

Preparatory Instruction

Tutorial: Beginner's guide to microbiome analysis

CSBio2020

Pichahpuk Uthaipaisanwong, Pantakan Puengrang, Chalida Rangsiwutisak,
Photchanathorn Prombun, Athrisi Sitthipunya, Natchaphon Rajudom and Kanthida Kusunmano

Systems Biology and Bioinformatics Research Group,
King Mongkut's University of Technology Thonburi, Bangkok, Thailand

This instruction is for the “Beginner's guide to microbiome analysis” tutorial, 11th International Conference on Computational Systems-Biology and Bioinformatics (CSBio2020). In order to follow the tutorial smoothly, we provide a preparatory instruction of required hardware, software, datasets, analysis outputs and codes that will be used along the tutorial.

1. Minimum system requirements

- Processor: Intel Core i5 or AMD equivalent
- Memory (RAM): 8 GB or higher
- Available storage (HDD): 20 GB
- Network accessibility
- Operating systems (OS): Windows 10 or higher, OSX 10.14.x or higher

2. Software requirements and installations

- Mothur v.1.44.3 (for microbiome analysis)
- RStudio v1.3-1093 and required R packages (for diversity analysis and visualization)
- Text editor e.g. Sublime, Visual Studio Code, etc. (for browsing resulting files)

3. Data and software preparation

3.1 Tutorial Github repository

We provide data and scripts for the tutorial in Github repository (Figure 1 and 2). You can either download all contents as ZIP file or use *git clone* command (once you have git installed). Please put the decompressed folder on your desktop (recommended).

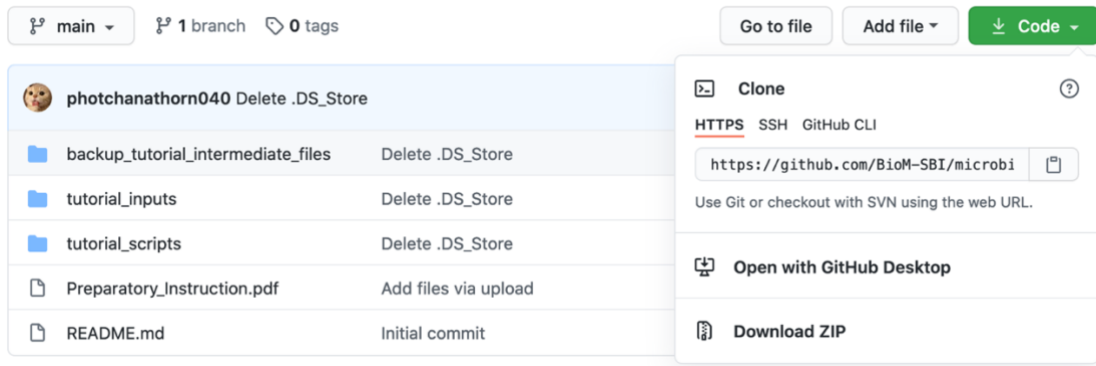


Figure 1. Github repository (https://github.com/BioM-SBI/microbiome_tutorial_csbio20) for the tutorial.

For git command, use the following command.

```
git clone https://github.com/BioM-SBI/microbiome_tutorial_csbio20.git
```

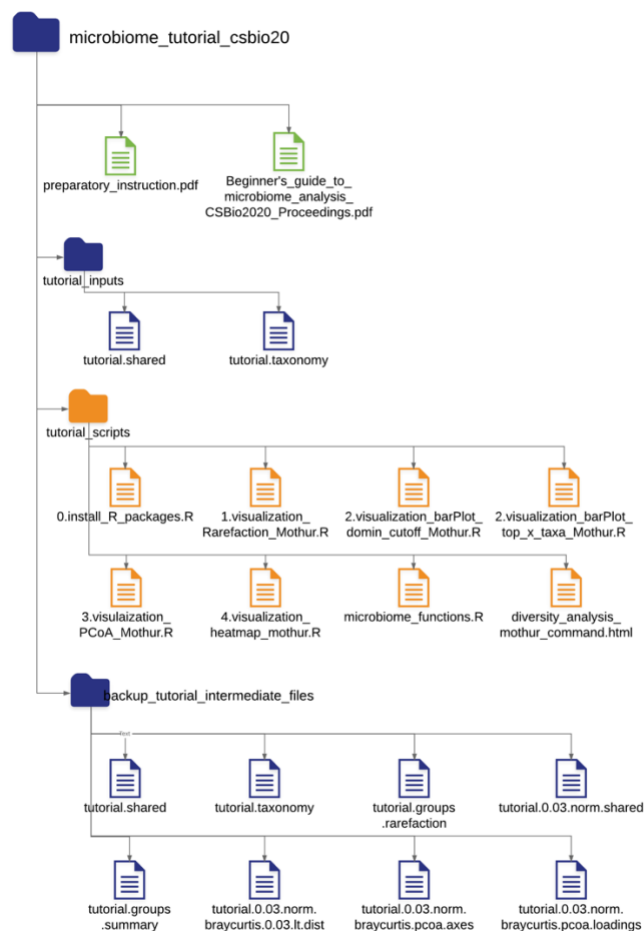


Figure 2. Tutorial files architecture. Blue icons indicate analysis files that will be produced/used during the microbiome analysis. Orange icons indicate scripts for the analysis.

3.2 Mothur [1]

1. Download `Mothur.xxx.zip` file based on your OS (Figure 3). The software is available at <https://github.com/mothur/mothur/releases/tag/v1.44.3>.

▼ Assets 11












 Mothur.linux.zip	19.2 MB
 Mothur.linux_noReadline.zip	19.2 MB
 Mothur.OSX-10.14.zip	24.9 MB
 Mothur.tools_linux.zip	12.4 MB
 Mothur.tools_OSX.zip	16.9 MB
 Mothur.tools_ubuntu.zip	12.4 MB
 Mothur.tools_win.zip	9.27 MB
 Mothur.Ubuntu_18.zip	18.9 MB
 Mothur.win.zip	13.8 MB
 Source code (zip)	
 Source code (tar.gz)	

Figure 3. Mothur software installation files. Underlines indicate the installation files for each OS; Linux, OSX, and Windows, respectively.

2. Decompress the downloaded zip file and put decompressed folder into the `microbiome_tutorial_csbio20` folder which is already downloaded from Github. The `mothur` file in decompressed folder could be used to execute the program (Figure 4).

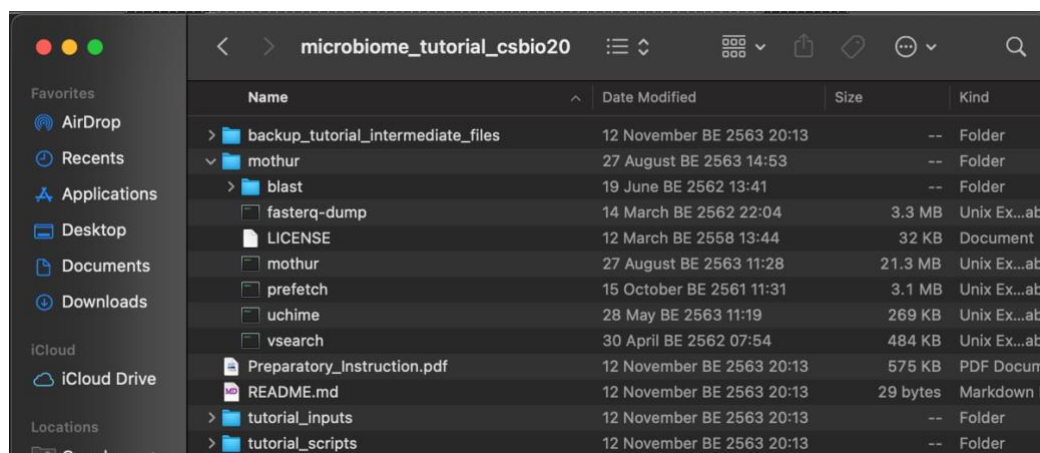


Figure 4. Decompress `mothur.xxx.zip` and put into `microbiome_tutorial_csbio20` folder. The `mothur` program could be executed directly.

3. For taxonomic assignment step, SILVA database could be used [2], in case users want to analyze data by themselves. The database for mothur software is available at https://mothur.s3.us-east-2.amazonaws.com/wiki/silva.nr_v132.tgz.

3.3 RStudio and required R packages

1. Before installing RStudio, please make sure that you already installed R. If not please download R from <http://mirrors.psu.ac.th/pub/cran/> and install it first based on your OS (Figure 5).

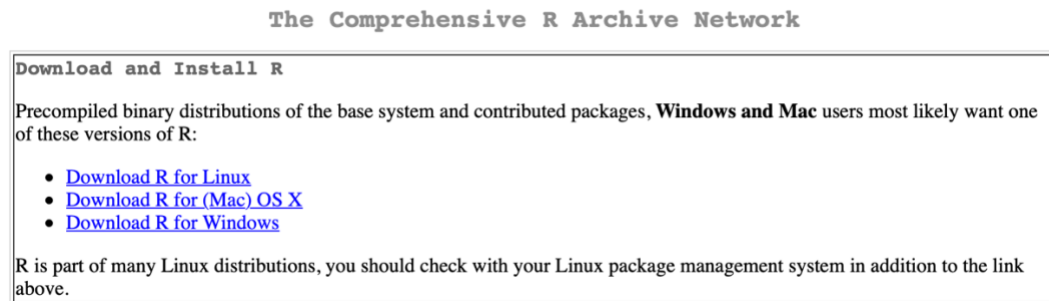


Figure 5. R installation files.

2. Download RStudio installation file and then install program based on your OS (Figure 6). The software available at <https://rstudio.com/products/rstudio/download/>.

All Installers

Linux users may need to [import RStudio's public code-signing key](#) prior to installation, depending on the operating system's security policy.

RStudio requires a 64-bit operating system. If you are on a 32 bit system, you can use an [older version of RStudio](#).

OS	Download	Size	SHA-256
Windows 10/8/7	RStudio-1.3.1093.exe	171.62 MB	62b9e60a
macOS 10.13+	RStudio-1.3.1093.dmg	148.66 MB	bdc4d3a4
Ubuntu 16	rstudio-1.3.1093-amd64.deb	124.33 MB	72f05048
Ubuntu 18/Debian 10	rstudio-1.3.1093-amd64.deb	126.80 MB	ff222177
Fedora 19/Red Hat 7	rstudio-1.3.1093-x86_64.rpm	146.96 MB	ed1f6ef8
Fedora 28/Red Hat 8	rstudio-1.3.1093-x86_64.rpm	151.05 MB	01a978f3
Debian 9	rstudio-1.3.1093-amd64.deb	127.00 MB	a747f9f9
SLES/OpenSUSE 12	rstudio-1.3.1093-x86_64.rpm	119.43 MB	5016cbcf
OpenSUSE 15	rstudio-1.3.1093-x86_64.rpm	128.40 MB	cf47e32d

Figure 6. RStudio software installation files. Underlines indicate the installation files for each OS; Windows, and OSX respectively.

3. After installing RStudio, please install all packages needed for data visualization which are ggplot2, dplyr, tidyr, stringr, RColorBrewer, pheatmap, gplots, vegan, and reshape2. For packages installation, we provide R script in `~/tutorial_scripts/0.install_R_packages.R` (Figure 7).

```
#!/usr/bin/env Rscript

#####
# Author: Pantakan Puengrang
# Modified Date: 21st October 2020
#####

#install R packages
install.packages("ggplot2")
install.packages("dplyr")
install.packages("tidyr")
install.packages("stringr")
install.packages("RColorBrewer")
install.packages("pheatmap")
install.packages("gplots")
install.packages("vegan")
install.packages("reshape2")

##### ----- Finish ----- #####
```

Figure 7. Preparatory script for packages installation (0.install_R_packages.R).

3.4 Text editors

A text editor will be used for browsing resulting files during the tutorial. Here we suggest to use Sublime Text or Visual Studio Code software. However, please feel free for any choices. Installation files for the suggested text editors are provided in the following links. Please install and follow the instruction of each software.

- Sublime Text: <https://www.sublimetext.com/3>.
- Visual Studio Code: <https://code.visualstudio.com/download>.

3.5 Datasets

Please download the dataset which will be used in this tutorial at <https://bit.ly/3p85JER>. It is microbiome data of 16S rRNA gene sequences sub-sampled from the study of microbial communities in wastewater treatment systems will be utilized in this tutorial [3]; Data was deposited in European Nucleotide Archive (ENA) under accession number ERP113548. The data is paired-end sequences of V3-V4 hypervariable regions based on Illumina sequencing platform. To provide less time and computational power through the tutorial session, the data were sub-sampled (30%) to reduce the size of the original dataset.

4. Analysis outputs

All microbiome analysis outputs are stored in github and could be downloaded from https://github.com/BioM-SBI/microbiome_tutorial_csbio20 (Figure 1). You can download entire repository using git clone or download compressed file (Figure 2). These are outputs that will be produced from the analyses. In case users cannot follow any step during the tutorial session, ones could jump for the output file from each step and continue the next step from backup_tutorial_intermediate_files folder.

5. Codes

All scripts for the microbiome analysis are stored in github and could be downloaded from https://github.com/BioM-SBI/microbiome_tutorial_csbio20 (Figure 1). You can download either entire repository or compressed file same as analysis outputs (Figure 2). Scripts can be found at tutorial_scripts folder.

References

1. Schloss, P.D., Westcott, S.L., Ryabin, T., Hall, J.R., Hartmann, M., Hollister, E.B., Lesniewski, R.A., Oakley, B.B., Parks, D.H., Robinson, C.J., Sahl, J.W., Stres, B., Thallinger, G.G., Van Horn, D.J., and Weber, C.F., 2009, "Introducing Mothur: Open-Source, Platform-Independent, Community-Supported Software for Describing and Comparing Microbial Communities", **Appl Environ Microbiol**, Vol. 75, No. 23, pp. 7537-7541.
2. Quast, C., Pruesse, E., Yilmaz, P., Gerken, J., Schweer, T., Yarza, P., Peplies, J., and Glockner, F.O., 2013, "The Silva Ribosomal Rna Gene Database Project: Improved Data Processing and Web-Based Tools", **Nucleic Acids Res**, Vol. 41, No. Database issue, pp. D590-596.
3. Puengrang, P., Suraraksa, B., Prommeenat, P., Boonapatcharoen, N., Cheevadhanarak, S., Tanticharoen, M., and Kusonmano, K., 2020, "Diverse Microbial Community Profiles of Propionate-Degrading Cultures Derived from Different Sludge Sources of Anaerobic Wastewater Treatment Plants", **Microorganisms**, Vol. 8, No. 2.