sequences in a given sample
 24 # and multiplication by the minimum sum across all samples. It is used instead of the classic rarefaction approach
 25 # to avoid unnecessary variation due to the random subsampling and loss of information due to rounding.
 26 # The option of random subsampling is still available for normalization if deemed necessary by users.
 27 # Rarefaction curves are showing species richness with respect to sequencing depth (number of reads).
 28 # Undersequenced samples are those that their rarefaction curve is not reaching plateau at the available number of reads.
 29 # This indicates that additional less abundant species are probably in the sample
 30 # but were not covered by the available depth of sequencing.
 31 # The terminal slope of the curve for each sample is documented in the tab delimited file
 32 # as the number of species added in richness by the last 100 reads
 33 #
 34 # Note:
 35 # Files are stored in the current folder
 36 # If a file is needed for downstream analysis, it is also automatically added to the appropriate folder
 37 # under the condition that original folder structure of Rhea is maintained.
 38 # If the evaluation of sufficiency of sequencing depth led to the removal of samples from the analysis
 39 # the normalization script should be re-run with the new otu table.
 40 #
 41 ##### Set parameters in this section manually #####
 42 #
 43 #
 44 #
 45 # Please set the directory of the script as the working folder (e.g. D:/studynome/NGS-Data/Rhea/normalize/)
 46 # Note: the path is denoted by forward slash "/"
 47 setwd("D:/tra/tslamiis-Rhea/Error/Rhea/1.Normalization") #<-- CHANGE ACCORDINGLY
 48 #
 49 # Please give the file name of the original otu-table with taxonomic classification
 50 file_name<- "OTUs-Table.tab" #<-- CHANGE ACCORDINGLY
 51 #
 52 # Please select the normalisation method
 53 # 0 = No random subsampling, no rounding
 54 # 1 = Random subsampling with rounding
 55 method <- 0 #<-- CHANGE ACCORDINGLY
 56 #
 57 #
 58 # Please choose the number of samples with the steepest rarefaction curves to be selectively plotted
 59 # The default number of samples presented separately is 5
 60 #
 61 #
 62 #
 63 #
 64 #
 65 #
 66 #
 67 #
 68 #
 69 #
 70 #
 71 #
 72 #
 73 #
 74 #
 75 #
 76 #
 77 #
 78 #
 79 #
 80 #
 81 #
 82 #
 83 #
 84 #
 85 #
 86 #
 87 #
 88 #
 89 #
 90 #
 91 #
 92 #
 93 #
 94 #
 95 #
 96 #
 97 #
 98 #
 99 #
 100 #
 101 #
 102 #
 103 #
 104 #
 105 #
 106 #
 107 #
 108 #
 109 #
 110 #
 111 #
 112 #
 113 #
 114 #
 115 #
 116 #
 117 #
 118 #
 119 #
 120 #
 121 #
 122 #
 123 #
 124 #
 125 #
 126 #
 127 #
 128 #
 129 #
 130 #
 131 #
 132 #
 133 #
 134 #
 135 #
 136 #
 137 #
 138 #
 139 #
 140 #
 141 #
 142 #
 143 #
 144 #
 145 #
 146 #
 147 #
 148 #
 149 #
 150 #
 151 #
 152 #
 153 #
 154 #
 155 #
 156 #
 157 #
 158 #
 159 #
 160 #
 161 #
 162 #
 163 #
 164 #
 165 #
 166 #
 167 #
 168 #
 169 #
 170 #
 171 #
 172 #
 173 #
 174 #
 175 #
 176 #
 177 #
 178 #
 179 #
 180 #
 181 #
 182 #
 183 #
 184 #
 185 #
 186 #
 187 #
 188 #
 189 #
 190 #
 191 #
 192 #
 193 #
 194 #
 195 #
 196 #
 197 #
 198 #
 199 #
 200 #
 201 #
 202 #
 203 #
 204 #
 205 #
 206 #
 207 #
 208 #
 209 #
 210 #
 211 #
 212 #
 213 #
 214 #
 215 #
 216 #
 217 #
 218 #
 219 #
 220 #
 221 #
 222 #
 223 #
 224 #
 225 #
 226 #
 227 #
 228 #
 229 #
 230 #
 231 #
 232 #
 233 #
 234 #
 235 #
 236 #
 237 #
 238 #
 239 #
 240 #
 241 #
 242 #
 243 #
 244 #
 245 #
 246 #
 247 #
 248 #
 249 #
 250 #
 251 #
 252 #
 253 #
 254 #
 255 #
 256 #
 257 #
 258 #
 259 #
 250 #
 251 #
 252 #
 253 #
 254 #
 255 #
 256 #
 257 #
 258 #
 259 #
 260 #
 261 #
 262 #
 263 #
 264 #
 265 #
 266 #
 267 #
 268 #
 269 #
 270 #
 271 #
 272 #
 273 #
 274 #
 275 #
 276 #
 277 #
 278 #
 279 #
 280 #
 281 #
 282 #
 283 #
 284 #
 285 #
 286 #
 287 #
 288 #
 289 #
 280 #
 281 #
 282 #
 283 #
 284 #
 285 #
 286 #
 287 #
 288 #
 289 #
 290 #
 291 #
 292 #
 293 #
 294 #
 295 #
 296 #
 297 #
 298 #
 299 #
 300 #
 301 #
 302 #
 303 #
 304 #
 305 #
 306 #
 307 #
 308 #
 309 #
 310 #
 311 #
 312 #
 313 #
 314 #
 315 #
 316 #
 317 #
 318 #
 319 #
 320 #
 321 #
 322 #
 323 #
 324 #
 325 #
 326 #
 327 #
 328 #
 329 #
 330 #
 331 #
 332 #
 333 #
 334 #
 335 #
 336 #
 337 #
 338 #
 339 #
 330 #
 331 #
 332 #
 333 #
 334 #
 335 #
 336 #
 337 #
 338 #
 339 #
 340 #
 341 #
 342 #
 343 #
 344 #
 345 #
 346 #
 347 #
 348 #
 349 #
 350 #
 351 #
 352 #
 353 #
 354 #
 355 #
 356 #
 357 #
 358 #
 359 #
 360 #
 361 #
 362 #
 363 #
 364 #
 365 #
 366 #
 367 #
 368 #
 369 #
 370 #
 371 #
 372 #
 373 #
 374 #
 375 #
 376 #
 377 #
 378 #
 379 #
 380 #
 381 #
 382 #
 383 #
 384 #
 385 #
 386 #
 387 #
 388 #
 389 #
 390 #
 391 #
 392 #
 393 #
 394 #
 395 #
 396 #
 397 #
 398 #
 399 #
 400 #
 401 #
 402 #
 403 #
 404 #
 405 #
 406 #
 407 #
 408 #
 409 #
 410 #
 411 #
 412 #
 413 #
 414 #
 415 #
 416 #
 417 #
 418 #
 419 #
 420 #
 421 #
 422 #
 423 #
 424 #
 425 #
 426 #
 427 #
 428 #
 429 #
 430 #
 431 #
 432 #
 433 #
 434 #
 435 #
 436 #
 437 #
 438 #
 439 #
 440 #
 441 #
 442 #
 443 #
 444 #
 445 #
 446 #
 447 #
 448 #
 449 #
 450 #
 451 #
 452 #
 453 #
 454 #
 455 #
 456 #
 457 #
 458 #
 459 #
 460 #
 461 #
 462 #
 463 #
 464 #
 465 #
 466 #
 467 #
 468 #
 469 #
 470 #
 471 #
 472 #
 473 #
 474 #
 475 #
 476 #
 477 #
 478 #
 479 #
 480 #
 481 #
 482 #
 483 #
 484 #
 485 #
 486 #
 487 #
 488 #
 489 #
 490 #
 491 #
 492 #
 493 #
 494 #
 495 #
 496 #
 497 #
 498 #
 499 #
 500 #
 501 #
 502 #
 503 #
 504 #
 505 #
 506 #
 507 #
 508 #
 509 #
 510 #
 511 #
 512 #
 513 #
 514 #
 515 #
 516 #
 517 #
 518 #
 519 #
 520 #
 521 #
 522 #
 523 #
 524 #
 525 #
 526 #
 527 #
 528 #
 529 #
 530 #
 531 #
 532 #
 533 #
 534 #
 535 #
 536 #
 537 #
 538 #
 539 #
 540 #
 541 #
 542 #
 543 #
 544 #
 545 #
 546 #
 547 #
 548 #
 549 #
 550 #
 551 #
 552 #
 553 #
 554 #
 555 #
 556 #
 557 #
 558 #
 559 #
 560 #
 561 #
 562 #
 563 #
 564 #
 565 #
 566 #
 567 #
 568 #
 569 #
 570 #
 571 #
 572 #
 573 #
 574 #
 575 #
 576 #
 577 #
 578 #
 579 #
 580 #
 581 #
 582 #
 583 #
 584 #
 585 #
 586 #
 587 #
 588 #
 589 #
 590 #
 591 #
 592 #
 593 #
 594 #
 595 #
 596 #
 597 #
 598 #
 599 #
 600 #
 601 #
 602 #
 603 #
 604 #
 605 #
 606 #
 607 #
 608 #
 609 #
 610 #
 611 #
 612 #
 613 #
 614 #
 615 #
 616 #
 617 #
 618 #
 619 #
 620 #
 621 #
 622 #
 623 #
 624 #
 625 #
 626 #
 627 #
 628 #
 629 #
 630 #
 631 #
 632 #
 633 #
 634 #
 635 #
 636 #
 637 #
 638 #
 639 #
 640 #
 641 #
 642 #
 643 #
 644 #
 645 #
 646 #
 647 #
 648 #
 649 #
 650 #
 651 #
 652 #
 653 #
 654 #
 655 #
 656 #
 657 #
 658 #
 659 #
 660 #
 661 #
 662 #
 663 #
 664 #
 665 #
 666 #
 667 #
 668 #
 669 #
 670 #
 671 #
 672 #
 673 #
 674 #
 675 #
 676 #
 677 #
 678 #
 679 #
 680 #
 681 #
 682 #
 683 #
 684 #
 685 #
 686 #
 687 #
 688 #
 689 #
 690 #
 691 #
 692 #
 693 #
 694 #
 695 #
 696 #
 697 #
 698 #
 699 #
 700 #
 701 #
 702 #
 703 #
 704 #
 705 #
 706 #
 707 #
 708 #
 709 #
 710 #
 711 #
 712 #
 713 #
 714 #
 715 #
 716 #
 717 #
 718 #
 719 #
 720 #
 721 #
 722 #
 723 #
 724 #
 725 #
 726 #
 727 #
 728 #
 729 #
 730 #
 731 #
 732 #
 733 #
 734 #
 735 #
 736 #
 737 #
 738 #
 739 #
 740 #
 741 #
 742 #
 743 #
 744 #
 745 #
 746 #
 747 #
 748 #
 749 #
 750 #
 751 #
 752 #
 753 #
 754 #
 755 #
 756 #
 757 #
 758 #
 759 #
 760 #
 761 #
 762 #
 763 #
 764 #
 765 #
 766 #
 767 #
 768 #
 769 #
 770 #
 771 #
 772 #
 773 #
 774 #
 775 #
 776 #
 777 #
 778 #
 779 #
 780 #
 781 #
 782 #
 783 #
 784 #
 785 #
 786 #
 787 #
 788 #
 789 #
 780 #
 781 #
 782 #
 783 #
 784 #
 785 #
 786 #
 787 #
 788 #
 789 #
 790 #
 791 #
 792 #
 793 #
 794 #
 795 #
 796 #
 797 #
 798 #
 799 #
 800 #
 801 #
 802 #
 803 #
 804 #
 805 #
 806 #
 807 #
 808 #
 809 #
 810 #
 811 #
 812 #
 813 #
 814 #
 815 #
 816 #
 817 #
 818 #
 819 #
 820 #
 821 #
 822 #
 823 #
 824 #
 825 #
 826 #
 827 #
 828 #
 829 #
 830 #
 831 #
 832 #
 833 #
 834 #
 835 #
 836 #
 837 #
 838 #
 839 #
 840 #
 841 #
 842 #
 843 #
 844 #
 845 #
 846 #
 847 #
 848 #
 849 #
 850 #
 851 #
 852 #
 853 #
 854 #
 855 #
 856 #
 857 #
 858 #
 859 #
 860 #
 861 #
 862 #
 863 #
 864 #
 865 #
 866 #
 867 #
 868 #
 869 #
 870 #
 871 #
 872 #
 873 #
 874 #
 875 #
 876 #
 877 #
 878 #
 879 #
 880 #
 881 #
 882 #
 883 #
 884 #
 885 #
 886 #
 887 #
 888 #
 889 #
 880 #
 881 #
 882 #
 883 #
 884 #
 885 #
 886 #
 887 #
 888 #
 889 #
 890 #
 891 #
 892 #
 893 #
 894 #
 895 #
 896 #
 897 #
 898 #
 899 #
 900 #
 901 #
 902 #
 903 #
 904 #
 905 #
 906 #
 907 #
 908 #
 909 #
 910 #
 911 #
 912 #
 913 #
 914 #
 915 #
 916 #
 917 #
 918 #
 919 #
 920 #
 921 #
 922 #
 923 #
 924 #
 925 #
 926 #
 927 #
 928 #
 929 #
 930 #
 931 #
 932 #
 933 #
 934 #
 935 #
 936 #
 937 #
 938 #
 939 #
 940 #
 941 #
 942 #
 943 #
 944 #
 945 #
 946 #
 947 #
 948 #
 949 #
 950 #
 951 #
 952 #
 953 #
 954 #
 955 #
 956 #
 957 #
 958 #
 959 #
 960 #
 961 #
 962 #
 963 #
 964 #
 965 #
 966 #
 967 #
 968 #
 969 #
 970 #
 971 #
 972 #
 973 #
 974 #
 975 #
 976 #
 977 #
 978 #
 979 #
 980 #
 981 #
 982 #
 983 #
 984 #
 985 #
 986 #
 987 #
 988 #
 989 #
 990 #
 991 #
 992 #
 993 #
 994 #
 995 #
 996 #
 997 #
 998 #
 999 #
 1000 #
 1001 #
 1002 #
 1003 #
 1004 #
 1005 #
 1006 #
 1007 #
 1008 #
 1009 #
 10010 #
 10011 #
 10012 #
 10013 #
 10014 #
 10015 #
 10016 #
 10017 #
 10018 #
 10019 #
 10020 #
 10021 #
 10022 #
 10023 #
 10024 #
 10025 #
 10026 #
 10027 #
 10028 #
 10029 #
 10030 #
 10031 #
 10032 #
 10033 #
 10034 #
 10035 #
 10036 #
 10037 #
 10038 #
 10039 #
 10040 #
 10041 #
 10042 #
 10043 #
 10044 #
 10045 #
 10046 #
 10047 #
 10048 #
 10049 #
 10050 #
 10051 #
 10052 #
 10053 #
 10054 #
 10055 #
 10056 #
 10057 #
 10058 #
 10059 #
 10060 #
 10061 #
 10062 #
 10063 #
 10064 #
 10065 #
 10066 #
 10067 #
 10068 #
 10069 #
 10070 #
 10071 #
 10072 #
 10073 #
 10074 #
 10075 #
 10076 #
 10077 #
 10078 #
 10079 #
 10080 #
 10081 #
 10082 #
 10083 #
 10084 #
 10085 #
 10086 #
 10087 #
 10088 #
 10089 #
 10080 #
 10081 #
 10082 #
 10083 #
 10084 #
 10085 #
 10086 #
 10087 #
 10088 #
 10089 #
 10090 #
 10091 #
 10092 #
 10093 #
 10094 #
 10095 #
 10096 #
 10097 #
 10098 #
 10099 #
 100100 #
 100101 #
 100102 #
 100103 #
 100104 #
 100105 #
 100106 #
 100107 #
 100108 #
 100109 #
 100110 #
 100111 #
 100112 #
 100113 #
 100114 #
 100115 #
 100116 #
 100117 #
 100118 #
 100119 #
 100120 #
 100121 #
 100122 #
 100123 #
 100124 #
 100125 #
 100126 #
 100127 #
 100128 #
 100129 #
 100130 #
 100131 #
 100132 #
 100133 #
 100134 #
 100135 #
 100136 #
 100137 #
 100138 #
 100139 #
 100140 #
 100141 #
 100142 #
 100143 #
 100144 #
 100145 #
 100146 #
 100147 #
 100148 #
 100149 #
 100150 #
 100151 #
 100152 #
 100153 #
 100154 #
 100155 #
 100156 #
 100157 #
 100158 #
 100159 #
 100160 #
 100161 #
 100162 #
 100163 #
 100164 #
 100165 #
 100166 #
 100167 #
 100168 #
 100169 #
 100170 #
 100171 #
 100172 #
 100173 #
 100174 #
 100175 #
 100176 #
 100177 #
 100178 #
 100179 #
 100180 #
 100181 #
 100182 #
 100183 #
 100184 #
 100185 #
 100186 #
 100187 #
 100188 #
 100189 #
 100190 #
 100191 #
 100192 #
 100193 #
 100194 #
 100195 #
 100196 #
 100197 #
 100198 #
 100199 #
 1001910 #
 1001911 #
 1001912 #
 1001913 #
 1001914 #
 1001915 #
 1001916 #
 1001917 #
 1001918 #
 1001919 #
 1001920 #
 1001921 #
 1001922 #
 1001923 #
 1001924 #
 1001925 #
 1001926 #
 1001927 #
 1001928 #
 1001929 #
 1001930 #
 1001931 #
 1001932 #
 1001933 #
 1001934 #
 1001935 #
 1001936 #
 1001937 #
 1001938 #
 1001939 #
 1001940 #
 1001941 #
 1001942 #
 1001943 #
 1001944 #
 1001945 #
 1001946 #
 1001947 #
 1001948 #
 1001949 #
 1001950 #
 1001951 #
 1001952 #
 1001953 #
 1001954 #
 1001955 #
 1001956 #
 1001957 #
 1001958 #
 1001959 #
 1001960 #
 1001961 #
 1001962 #
 1001963 #
 1001964 #
 1001965 #
 1001966 #
 1001967 #
 1001968 #
 1001969 #
 1001970 #
 1001971 #
 1001972 #
 1001973 #
 1001974 #
 1001975 #
 1001976 #
 1001977 #
 1001978 #
 1001979 #
 1001980 #
 1001981 #
 1001982 #
 1001983 #
 1001984 #
 1001985 #
 1001986 #
 1001987 #
 1001988 #
 1001989 #
 1001990 #
 1001991 #
 1001992 #
 1001993 #
 1001994 #
 1001995 #
 1001996 #
 1001997 #
 1001998 #
 1001999 #
 10019100 #
 10019101 #
 10019102 #
 10019103 #
 10019104 #
 10019105 #
 10019106 #
 10019107 #
 10019108 #
 10019109 #
 10019110 #
 10019111 #
 10019112 #
 10019113 #
 10019114 #
 10019115 #
 10019116 #
 10019117 #
 10019118 #
 10019119 #
 10019120 #
 10019121 #
 10019122 #
 10019123 #
 10019124 #
 10019125 #
 10019126 #
 10019127 #
 10019128 #
 10019129 #
 10019130 #
 10019131 #
 10019132 #
 10019133 #
 10019134 #
 10019135 #
 10019136 #
 10019137 #
 10019138 #
 10019139 #
 10019140 #
 10019141 #
 10019142 #
 10019143 #
 10019144 #
 10019145 #
 10019146 #
 10019147 #
 10019148 #
 10019149 #
 10019150 #
 10019151 #
 10019152 #
 10019153 #
 10019154 #
 10019155 #
 10019156 #
 10019157 #
 10019158 #
 10019159 #
 10019160 #
 10019161 #
 10019162 #
 10019163 #
 10019164 #
 10019165 #
 10019166 #
 10019167 #
 10019168 #
 10019169 #
 10019170 #
 10019171 #
 10019172 #
 10019173 #
 10019174 #
 10019175 #
 10019176 #
 10019177 #
 10019178 #
 10019179 #
 10019180 #
 10019181 #
 10019182 #
 10019183 #
 10019184 #
 10019185 #
 10019186 #
 10019187 #
 10019188 #
 10019189 #
 10019190 #
 10019191 #
 10019192 #
 10019193 #
 10019194 #
 10019195 #
 10019196 #
 10019197 #
 10019198 #
 10019199 #
 100191100 #
 100191101 #
 100191102 #
 100191103 #
 100191104 #
 100191105 #
 100191106 #
 100191107 #
 100191108 #
 100191109 #
 100191110 #
 100191111 #
 100191112 #
 100191113 #
 100191114 #
 100191115 #
 100191116 #
 100191117 #
 100191118 #
 100191119 #
 100191120 #
 100191121 #
 100191122 #
 100191123 #
 100191124 #
 100191125 #
 100191126 #
 100191127 #
 100191128 #
 100191129 #
 100191130 #
 100191131 #
 100191132 #
 100191133 #
 100191134 #
 100191135 #
 100191136 #
 100191137 #
 100191138 #
 100191139 #
 100191140 #
 100191141 #
 100191142 #
 100191143 #
 100191144 #
 100191145 #
 100191146 #
 100191147 #
 100191148 #
 100191149 #
 100191150 #
 100191151 #
 100191152 #
 100191153 #
 100191154 #
 100191155 #
 100191156 #
 100191157 #
 100191158 #
 100191159 #
 100191160 #
 100191161 #
 100191162 #
 100191163 #
 100191164 #
 100191165 #
 100191166 #
 100191167 #
 100191168 #
 100191169 #
 100191170 #
 100191171 #
 100191172 #
 100191173 #
 100191174 #
 100191175 #
 100191176 #
 100191177 #
 100191178 #
 100191179 #
 100191180 #
 100191181 #
 100191182 #
 100191183 #
 100191184 #
 100191185 #
 100191186 #
 100191187 #
 100191188 #
 100191189 #
 100191190 #
 100191191 #
 100191192 #
 100191193 #
 100191194 #
 100191195 #
 100191196 #
 100191197 #
 100191198 #
 100191199 #
 1001911100 #
 1001911101 #
 1001911102 #
 1001911103 #
 1001911104 #
 1001911105 #
 1001911106 #
 1001911107 #
 1001911108 #
 1001911109 #
 1001911110 #
 1001911111 #
 1001911112 #
 1001911113 #
 1001911114 #
 1001911115 #
 1001911116 #
 1001911117 #
 1001911118 #
 1001911119 #
 1001911120 #
 1001911121 #
 1001911122 #
 1001911123 #
 1001911124 #
 1001911125 #
 1001911126 #
 1001911127 #
 1001911128 #
 1001911129 #
 1001911130 #
 1001911131 #
 1001911132 #
 1001911133 #
 1001911134 #
 1001911135 #
 1001911136 #
 1001911137 #
 1001911138 #
 1001911139 #
 1001911140 #
 1001911141 #
 1001911142 #
 1001911143 #
 1001911144 #
 1001911145 #
 1001911146 #
 1001911147 #
 1001911148 #
 1001911149 #
 1001911150 #
 1001911151 #
 1001911152 #
 1001911153 #
 1001911154 #
 1001911155 #
 1001911156 #
 1001911157 #
 1001911158 #
 1001911159 #
 1001911160 #
 1001911161 #
 1001911162 #
 1001911163 #
 1001911164 #
 1001911165 #
 1001911166 #
 1001911167 #
 1001911168 #
 1001911169 #
 1001911170 #
 1001911171 #
 1001911172 #
 1001911173 #
 1001911174 #
 1001911175 #
 1001911176 #
 1001911177 #
 1001911178 #
 1001911179 #
 1001911180 #
 1001911181 #
 1001911182 #
 1001911183 #
 1001911184 #
 1001911185 #
 1001911186 #
 1001911187 #
 1001911188 #
 1001911189 #
 1001911190 #
 1001911191 #
 1001911192 #
 1001911193 #
 1001911194 #
 1001911195 #
 1001911196 #
 1001911197 #
 1001911198 #
 1001911199 #
 10019111100 #
 10019111101 #
 10019111102 #
 10019111103 #
 10019111104 #
 10019111105 #
 10019111106 #
 10019111107 #
 10019111108 #
 10019111109 #
 10019111110 #
 10019111111 #
 10019111112 #
 10019111113 #
 10019111114 #
 10019111115 #
 10019111116 #
 10019111117 #
 10019111118 #
 10019111119 #
 10019111120 #
 10019111121 #
 10019111122 #
 10019111123 #
 10019111124 #
 10019111125 #
 10019111126 #
 1
```

After running „Normalization.R“ (remember: select all in the Script window, press run, wait until finished)

The script „Normalization.R“ creates a data output, most of which is placed in the folder 1.Normalization.

①

②

③

open „RarefactionCurve.pdf“

Name	Änderungsdatum	Typ	Größe
.Rhistory	20.04.2018 11:01	RHISTORY-Datei	15 KB
Normalization Script ReadMe.pdf	19.01.2018 17:17	Adobe Acrobat-D...	154 KB
Normalization.R	19.01.2018 17:17	R-Datei	11 KB
OTUs_Table-norm.tab	20.04.2018 11:40	TAB-Datei	245 KB
OTUs_Table-norm-rel.tab	20.04.2018 11:40	TAB-Datei	266 KB
OTUs_Table-norm-rel-tax.tab	20.04.2018 11:40	TAB-Datei	290 KB
OTUs_Table-norm-tax.tab	20.04.2018 11:40	TAB-Datei	270 KB
OTUs_Table.tab	09.02.2018 11:32	TAB-Datei	87 KB
RarefactionCurve.pdf	20.04.2018 11:40	Adobe Acrobat-D...	381 KB
RarefactionCurve.tab	20.04.2018 11:40	TAB-Datei	3 KB

Folder „2.Alpha-Diversity“ BEFORE running the script „Normalization.R“

Name	Änderungsdatum	Typ	Größe
Alpha Diversity Script ReadMe.pdf	09.02.2018 10:00	Adobe Acrobat-D...	96 KB
Alpha-Diversity.R	09.02.2018 10:00	R-Datei	4 KB

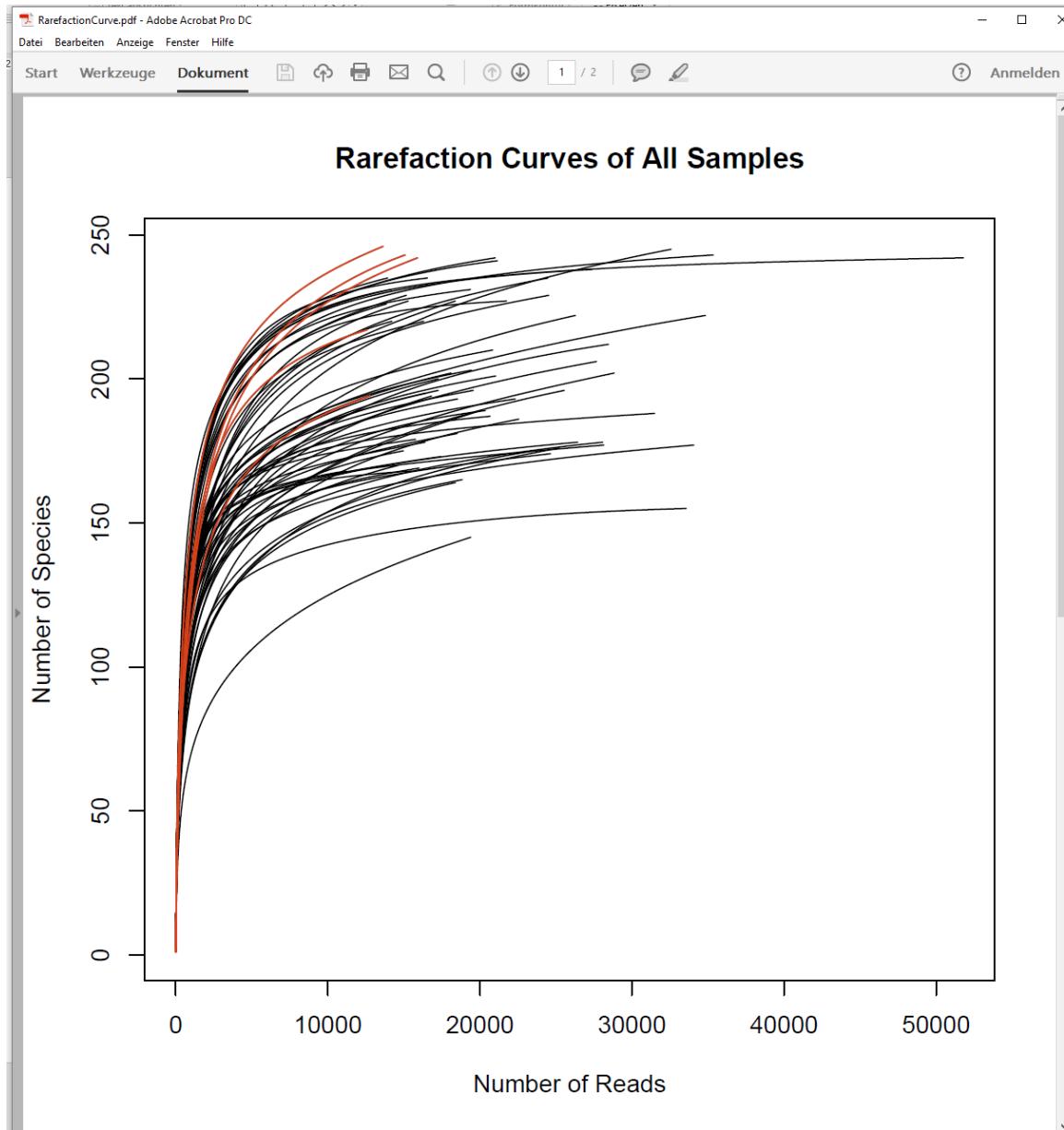
②

③

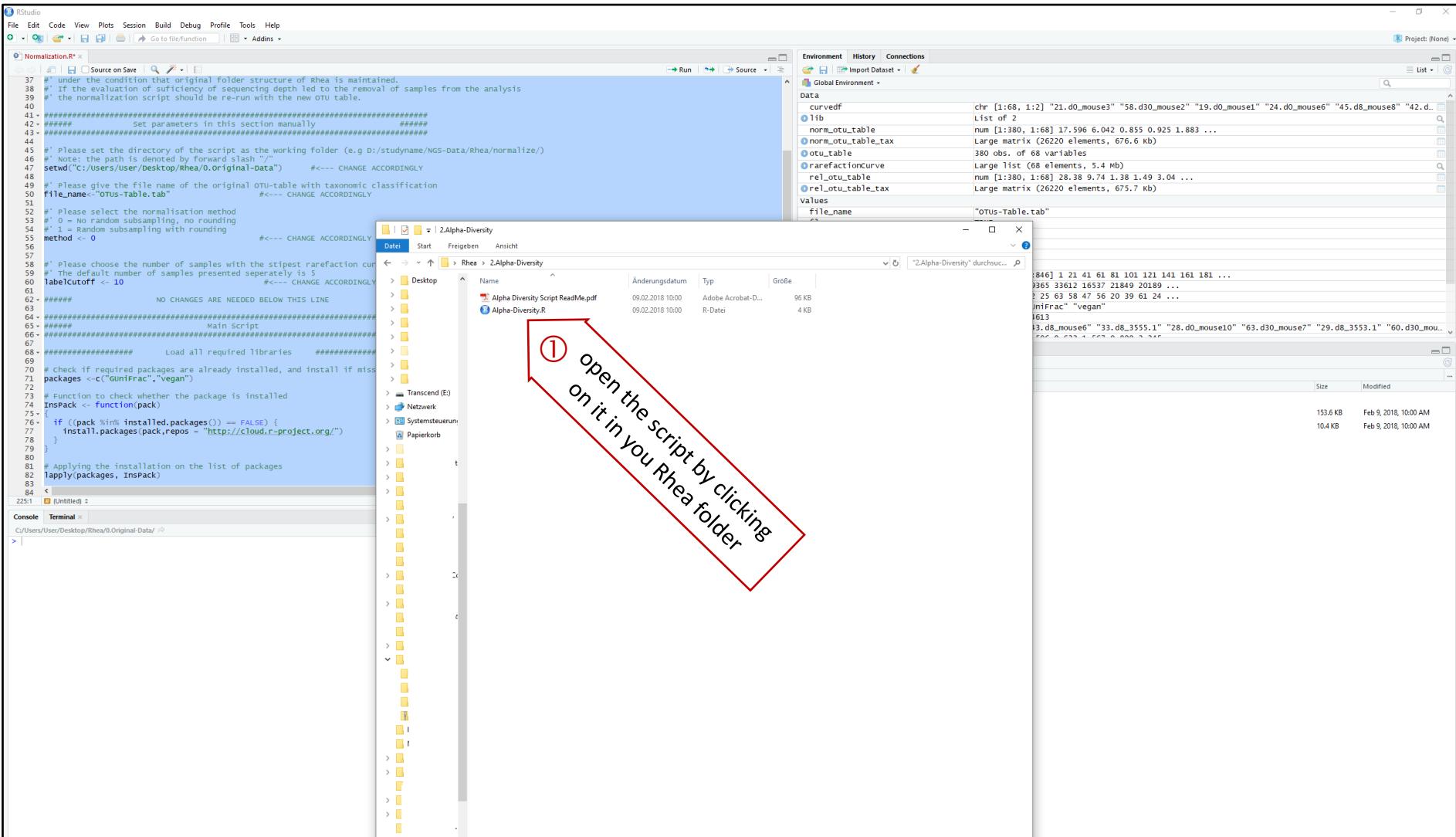
Folder „2.Alpha-Diversity“ AFTER running the script „Normalization.R“. The output „OTUs_Table-norm.tab“ is put here by the script.

Name	Änderungsdatum	Typ	Größe
Alpha Diversity Script ReadMe.pdf	09.02.2018 10:00	Adobe Acrobat-D...	96 KB
Alpha-Diversity.R	09.02.2018 10:00	R-Datei	4 KB
OTUs_Table-norm.tab	09.02.2018 10:27	TAB-Datei	276 KB

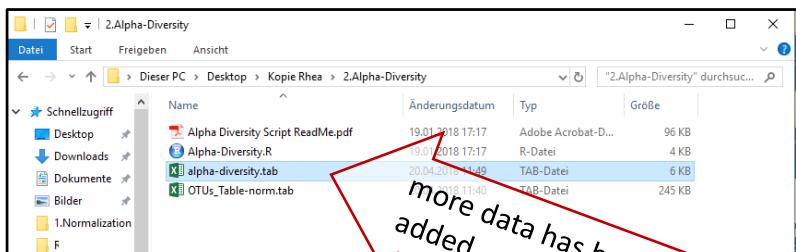
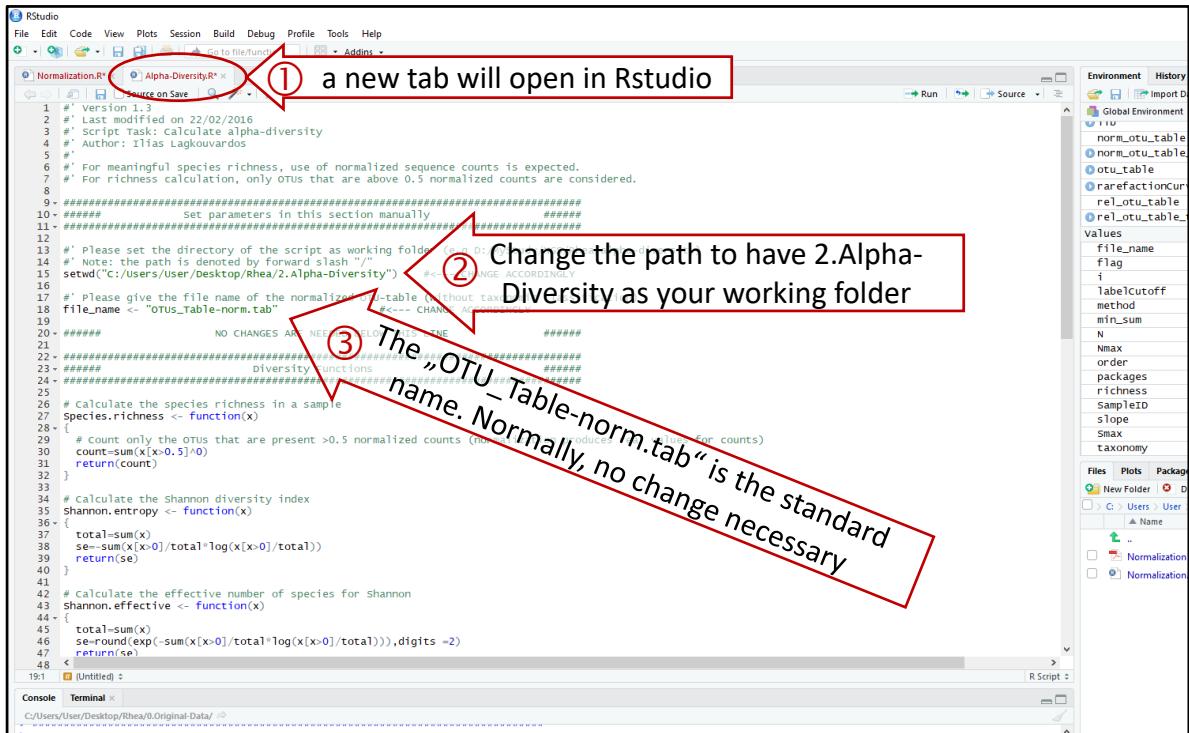
The first graphical output of your data are the rarefaction curve. The 5 worst curves are marked in red.



Run the next script „Alpha-Diversity.R“ - I



Run the next script „Alpha-Diversity.R“ - II

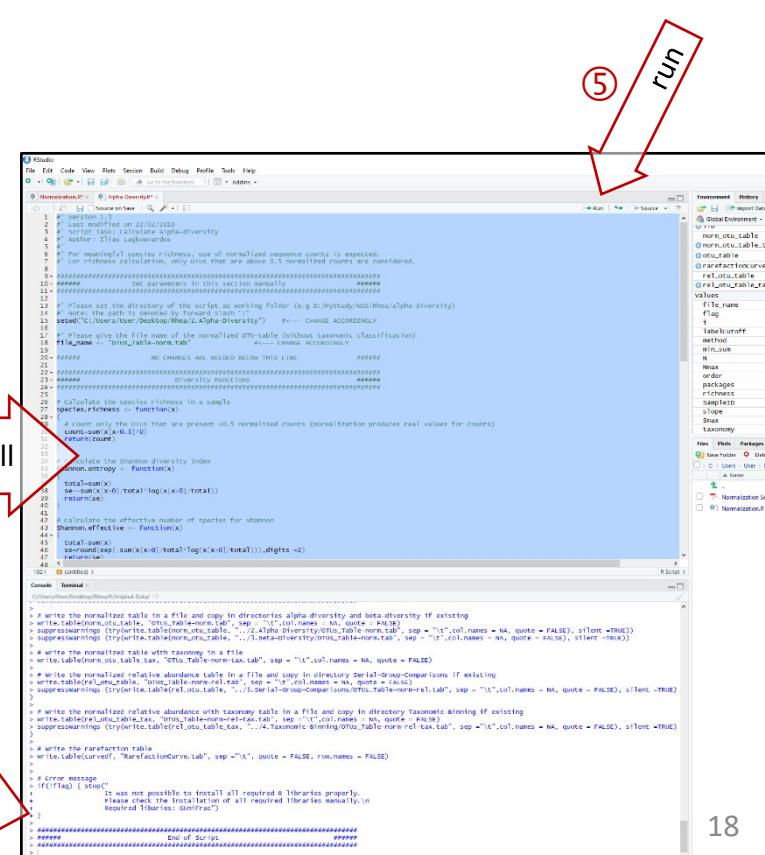


more data has been added

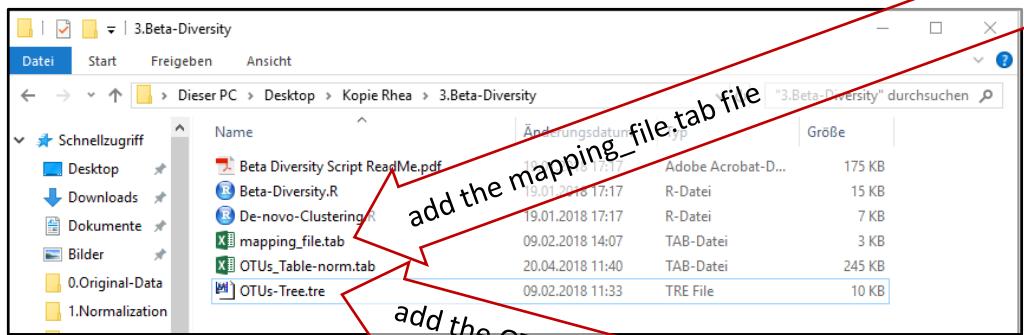
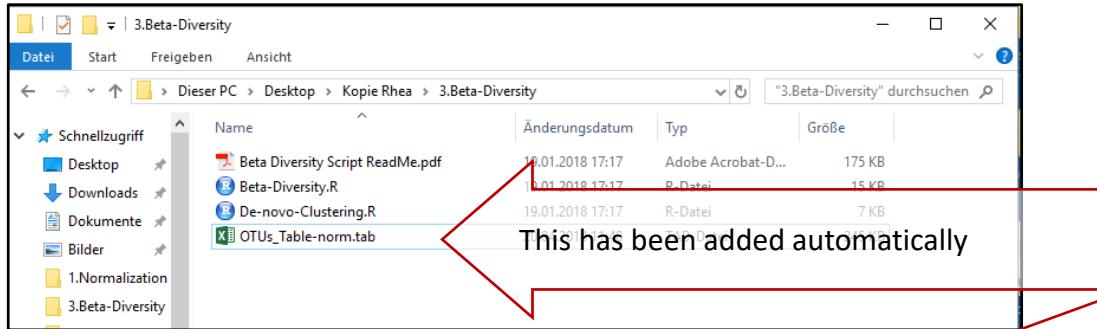
7

Select all

⑥ wait until script ended

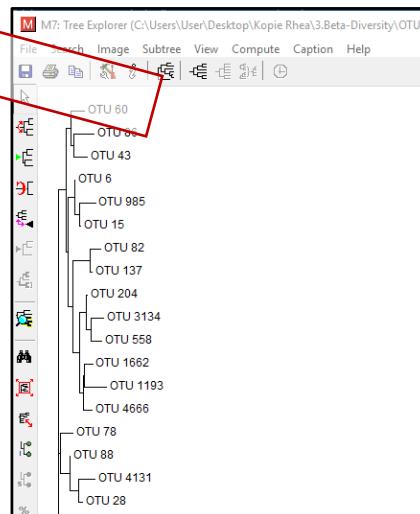


Proceed with the beta diversity ...



The mapping file contains the information, which sample belongs to which condition, cage, day, patient, group, etc. – this depends on your experiment!

The OTU tree is constructed by IMNGS (advanced: yourself).



A	B	C	D	E	F	G
#Sample	Day	Mouse	Cage	Facility	cage=mouse	
1						
2	01.d0_3250.1	0	2	3250	32501	
3	02.d0_3250.2	0	2	3250	32502	
4	03.d0_3250.3	0	3	3250	32503	
5	04.d0_3250.4	0	4	3250	32504	
6	05.d0_3250.5	0	5	3250	32505	
7	06.d0_3553.1	0	1	3553	35531	
8	07.d0_3553.2	0	2	3553	35532	
9	08.d0_3553.3	0	3	3553	35533	
10	09.d0_3553.4	0	4	3553	35534	
11	10.d0_3555.1	0	1	3555	35551	
12	11.d0_3555.2	0	2	3555	35552	
13	12.d0_3555.3	0	3	3555	35553	
14	13.d0_3555.4	0	4	3555	35554	
15	14.d0_3555.5	0	5	3555	35555	
16	15.d0_3556.1	0	1	3556	35561	
17	16.d0_3556.2	0	2	3556	35562	
18	17.d0_3556.3	0	3	3556	35563	
19	18.d0_3556.4	0	4	3556	35564	
20	19.d0_mouse1	0	1	1111 LSE	11101	
21	20.d0_mouse2	0	2	1111 LSE	11102	
22	21.d0_mouse3	0	3	1111 LSE	11103	
23	22.d0_mouse4	0	4	1111 LSE	11104	
24	23.d0_mouse5	0	5	1111 LSE	11105	
25	24.d0_mouse6	0	6	1111 LSE	11106	
26	25.d0_mouse7	0	7	1111 LSE	11107	
27	26.d0_mouse8	0	8	1111 LSE	11108	
28	27.d0_mouse9	0	9	1111 LSE	11109	
29	28.d0_mouse10	0	10	1111 LSE	11110	
30	29.d0_3553.1	8	1	3553	35531	
31	30.d0_3553.2	8	2	3553	35532	
32	31.d0_3553.3	8	3	3553	35533	
33	32.d0_3553.4	8	4	3553	35534	
34	33.d0_3555.1	8	1	3555	35551	
35	34.d0_3555.2	8	2	3555	35552	
36	35.d0_3555.3	8	3	3555	35553	
37	36.d0_3555.4	8	4	3555	35554	
38	37.d0_3555.5	8	5	3555	35555	
39	38.d0_mouse1	8	1	1111 LSE	11101	
40	39.d0_mouse2	8	2	1111 LSE	11102	
41	40.d0_mouse3	8	3	1111 LSE	11103	
42	41.d0_mouse4	8	4	1111 LSE	11104	
43	42.d0_mouse5	8	5	1111 LSE	11105	
44	43.d0_mouse6	8	6	1111 LSE	11106	
45	44.d0_mouse7	8	7	1111 LSE	11107	
46	45.d0_mouse8	8	8	1111 LSE	11108	
47	46.d0_mouse9	8	9	1111 LSE	11109	
48	47.d0_mouse10	8	10	1111 LSE	11110	
49	48.d0_3250.1	30	1	3250	32501	
50	49.d0_3250.2	30	2	3250	32502	
51	50.d0_3250.3	30	3	3250	32503	
52	51.d0_3250.4	30	4	3250	32504	
53	52.d0_3250.5	30	5	3250	32505	
54	53.d0_3556.1	30	1	3556	35561	
55	54.d0_3556.2	30	2	3556	35562	
56	55.d0_3556.3	30	3	3556	35563	
57	56.d0_3556.4	30	4	3556	35564	
58	57.d0_mouse1	30	1	1111 LSE	11101	
59	58.d0_mouse2	30	2	1111 LSE	11102	

Proceed with the beta diversity ... click the script „Beta-Diversity.R“

① a new tab will open in Rstudio

② Change the path

Normally, no change necessary

Normally, no change necessary

③ Change to the group you want to analyze! Our mapping file contains mice grouped in cages. We are interested in differences between cages, thus, insert „Cage“ (exactly as in the mapping file, case sensitive). You could choose another variable.

```
install.packages.R  Normalization.R  Beta-Diversity.R
```

```
1 # 1. The distance matrix
2 # 2. The unifrac distance matrix
3 # 3. The distance matrix calculated based on the generalized unifrac approach
4 # 4. The distance matrix
5 # 5. Plot showing the optimal number of clusters
6 # 6. Dendrogram for all samples in a newick tree file
7 # 7. Concept:
8 # 8. Taxonomic distances are calculated based on the generalized unifrac approach
9 # 9. Chen, J., et al. Associating microbiome composition with environmental covariates using generalized unifrac distances. 2012
10 # 10. Samples are clustered based on the distance matrix using the ward's hierarchical clustering method
11 # 11. To determine similarities between samples, a multivariate analysis is applied
12 # 12. and sample distribution is illustrated by means of MDS and NMDS (non-metric) plots
13 # 13. The Calinski-Harabasz (CH) Index is used to assess the optimal number of clusters the dataset was most robustly partitioned into
14 # 14. The CH index is calculated based on the generalized unifrac approach
15 # 15. The distance matrix is calculated based on the generalized unifrac approach
16 # 16. The distance matrix is calculated based on the generalized unifrac approach
17 # 17. The distance matrix is calculated based on the generalized unifrac approach
18 # 18. The distance matrix is calculated based on the generalized unifrac approach
19 # 19. The distance matrix is calculated based on the generalized unifrac approach
20 # 20. The distance matrix is calculated based on the generalized unifrac approach
21 # 21. The distance matrix is calculated based on the generalized unifrac approach
22 # 22. The distance matrix is calculated based on the generalized unifrac approach
23 # 23. The distance matrix is calculated based on the generalized unifrac approach
24 # 24. The distance matrix is calculated based on the generalized unifrac approach
25 # 25. The distance matrix is calculated based on the generalized unifrac approach
26 # 26. The distance matrix is calculated based on the generalized unifrac approach
27 # 27. Chen, J., et al. Associating microbiome composition with environmental covariates using generalized unifrac distances. 2012
28 # 28. Samples are clustered based on the distance matrix using the ward's hierarchical clustering method
29 # 29. To determine similarities between samples, a multivariate analysis is applied
30 # 30. and sample distribution is illustrated by means of MDS and NMDS (non-metric) plots
31 # 31. The Calinski-Harabasz (CH) Index is used to assess the optimal number of clusters the dataset was most robustly partitioned into
32 # 32. The CH index is calculated based on the generalized unifrac approach
33 #####
34 ##### Set parameters in this section manually #####
35 #####
36 # Please set the directory of the script as the working folder <-- n<-getwd();n<-c("C:/Users/User/Desktop/Kopie Rhea/3.Beta-Diversity")
37 # Note: the path is denoted by forward slash "/"
38 setwd("C:/users/User/Desktop/Kopie Rhea/3.Beta-Diversity") <-- C:/users/User/Desktop/Kopie Rhea/3.Beta-Diversity
39 # Please give the file name of the normalized otu-table without taxonomic classification
40 input_otu = "OTU_Table-norm.tab" <-- CHANGE ACCORDINGLY!!!
41 # Please give the name of the mapping_file.tab which contains individual sample information
42 input_meta = "mapping_file.tab" <-- CHANGE ACCORDINGLY!!!
43 # Please give the name of the phylogenetic tree file
44 input_tree = "OTU-tree.tre"
45 # Please give the column name (e.g. genotype) on which the samples are grouped (e.g. genotype)
46 group_name = "Cage"
47 ##### Additional parameters #####
48 # Turn on sample labeling
49 # 0 - Samples are not labeled in the MDS/NMDS plots
50 # 1 - All samples are labeled in the MDS/NMDS plots
51 label_samples =
52
53 # Determine which sample label should appear
54 # write the name of samples (in quotation marks), which should appear in the MDS/NMDS plots, in the vector (c) below
55 # If more than one sample should be plotted, please separate their IDs by comma (e.g. c("sample1","sample2"))
56 label_id <-c()
57
58 # De-Novo Clustering will be performed for the number of samples or maximal for the set limit
59 # Default Limit is 100
60 kmers_limit=100
61
62 <-
63
64
65
66
67
68
69
70
71 <
72 >
```

Environment History Connections

Data

- adonis
- all_fit
- lib
- meta
- meta_file
- otu_file
- paired_dist
- paired_matrix
- pairedMatrixList
- rooted_tree
- tree
- tree_file
- unifrac_dist
- unifrac_dist_comp

Values

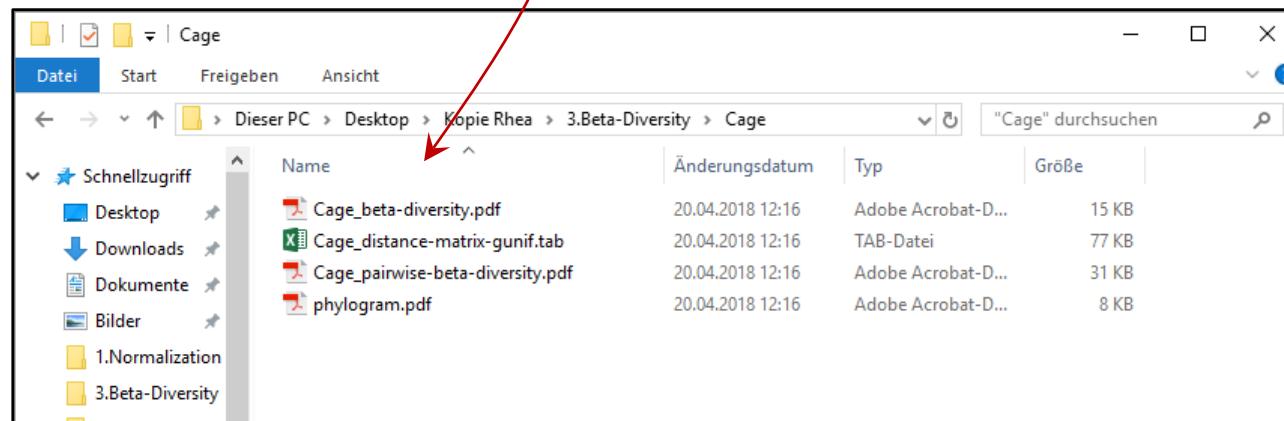
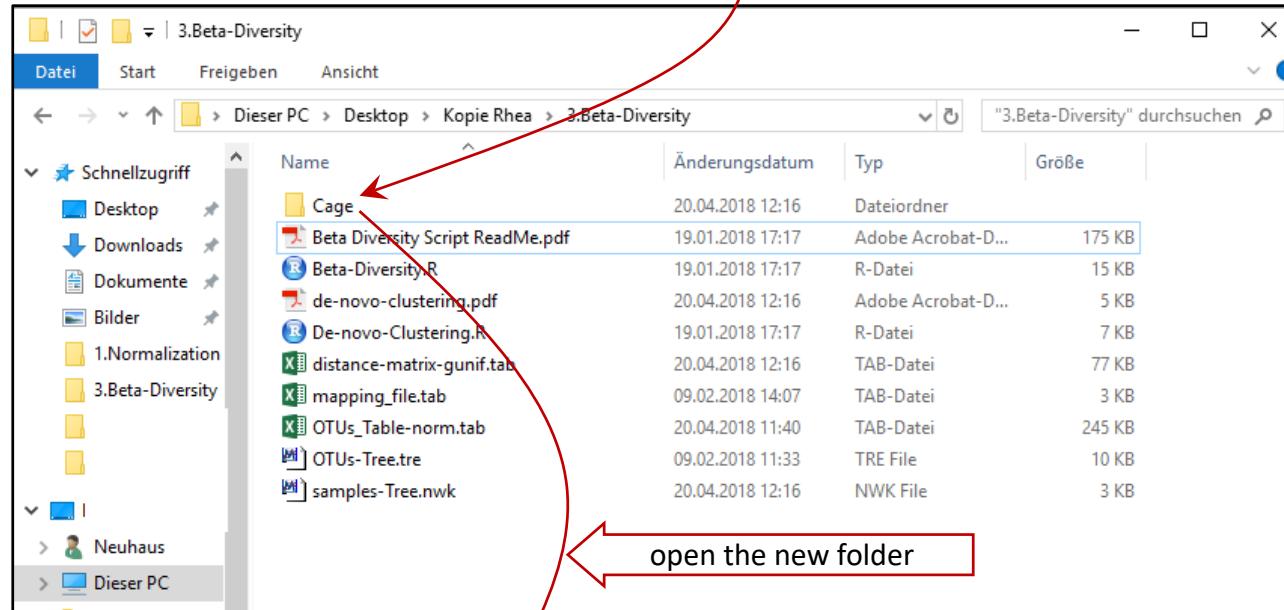
- all_dist_matrix
- all_groups
- all_groups_comp
- data_cluster
- file_name
- fin

Files Plots Packages Help Viewer

C:\Users\User\Desktop\Kopie Rhea

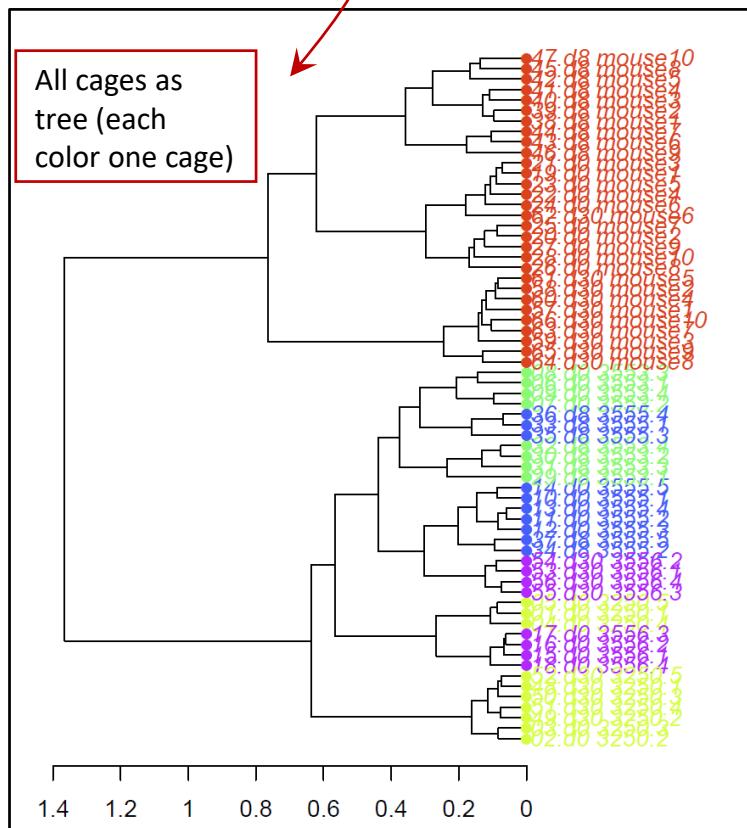
① ② ③

The script „Beta-Diversity.R“ produces a new folder for each group you analyzed!



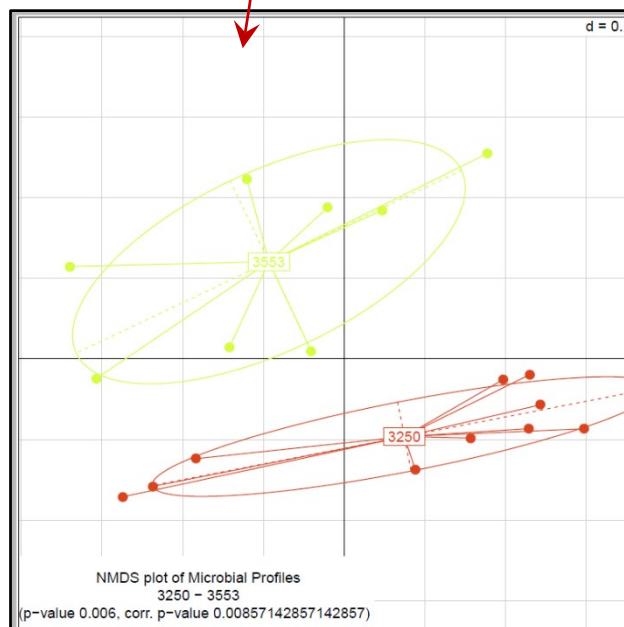
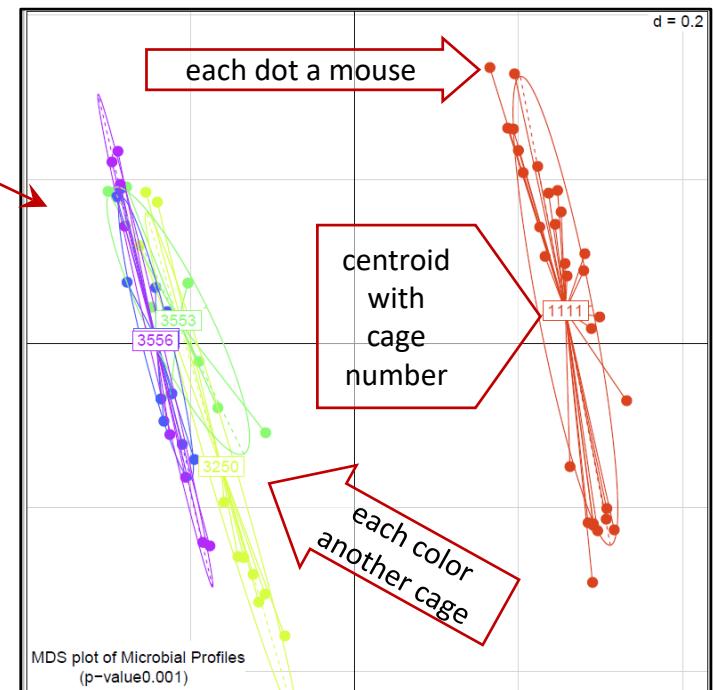
Cage

	Name	Änderungsdatum	Typ	Größe
Desktop	Cage_beta-diversity.pdf	20.04.2018 12:16	Adobe Acrobat-Dokument	15 KB
Downloads	Cage_distance-matrix-gunif.tab	20.04.2018 12:16	TAB-Datei	77 KB
Dokumente	Cage_pairwise-beta-diversity.pdf	20.04.2018 12:16	Adobe Acrobat-Dokument	31 KB
Bilder	phylogram.pdf	20.04.2018 12:16	Adobe Acrobat-Dokument	8 KB



All cages (each dot is one mouse) in one go as MDS plot

Cages pairwise (one example)



Outputs of Beta-Diversity.R

The script „Taxonomic-Binning.R“

① a new tab will open in Rstudio

② Change the path

③ Normally, no change necessary

④ Select all and run

⑤ a new folder was created

```

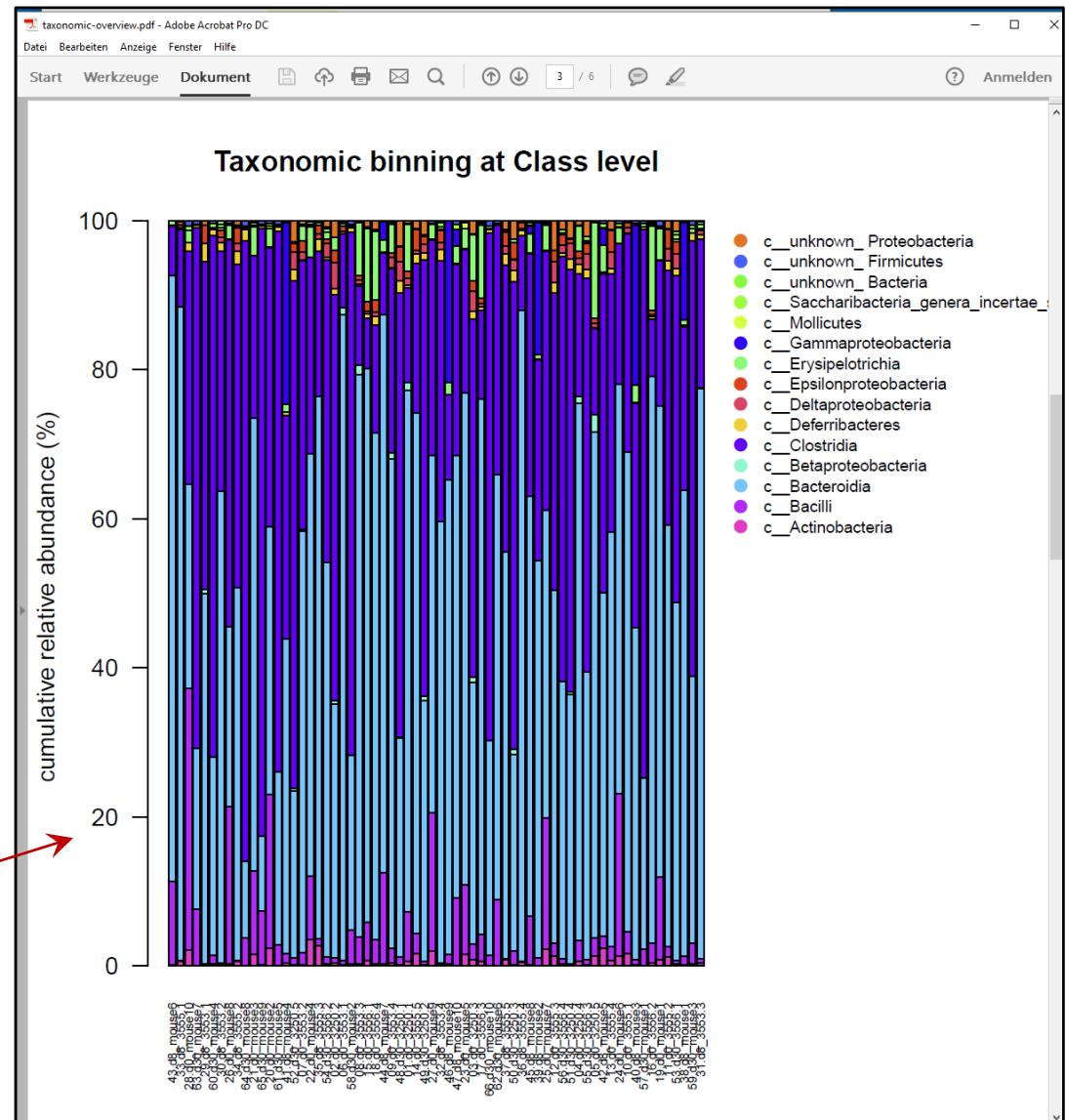
# 2. write the name of the examined otu file
# 3. Output: The script is generating seven tab-delimited files
# 4. 1. Relative taxonomic abundance at the kingdom level for each sample
# 5. 2. Relative taxonomic abundance at the phyla level for each sample
# 6. Relative taxonomic abundance at the class level for each sample
# 7. Relative taxonomic abundance at the order level for each sample
# 8. Relative taxonomic abundance at the family level for each sample
# 9. Relative taxonomic abundance at the genera level for each sample
# 10. Relative taxonomic abundance at all taxonomic levels for each sample
# 11. Graphical output:
# 12. Distribution of taxonomic relative abundances across all taxonomic groups for all samples
# 13. Concept:
# 14. Taxonomic information is split into the different taxonomic levels for each OTU
# 15. Individual abundance values of OTUs belonging to the same taxonomic group are summed up
# 16. Individual taxonomic composition for each sample is generated
# 17. Note:
# 18. If taxonomic information at a specific level is missing, the entry is replaced by
# 19. the last available taxonomic level including the prefix "unknown"
# 20. #####
# 21. Set parameters in this section manually #####
# 22. #####
# 23. Note:
# 24. If taxonomic information at a specific level is missing, the entry is replaced by
# 25. the last available taxonomic level including the prefix "unknown"
# 26. Individual taxonomic composition for each sample is generated
# 27. #####
# 28. Note:
# 29. If taxonomic information at a specific level is missing, the entry is replaced by
# 30. the last available taxonomic level including the prefix "unknown"
# 31. #####
# 32. ##### Please set the directory of the script as the working folder C:/Users/User/Desktop/Kopie Rhea/4.Taxonomic-Binning #####
# 33. Note: the path is denoted by forward slash "/"
# 34. setwd("C:/Users/User/Desktop/kopie Rhea/4.Taxonomic-Binning") --- CHANGE ORIGINALLY HERE
# 35. Please give the file name of the OTU-table containing relative abundances and taxonomic classification
# 36. otu_file<- "otus_Table-norm-rel-tax.tab" --- CHANGE ORIGINALLY HERE
# 37. "C:/Users/User/Desktop/koenig_rheinbeck_16S_RawData/OTUTables/OTUtable_norm_rel_tax.tab"
# 38. "C:/Users/User/Desktop/koenig_rheinbeck_16S_RawData/OTUTables/OTUtable_norm_rel_tax.tab"
# 39. "C:/Users/User/Desktop/koenig_rheinbeck_16S_RawData/OTUTables/OTUtable_norm_rel_tax.tab"
# 40. "C:/Users/User/Desktop/koenig_rheinbeck_16S_RawData/OTUTables/OTUtable_norm_rel_tax.tab"
# 41. "C:/Users/User/Desktop/koenig_rheinbeck_16S_RawData/OTUTables/OTUtable_norm_rel_tax.tab"
# 42. "C:/Users/User/Desktop/koenig_rheinbeck_16S_RawData/OTUTables/OTUtable_norm_rel_tax.tab"
# 43. "C:/Users/User/Desktop/koenig_rheinbeck_16S_RawData/OTUTables/OTUtable_norm_rel_tax.tab"
# 44. ##### NO CHANGES ARE NEEDED BELOW THIS LINE #####
# 45. ##### Main Script #####
# 46. #####
# 47. ##### Read input table #####
# 48. #####
# 49. #####
# 50. ##### Load the tab-delimited file containing the abundances and taxonomic information to be checked (rownames in the first column)
# 51. otu_table <- read.table(otu_file,
# 52. check.names = FALSE,
# 53. header = TRUE,
# 54. dec = ".",
# 55. sep = "\t",
# 56. row.names = 1,
# 57. na.strings = "unknown")
# 58. <-
# 59. <
# 3860 (Untitled) : R Script
# Console Terminal
# C:/Users/User/Desktop/Kopie Rhea/4.Taxonomic-Binning/Taxonomic-Binning/
# + c_colsample(c_col), cex.names=0.5,ylab="cumulative relative abundance (%)",las=2, main="Taxonomic binning at Class level")
# + legend(par('usr')[2], par('usr')[4], bty="n",rev(row.names(classes)),cex=0.7,col = rev(c_col),pch = 15,pt.cex = 1.2)
# +
# + #Orders
# + o_col=rainbow(dim(orders)[1])
# + o_col=sample(o_col)
# + barplot(orders,col=o_col, cex.names=0.5,ylab="cumulative relative abundance (%)",las=2, main="Taxonomic binning at Order level")
# + legend(par('usr')[2], par('usr')[4], bty="n",rev(row.names(orders)),cex=0.7,col = rev(o_col),pch = 16,pt.cex = 1.2)
# +
# + #Families
# + f_col=rainbow(dim(families)[1])
# + f_col=sample(f_col)
# + barplot(families,col=f_col, cex.names=0.5,ylab="cumulative relative abundance (%)",las=2, main="Taxonomic binning at Family level")
# + legend(par('usr')[2], par('usr')[4], bty="n",rev(row.names(families)),cex=0.7,col = rev(f_col),pch = 16,pt.cex = 1.2)
# +
# + #Genera
# + g_col=rainbow(dim(genera)[1])
# + g_col=sample(g_col)
# + barplot(genera,col=g_col, cex.names=0.5,ylab="cumulative relative abundance (%)",las=2, main="Taxonomic binning at Genus level")
# + legend(par('usr')[2], par('usr')[4], bty="n",rev(row.names(genera)),cex=0.7,col = rev(g_col),pch = 16,pt.cex = 1.2)
# +
# + dev.off()
# +
null device
1
> #####
# End of script #####
# #####
# | 4.Taxonomic-Binning
# |
# | Datei Start Freigeben Ansicht
# | < > & & D Dieser PC > Desktop > Kopie Rhea > 4.Taxonomic-Binning & "4.Taxonomic-Binning" durch... & i
# | Name Änderungsdatum Typ Größe
# | Desktop 20.04.2018 12:45 Dateordner
# | Downloads 20.04.2018 11:40 TAB-Datei 290 KB
# | Dokumente 19.01.2018 17:17 Adobe Acrobat-D... 105 KB
# | Bilder 19.01.2018 17:17 R-Datei 14 KB
# | 3.Beta-Diversity
# | Cage
# Global Environment
# meta
# meta_file
# orders
# otu_table
# otufile
# paired_dist
# pairedMatrixList
# phylo
# rooted_tree
# sample_list
# sub_sample_tax
# tax_summary
# taxonomy_new
# tree
# tree_file
# unifrac_dist
# unifrac_dist_comp
# Values
# a
# all_dist_matrix
# 
```

Results of „Taxonomic-Binning.R“

4.Taxonomic-Binning			
	Name	Änderungsdatum	Typ
✓ Schnellzugriff	Taxonomic-Binning	20.04.2018 12:45	Dateiordner
Desktop	OTUs_Table-norm-rel-tax.tab	20.04.2018 11:40	TAB-Datei
Downloads	Taxonomic Binning Script ReadMe.pdf	19.01.2018 17:17	Adobe Acrobat-Dokument
Dokumente	Taxonomic-Binning.R	19.01.2018 17:17	R-Datei
Bilder			
3.Beta-Diversity			
Cage			

Taxonomic-Binning			
	Name	Änderungsdatum	Typ
✓ Schnellzugriff	0.Kingdom.all.tab	20.04.2018 12:45	TA
Desktop	1.Phyla.all.tab	20.04.2018 12:45	TA
Downloads	2.Classes.all.tab	20.04.2018 12:45	TA
Dokumente	3.Orders.all.tab	20.04.2018 12:45	TA
Bilder	4.Families.all.tab	20.04.2018 12:45	TA
3.Beta-Diversity	5.Genera.all.tab	20.04.2018 12:45	TA
Cage	tax.summary.all.tab	20.04.2018 12:45	TA
	taxonomic-overview.pdf	20.04.2018 12:45	Ad

The taxonomic groups are binned for each mouse sample according to taxonomic level (genus, family, order, phyla, class, kingdom; shown is class level). All data are also found in the tables.



Before proceeding with either „Serial-Group-Comparisons.R“ or „Over-Time-Serial-Comparisons.R“, an input table must be created using „Create_Input_Table.R“.

The image illustrates the steps to prepare the 'Create_Input_Table.R' script for execution:

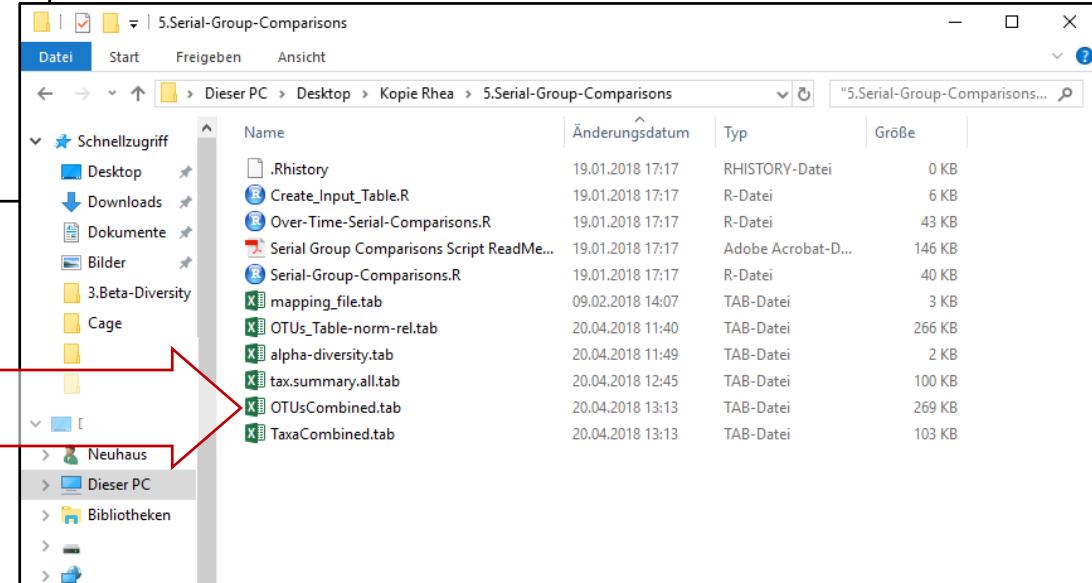
- ①** copy missing file into folder
- ②** start Create_Input_Table.R
- ③** a new tab will open in Rstudio
- ④** Change the path
- ⑤** Select all and run

Annotations in the RStudio window:

- "Change the path": Points to the line `setwd("C:/users/user/Desktop/kopie Rhea/5.Serial-Group-Comparisons")` in the script.
- "Normally, no change necessary": Points to the line `alpha <- "alpha-diversity.tab";`
- "Normally, no change necessary": Points to the line `OTUS_Table-norm-rel.tab`
- "Normally, no change necessary": Points to the line `TaxonomyAll <- "tax.summary.all.tab";`
- "Normally, no change necessary": Points to the line `Metafile <- "mapping_file.tab";`
- "Normally, no change necessary": Points to the line `main Script`
- "Load all required libraries": Points to the line `library("compar")`
- "Check if required packages are already installed, and install if missing": Points to the line `if (!requireNamespace("compar", quietly = TRUE)) {`
- "Function to check whether the package is installed": Points to the line `existsPack <- function(pack)`
- "if (existsPack(pack) == FALSE) {`": Points to the line `if (existsPack(pack) == FALSE) {`
- "install.packages(pack, repos = "http://cloud.r-project.org/")`": Points to the line `install.packages(pack, repos = "http://cloud.r-project.org/")`
- "Applying the installation on the list of packages": Points to the line `}`

Results of „Create_Input_Table.R“.

script has run



two new files: OTUsCombined.tab & TaxaCombined.tab

Starting „Serial-Group-Comparisons.R“

RStudio

File Edit Code View Plots Session Build Debug Profile Tools Help

install_packages.R* Normalization.R* Alpha-Diversity.R* Beta-Diversity.R* Taxonomic-Binning.R* Create_Input_Table.R* Serial-Group-Comparisons.R

Source on Save Run Source

```
25 ##### Set parameters in this section manually #####
26 #' Please set the directory of the script as the working folder (e.g D:/studynome/NGS-Data/Rhea/beta-diversity/)
27 #' Note: the path is denoted by forward slash "/"
28 setwd("D:/path/to/Rhea/5.serial-Group-Comparisons/") #< Change the path (not yet changed here) ②
29 #' Please give the file name generated by the Create_Input_Table.R script
30 input_filename = "OTUSCombined.tab" #< CHANGE ACCORDINGLY !!! ③
31 #' The name of the independant variable that the analysis will be based on
32 independant_variable_name <- "Diet"
33 #' Please enter the position in the table (column number) where the dependant variable starts
34 #' Note: the first column containing sample names does not count!
35 dependant_variables_start <- 4
36 #' Please enter the order of the group names
37 #' If no group names are writing groups are ordered automatically
38 group_order<-("") ④ See next page
39 #' Please enter the position in the table (column number) where relative abundances of OTUs or taxonomic groups start
40 #' Note: the first column containing sample names does not count!
41 taxonomic_variables_start <- 14
42
43 #' The cutoff of relative abundance; all values below this cutoff will be zeroed (default cutoff is 0.5 %)
44 abundance_cutoff <- 0.5
45
46 #' The prevalence cutoff; at least one group must have a number of samples above the selected threshold
47 #' for the variable to be tested (default cutoff is 0.3 = 30 % of samples are positive within a given group)
48 prevalence_cutoff <- 0.3
49
50 #' The minimum median abundance value that must be observed in at least one group before statistical test is performed
51 max_median_cutoff <- 1
52
53 #' Replace 0 value with NA
54 #' YES: Replace zeros with NA (default)
55 #' NO: Consider zeros in statistics
56 ReplaceZero = "YES" ⑤ Please read the ReadMe for all other settings
57
58 #' Set the graphical output parameter
59 #' 1 = without individual values as dots
60 #' 2 = with individual values as dots
61 #' 3 = with individual values as dots and with sample names
62 Plotoption = 1
63
64 #' set the significance cutoff level (default cutoff is 0.05 but it can be set lower)
65 sig_cutoff <- 0.05
66
67 ##### NO CHANGES ARE NEEDED BELOW THIS LINE #####
68
69
70
71
72
73
74
75
76
```

① a new tab will open in Rstudio

② Change the path (not yet changed here)

③ Normally, no change necessary

④ Change to the group you want to analyze! Our mapping file contains mice grouped in cages; thus, insert „Cage“ (exactly as in the mapping file, case sensitive) – not changed yet.

⑤ See next page

⑥ Please read the ReadMe for all other settings

70:15 (Untitled) R Script

Console Terminal

Count columns to change the parameters in „Serial-Group-Comparisons.R“

taxonomic_variables_start <- 14

dependant_variables_start <- 4

Insert correct values in script
„Serial-Group-Comparisons.R“

	0	1	2	3	4	5	6	7	8	9	10	K	L
1	SampleID	Day	Mouse	Cage	Facility	cage+mouse	Richness	Shannon.eff	Simpson.eff	OTU_5	OTU_1618	OTU_12	
2	43.d8_mouse	8	6	1111	LSE	111106	145	20.39	9.61	25.0115985	3.77854529	10.186092	
3	33.d8_3555.1	8	1	3555	FS	35551	150	37.8	20.82	0	0		
4	28.d0_mouse	0	10	1111	LSE	111110	235	44.44	16.13	2.37562715	0.33851176	0.7132926	
5	63.d30_mou	30	7	1111	LSE	111107	227	60.24	26.96	0.84589923	0.12412652	0.4321441	
6	29.d8_3553.1	8	1	3553	FS	35531	189	52.13	29.94	0.05483823	0.01495588	0.0199411	
7	60.d0_mou	30	4	1111	LSE	111104	235	66.11	30.01	0.94401544	0.16023166	0.2335907	
8	30.d8_3553.2	8	2	3553	FS	35532	196	62.82	29.93	0.02044572	0		
9	26.d0_mous	0	8	1111	LSE	111108	231	62.31	26.62	3.87301587	0.59259259	0.5008818	
10	34.d8_3555.2	8	2	3555	FS	35552	165	73.39	44.01	0	0		
11	64.d30_mou	30	8	1111	LSE	111108	229	69.22	33.9	0.46174142	0.1121372	0.3693931	
12	21.d0_mous	0	3	1111	LSE	111103	243	35.96	14.47	2.0037155	0.35164544	0.7961783	
13	65.d30_mou	30	9	1111	LSE	111109	241	84.55	39.08	0.55324381	0.07567572	0.3499148	
14	20.d0_mous	0	2	1111	LSE	111102	242	54.31	22.8	2.36134254	0.3475363	0.8997857	
15	61.d30_mou	30	5	1111	LSE	111105	230	70.94	32.08	1.0881075	0.178917	0.6572461	
16	41.d8_mous	8	4	1111	LSE	111104	211	41.48	15.29	10.873703	1.71916252	4.5680604	
17	52.d30_3250.	30	5	3250	FS	32505	178	62.01	27.11	0	0		
18	07.d0_3553.2	0	2	3553	FS	35532	203	63.19	29.33	0	0	0.0102917	
19	44.d0_mous	0	4	1111	FC	111101	207	14.02	2.26747545	0.4260543	0.9098786		

```

37 independant_variable_name <- "Diet"
38
39 #' Please enter the position in the table (column number) where the dependant
40 #' Note: the first column containing sample names does not count!
41 dependant_variables_start <- 7
42
43 #' Please enter the order of the group names
44 #' If no group names are writing groups are ordered automatically
45 group_order=c("")
46
47 #' Please enter the position in the table (column number) where relative abun
48 #' Note: the first column containing sample names does not count!
49 taxonomic_variables_start <- 10
50

```

A	B	C	D	E
1	43.d8_mouse	33.d8_3555.128.d0_mouse	63.d30_mou	29.d8
2	OTU_5	#####	0	#####
3	OTU_1618	#####	0	#####
4	OTU_12	#####	0	#####
5	OTU_1373	#####	0	#####
6	OTU_2805	#####	0	#####
7	OTU_1273	#####	0	#####
8	OTU_9	#####	0	#####
9	OTU_1121	#####	0	#####
10	OTU_1130	#####	0	#####
11	OTU_4059	#####	0	#####
12	OTU_3	#####	0	#####
13	OTU_4660	#####	0	0.5574

NOTE: The tab files use for the decimal point a dot (.) as separator and for the thousands separator a comma (,). However, German Excel Version use the comma (,) as decimal and the dot (.) as thousands separator, respectively.

Open an Excel file.
Goto File / Datei
Goto Options / Optionen
Goto Advanced / Erweitert
uncheck „Take separators from the system“ / „Trennzeichen vom Betriebssystem übernehmen“
insert . as decimal separator
insert , as thousands separator

Run „Serial-Group-Comparisons.R“ after setting all parameters

Select all and run

```

25 ##### In this section manually #####
26 # Please set the directory of the script as the working folder (e.g D:/studymane/NGS-Data/khea/beta-diversity/)
27 # Note: the path is denoted by forward slash "/"
28 #----- CHANGE ACCORDINGLY !!!
29 #----- CHOOSE THE PATH TO THE WORKING FOLDER
30 setwd("D:/path/to/Rhea/5.Serial-Group-Comparisons/")
31 #----- CHANGE ACCORDINGLY !!!
32 #----- CHOOSE THE PATH TO THE WORKING FOLDER
33 # Please give the file name generated by the Create_Input_Table.R script
34 input_filename = "OTUsCombined.tab" #<-- CHANGE ACCORDINGLY !!!
35 
36 # The name of the independent variable that the analysis will be performed on
37 independent_variable_name <- "Cage"
38 
39 # Please enter the position in the table (column number) where the dependant variable starts (e.g. richness)
40 # Note: The first column containing sample names does not count!
41 dependant_variables_start <- 7
42 
43 # Please enter the order of the group names
44 # If no group names are writing groups are ordered automatically
45 group_order<-c()
46 
47 # Please enter the position in the table (column number) where relative abundances of otus or taxonomic groups start
48 # Note: the first column containing sample names does not count!
49 taxonomic_variables_start <- 10
50 
51 # The cutoff of relative abundance; all values below this cutoff will be zeroed (default cutoff is 0.5%)
52 abundance_cutoff <- 0.5
53 
54 # The prevalence cutoff; at least one group must have a number of samples above the selected threshold
55 # For the variables to be tested (default cutoff is 0.3 = 30 % of samples are positive within a given group)
56 prevalence_cutoff <- 0.3
57 
58 # The minimum median abundance value that must be observed in at least one group before statistical test is performed
59 max_median_cutoff <- 0.05
60 
61 # Replace 0 Value with NA
62 # Replace zeros with NA (default)
63 # NO! consider zeros in statistics
64 ReplaceZero = "YES"
65 
66 # Set the graphical output parameter
67 # 1 = without individual values as dots
68 # 2 = with individual values as dots
69 # 3 = with individual values as dots and with sample names
70 Plotoption = 1
71 
72 # Set the significance cutoff level (default cutoff is 0.05 but it can be set lower)
73 sig_cutoff <- 0.05
74 
75 ##### NO CHANGES ARE NEEDED BELOW THIS LINE #####
76 <

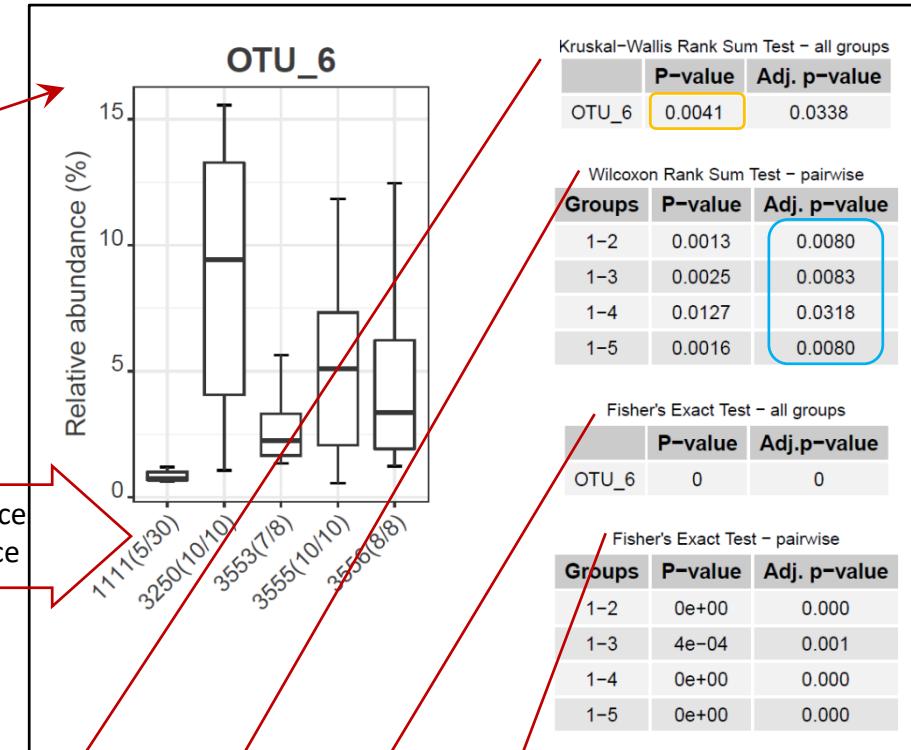
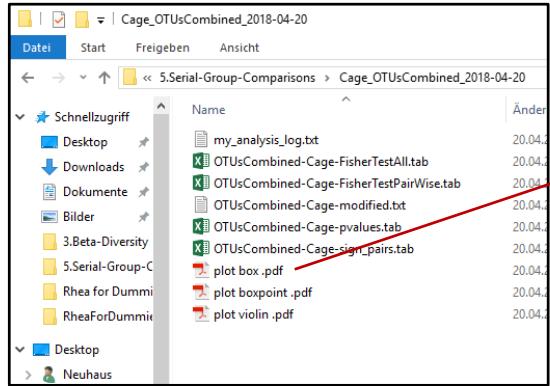
```

a new folder is created

Name	Änderungsdatum	Typ	Größe
Cage_OTUsCombined_2018-04-20	19.01.2018 17:17	RHISTORY-Datei	0 KB
Blhistory	19.01.2018 17:17	R-Datei	6 KB
Create_Input_Table.R	19.01.2018 17:17	R-Datei	43 KB
Over-Time-Serial-Comparisons.R	19.01.2018 17:17	Adobe Acrobat-Dokument	146 KB
Serial Group Comparisons Script ReadMe...	19.01.2018 17:17	R-Datei	40 KB
mapping_file.tab	09.02.2018 14:07	TAB-Datei	3 KB
OTUs_Table-norm-rel.tab	20.04.2018 11:40	TAB-Datei	266 KB
alpha-diversity.tab	20.04.2018 11:49	TAB-Datei	2 KB
tax.summary.all.tab	20.04.2018 12:45	TAB-Datei	100 KB
OTUsCombined.tab	20.04.2018 13:13	TAB-Datei	269 KB
TaxaCombined.tab	20.04.2018 13:13	TAB-Datei	103 KB

Name	Änderungsdatum	Typ	Größe
my_analysis_log.txt	20.04.2018 13:46	Textdokument	1 KB
OTUsCombined-Cage-FisherTestAll.tab	20.04.2018 13:46	TAB-Datei	2 KB
OTUsCombined-Cage-FisherTestPairWise.tab	20.04.2018 13:46	TAB-Datei	27 KB
OTUsCombined-Cage-modified.txt	20.04.2018 13:46	Textdokument	45 KB
OTUsCombined-Cage-pvalues.tab	20.04.2018 13:46	TAB-Datei	2 KB
OTUsCombined-Cage-sign_pairs.tab	20.04.2018 13:46	TAB-Datei	8 KB
plot box.pdf	20.04.2018 13:46	Adobe Acrobat-Dokument	162 KB
plot boxpoint.pdf	20.04.2018 13:46	Adobe Acrobat-Dokument	160 KB
plot violin.pdf	20.04.2018 13:46	Adobe Acrobat-Dokument	1.014 KB

Results of „Serial-Group-Comparisons.R“



p-values are either without (P-value) or with (Adj. p-value) correction for multiple testings.

The appropriate p-value should be selected by you, based on your hypothesis / experiment.

OTU_6 is only present in 5 of 30 mice in cage 1111 and has low abundance

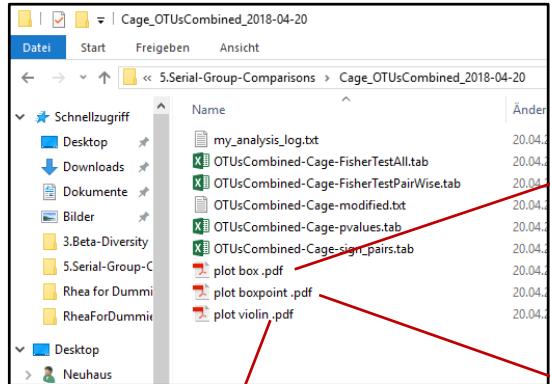
Kruskal-Wallis Rank Sum: Difference in relative **abundance** ("overall within all groups there is at least one group different to another one")

Wilcoxon Rank Sum Test – pairwise:
Difference in relative **abundance** between pairs of groups

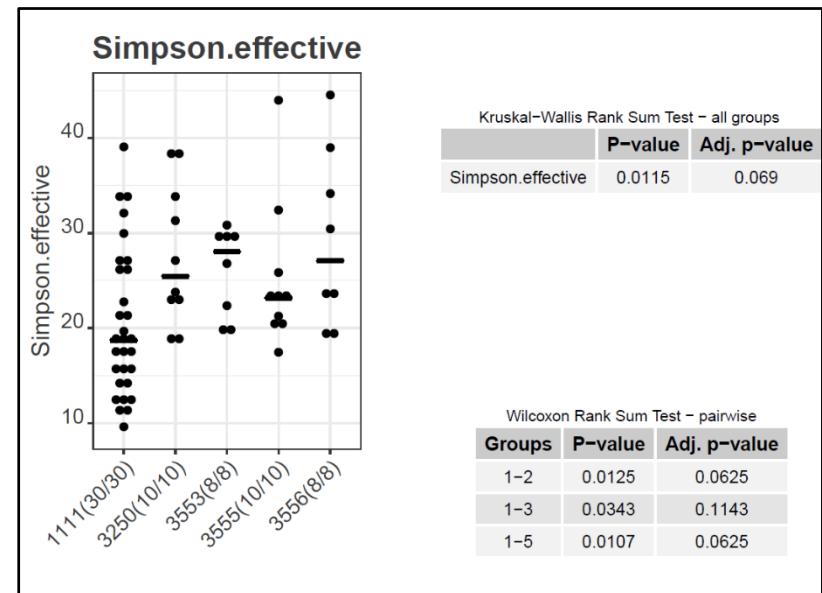
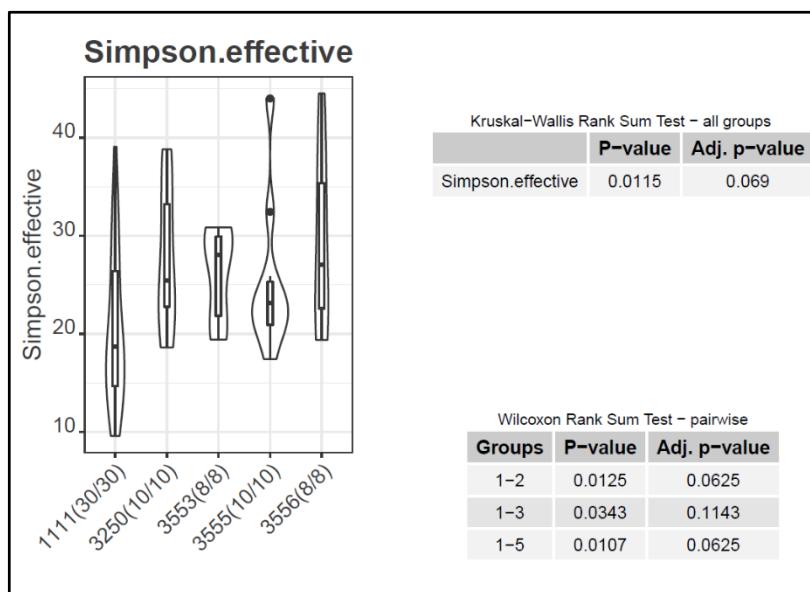
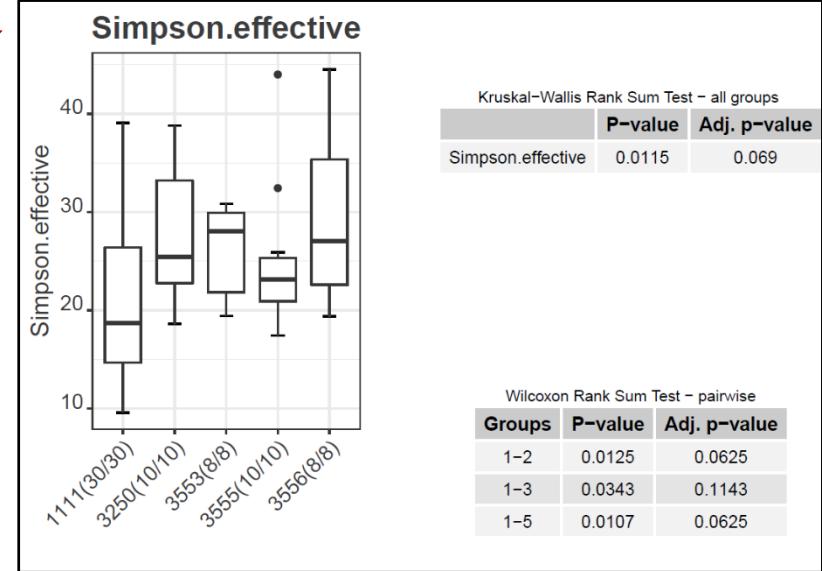
Fisher's Exact Test – pairwise:
Difference in relative **prevalence** between pairs of groups

Fisher's Exact Test - all groups: Difference in relative **prevalence** ("overall within all groups there is at least one group different to another one")

Results of „Serial-Group-Comparisons.R“



only one example is shown



Starting „Correlations.R“

RStudio
File Edit Code View Plots Session Build Debug Profile Tools Help
Source on Save Go to file/function Addins

install_packages.R* Normalization.R* Alpha-Diversity.R* Beta-Diversity.R* Taxonomic-Binning.R* Create_Input_Table.R* Serial-Group-Comparisons.R* Correlations.R* Run Source

```
25 #' 3. The asymptotic p-value
26 #' 4. Number of observations used for each pair of variables
27 #' 5. Pairwise information about correlation coefficients, p-value, number of observations, and corrected p-value
28 #' 6. Corrected table with pairwise information (p-value < 0.05)
29 #' 7. Graphical illustration of correlation matrix (with and without significance cutoff)
30 #' 8. Pairwise linear model for log-transformed data
31 #
32 #' Concept:
33 #' Calculate Pearson correlation coefficients
34 #' Significant pairs are determined by p-value (default cutoff is 0.05)
35 #' Graphical output is generated if absolute correlation value is >0.5
36 #
37 #' Note:
38 #' It is important to take into account the number of observations for the interpretation of results!
39
40 ##### Set parameters in this section manually #####
41 #
42 #
43
44 #' Please set the directory of the present script as the working folder for the R script (correlation/).
45 #' Note: the path is denoted by forward slash (/) as the separator
46 setwd("D:/path/to/Rhea/6.Correlations")
47
48 #' Please give the file name of the table containing the variables for analysis
49 input_file <- "OTUScombined_Corr_input_table.tab"
50
51 #' Please give the position where the taxonomic variables (OTUS or taxonomic groups) start!!
52 #' IMPORTANT: Since the first column in the input file will be used as row names, we do not count it!
53 otu_variables_start <- 10
54
55 ##### Optional parameters in this section #####
56 #
57 #
58 # unless users follow specific purposes and know exactly what to test, default values are recommended.
59
60 # Set the cutoff for significance
61 # Possible values are any real number between 0 and 1 (default is 0.05)
62 signif_cutoff <- 0.05
63
64 # calculate correlation among taxonomic variables
65 # If selected, this will result in many additional tests for correlation among taxonomic variables
66 # Possible parameters are 1 or 0 (default is 0):
67 # 1 = calculate correlations within OTUS or taxa
68 # 0 = NO test within taxonomic variables
69 includeTax <- 0
70
71 # Calculate correlation among meta-variables
72 # If selected, this will result in many additional tests for correlation among meta-variables
73 # Possible parameters are 1 or 0 (default is 0):
74 # 1 = calculate correlations within meta-variables
75 # 0 = NO test within meta-variables
```

① a new tab will open in Rstudio

② Change the path (not yet changed here)
Normally, no change necessary

③ count in table

Please read the ReadMe for all other settings

column counting

Richness

SampleID	Day	Mouse	Cage	Facility	cage+mouse	Richness	Shannon.eff	Simpson.eff	OTU_5	OTU_1618	OTU_12	OTU_13
43.d8_mouse	8	6	1111	LSE	111106	145	20.39	9.61	25.0115985	3.77854529	10.1860921	1.3763
33.d8_3555.1	8	1	3555	FS	35551	150	37.8	20.82	0	0	0	0
28.d0_mouse	0	10	1111	LSE	111110	235	44.44	16.13	2.37562715	0.33851176	0.71329263	0.1083
63.d30_mouse	30	7	1111	LSE	111107	227	60.24	26.96	0.84589923	0.12412652	0.43214417	0.0459
29.d8_3553.1	8	1	3553	FS	35531	189	52.13	29.94	0.05483823	0.01495588	0.01994117	0.0049
60.d30_mouse	30	4	1111	LSE	111104	235	66.11	30.01	0.94401544	0.16023166	0.23359073	0.0444
30.d8_3553.2	8	2	3553	FS	35532	196	62.82	29.93	0.02044572	0	0	0

Run „Correlations.R“

① a new tab will open in Rstudio

② Select all and run

warnings are ignored

two new files are created

The screenshot shows the RStudio interface with the 'Correlations.R' script open in the editor. The script contains R code for performing correlation analysis on OTU data. The RStudio environment pane shows various objects like 'a', 'clm', 'corr_pval_cutoff', etc. The global environment pane lists variables such as 'a', 'clm', 'corr_pval_cutoff', 'lib', 'matrix_names', 'my_corr_matrix', 'my_data', 'my_meta_data', 'my_meta_fixed', 'my_m_matrix', 'my_n_matrix', 'my_otu_fixed', 'my_pairs', 'my_pval_matrix', 'my_rcorr', 'my_scaled_data', 'pairs', 'transformed_data', and 'x'. The console pane shows the execution of the script, including the creation of a new directory '6.Correlations' and the generation of several PDF files: 'Correlations Script ReadMe.pdf', 'corplot.pdf', and 'linear_sign_pairs.pdf'. The file browser pane shows the contents of the '6.Correlations' folder, which includes the script file 'Correlations.R', the generated PDFs, and other files like 'install_packages.R', 'LICENSE', 'README.md', and 'missing_packages.txt'. Red annotations with arrows point to specific elements: a red circle around the 'Correlations.R' tab, a red arrow pointing to the 'Run' button, a red arrow pointing to the 'Run' button, a red arrow pointing to the warning messages in the console, and a red arrow pointing to the file browser.

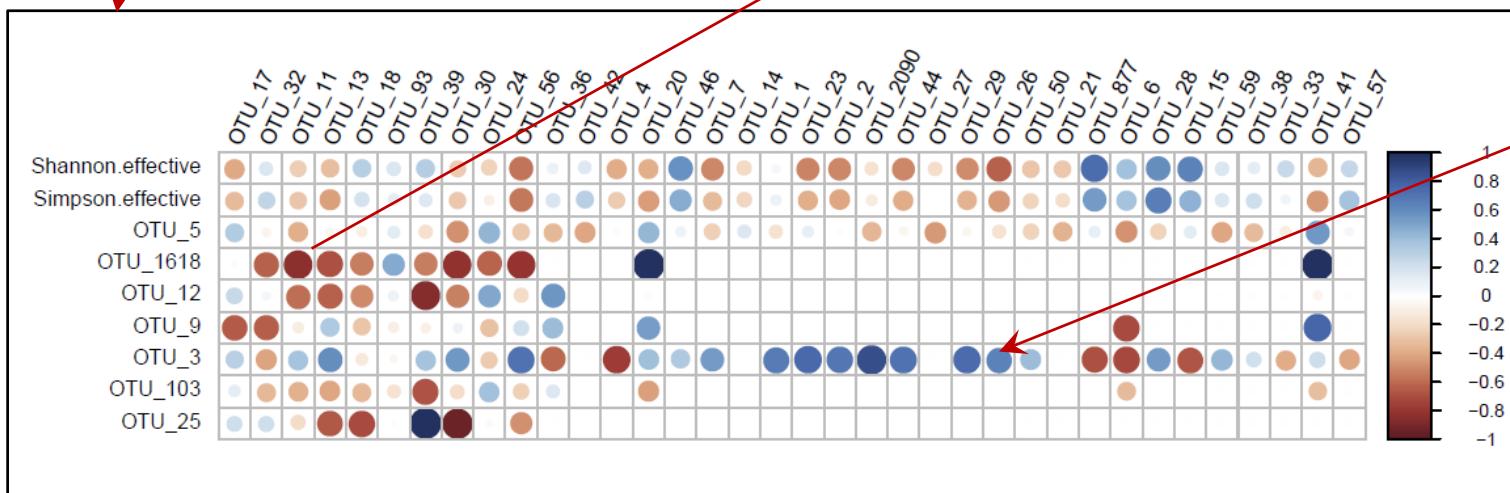
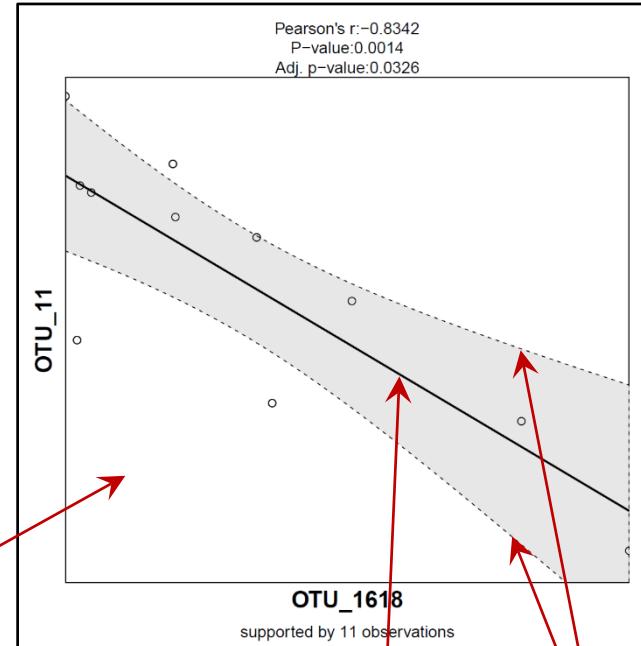
```
25 # 3. The asymptotic p-value
26 # 4. Number of observations used for each pair of variables
27 # 5. Pairwise information about correlation coefficients, p-value, number of observations, and corrected p-value
28 # 6. Corrected table with pairwise information (p-value < 0.05)
29 # 7. Graphical illustration of correlation matrix (with and without significance cutoff)
30 # 8. Pairwise linear model for log-transformed data
31 #
32 # Concept:
33 # calculate Pearson correlation coefficients
34 # Significant pairs are determined by p-value (default cutoff is 0.05)
35 # Graphical output is generated if absolute correlation value is >= 0.5
36 #
37 # Note:
38 # It is important to take into account the number of observations for the interpretation of results!
39 #
40 #####
41 #####
42 #####
43 #####
44 # Please set the directory of the present script as the working folder (e.g. D:/studynome/NGS-Data/Rhea/correlation/)
45 # Note: the path is denoted by forward slash "/"
46 setwd("C:/Users/User/Desktop/kopie Rhea/6.correlations") #<-- CHANGE ACCORDINGLY !!!
47 #
48 # Please give the file name of the table containing the variables for analysis
49 input_file <- "OTUscombined_Corr_Input_table.tab" #<-- CHANGE ACCORDINGLY !!!
50 #
51 # Please give the position where the taxonomic variables (OTUs or taxonomic groups) start!!!
52 # IMPORTANT: Since the first column in the input file will be used as row names, we do not count it!
53 otu_variables_start <- 10 #<-- CHANGE ACCORDINGLY !!!
54 #
55 #####
56 ##### optional parameters in this section #####
57 #####
58 # Unless users follow specific purposes and know exactly what to test, default values are recommended.
59 #
60 # set the cutoff for significance
61 # Possible values are any real number between 0 and 1 (Default is 0.05)
62 signif_cutoff <- 0.05
63 #
64 # Calculate correlation among taxonomic variables
65 # If selected, this will result in many additional tests for correlation among the taxonomic data
66 # Possible parameters are 1 or 0 (Default is 0):
67 # 1 = calculate correlations within OTUs or taxa
68 # 0 = NO test within taxonomic variables
69 includeTax <- 0
70 #
71 # Calculate correlation among meta-variables
72 # If selected, this will result in many additional tests for correlation among the meta-data
73 # Possible parameters are 1 or 0 (Default is 0):
74 # 1 = calculate correlations within meta-variables
75 # 0 = NO test within meta-variables
76 #
77 # Centre and scale the values
78 my_scaled_data <- scale(transformed_data, center = TRUE, scale = TRUE)
79 #
80 # calculate all pairwise correlations using Pearson correlation method
81 my_rcorr <- cor(my_scaled_data, type = "pearson")
82 Warning messages:
83 1: In sqrt(npair - 2) : Nans produced
84 2: In sqrt(1 - h) : Nans produced
85 3: In pta(b(h)) * sqrt(npair - 2) / sqrt(1 - h), npair - 2 : 
NANS wurden erzeugt
86 #
87 # Generate vector with variable names
88 var_names <- row.names(my_rcorr)
89 #
90 # Depending on which parameters were set at the beginning, one query type is selected
91 # In each query type, three matrices are generated: p-value matrix, correlation matrix, support matrix
92 # All possible pairs are saved in a vector (pairs)
93 if(includeTax == 1 & includeMeta == 0){
94   #
95   # Correlation among OTUs and NO correlation among meta-variables
96   # row_names <- var_names[c(otu_variables_start:dim(my_rcorr)[1])]
97   # col_names <- var_names
98   # pairs <- expand.grid(row_names, col_names)
99   # my_corr_matrix <- my_rcorr[,c(otu_variables_start:dim(my_rcorr)[1]),]
100  # my_pval_matrix <- my_rcorr[,c(otu_variables_start:dim(my_rcorr)[1]),]
101  # my_supp_matrix <- my_rcorr[,c(otu_variables_start:dim(my_rcorr)[1]),]
```

Results „Correlations.R“

6.Correlations			
	Name	Änderungsdatum	Typ
Desktop	OTUSCombined_Corr_input_table_2018-0...	20.04.2018 14:16	Dateiordner
Downloads	Correlations Script ReadMe.pdf	19.01.2018 17:17	Adobe Acrobat-D...
Dokumente	Correlations.R	19.01.2018 17:17	R-Datei
Bilder	corplot.pdf	20.04.2018 14:16	Adobe Acrobat-D...
5.Serial-Group-C	linear_sign_pairs.pdf	20.04.2018 14:16	Adobe Acrobat-D...
6.Correlations	OTUSCombined_Corr_input_table.tab	20.04.2018 14:15	TAB-Datei

Shows which parameter or OTU has a negative or positive correlation (see color scale to the right)

Shows the actual data points (observations) on a dimensionless scale for the correlation of OTU_11 and OTU_1618



The size of the circle gives the significance and the darkness of the color indicates the strength of the correlation according the the color code shown.

Build a better tree / taxonomy

① Goto: <https://www.arb-silva.de/aligner/>

The screenshot shows the ARB-SILVA Alignment Service interface. At the top, there's a logo for silva (high quality ribosomal RNA databases) and de.NBI (GERMAN NETWORK FOR BIOINFORMATICS INFRASTRUCTURE). Below the header, there are links for Home, SILVAngs, Browser, Search, ACT, Download, Documentation, Projects, FISH & Probes, and Contact. The main title is "ACT: Alignment, Classification and Tree Service".
The interface includes:

- An input area with a text box for "Paste your FASTA sequence here" and a "upload an FASTA file" button.
- "Basic alignment parameters" section with dropdowns for Gene (set to SSU), Min. identity with query sequence (set to 0.95), and Number of neighbours (set to 10).
- A "Compute tree" section with dropdowns for Program (FastTree), Model (GTR), and Rate model (Gamma).
- "Output settings" section with options for Format (FASTA selected), Compression (zip selected), and Reject sequences below identity (set to 70%).
- An "Advanced alignment parameters" section.
- A "Job Name" input field set to "unnamed aligner job" and buttons for "Reset Settings" and "Run Aligner".
- An "Aligner Taskmanager" section showing a table with columns: #, Job Name, Creation Time, Job Type, Status, Quantity, Progress, Status Message, Elapsed Time, and Queue. It displays "none found...".

The screenshot shows the ACT: Alignment, Classification and Tree Service interface. At the top, there are links for Home, SILVAngs, Browser, Search, ACT (highlighted in orange), Download, Documentation, and Project. The main title is "ACT: Alignment, Classification and Tree Service".
The interface includes:

- An input area with a text box containing a long DNA sequence and a "upload an FASTA file" button.
- "Basic alignment parameters" section with a "Search and classify" checkbox (unchecked).
- A "Tip: hovering over the options shows enhanced descriptions." note.

The screenshot shows the ACT: Alignment, Classification and Tree Service interface. The main title is "ACT: Alignment, Classification and Tree Service".
The interface includes:

- An input area with a text box containing a long DNA sequence and a "upload an FASTA file" button.
- "Basic alignment parameters" section with a "Search and classify" checkbox (unchecked).
- "Advanced search and classification parameters" section:
 - A note: "The Sequence Collection that is used for search and classification:" with radio buttons for REF NR (selected), Ref, and SILVA.
 - A list of taxonomies for classification: RDP, greengenes, LTP, and EMBL, each with a checkbox. The "SILVA" checkbox is checked and highlighted with a red circle.
 - Parameter settings: search-kmer-candidates (1000), lca-quorum (0.8), search-kmer-len (10), search-kmer-mm (0), search-no-fast (unchecked), and search-kmer-noref (unchecked).
- A "Job Name" input field set to "unnamed aligner job" and buttons for "Reset Settings" and "Run Aligner".
- An "Aligner Taskmanager" section showing a table with columns: #, Job Name, Creation Time, Job Type, Status, Quantity, Progress, Status Message, Elapsed Time, and Queue. It displays "none found...".

① Results are finished

Basic alignment parameters

Job Name: unnamed aligner job

Aligner Taskmanager

#	Job Name	Creation Time	Job Type	Status	Quantity	Progress	Status Message	Elapsed Time	Queue
1		2018-05-08 15:17	Align	Finished	2	2		00:04:12	0

Alignment Result Table

Jobid	#	Sequence Identifier	Full Name	Identity	Score	Cutoff Head	Cutoff Tail	E.coli Pos.	Gene Bps	Turn
524494	1	1	None	95.37	97	0	0	347	431	none
524494	2	2	None	96.03	98	0	0	347	428	none

Export To CSV

② Please select a job

Job Name: unnamed aligner job

Reset Settings Run Aligner

Aligner Taskmanager

#	Job Name	Creation Time	Job Type	Status	Quantity	Progress	Status Message	Elapsed Time	Queue
1		2018-05-08 15:17	Align	Finished	2	2		00:04:12	0

Show result

Download file

Add neighbors to cart

Alignment Result Table

Jobid	#	Sequence Identifier	Full Name	Identity	Score	Cutoff Head	Cutoff Tail	E.coli Pos.	Gene Bps	Turn
524494	1	1	None	95.37	97	0	0	347	431	none
524494	2	2	None	96.03	98	0	0	347	428	none

When selecting a job,
you get an additional
menu

③ Download: choose zip

Align

... Align

... Align

Download file

Add neighbors to

zip (1 Kb)

log (8 Kb)

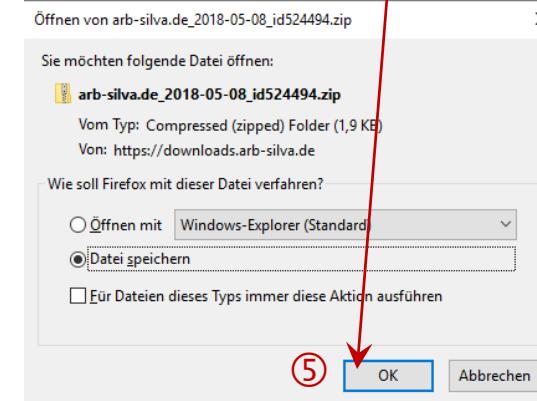
Display Scores

Display Classification

Full Name

Identity

Score



④ You find your zip in the download folder.

① In addition, download CSV

② Export To CSV

③ Öffnen von arb-silva.de_align_resultlist_524494.csv

Sie möchten folgende Datei öffnen:
arb-silva.de_align_resultlist_524494.csv
Vom Typ: Microsoft Excel-CSV-Datei
Von: https://www.arb-silva.de

Wie soll Firefox mit dieser Datei verfahren?
 Öffnen mit Microsoft Excel (Standard)
 Datei speichern
 Für Dateien dieses Typs immer diese Aktion ausführen

④ Downloads

arb-silva.de_align_resultlist_524494.csv
arb-silva.de_2018-05-08_id524494.zip

⑤ In the csv is your SiLVA taxonomy

⑥ In the zip is your aligned(!) fasta