Preparatory Instruction
Tutorial: Beginner's guide to microbiome analysis
CSBio2020

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

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)

2.1 Mothur [1]
1. Download Mothur.zip file based on your OS (Figure 1). The software is available at https://github.com/mothur/mothur/releases/tag/v1.44.3.

Figure 1. Mothur software installation files. Underlines indicate the installation files for each OS; Linux, OSX, and Windows, respectively.

2. Decompress the downloaded zip file. The 'mothur' file in decompressed folder could be used to execute the program.
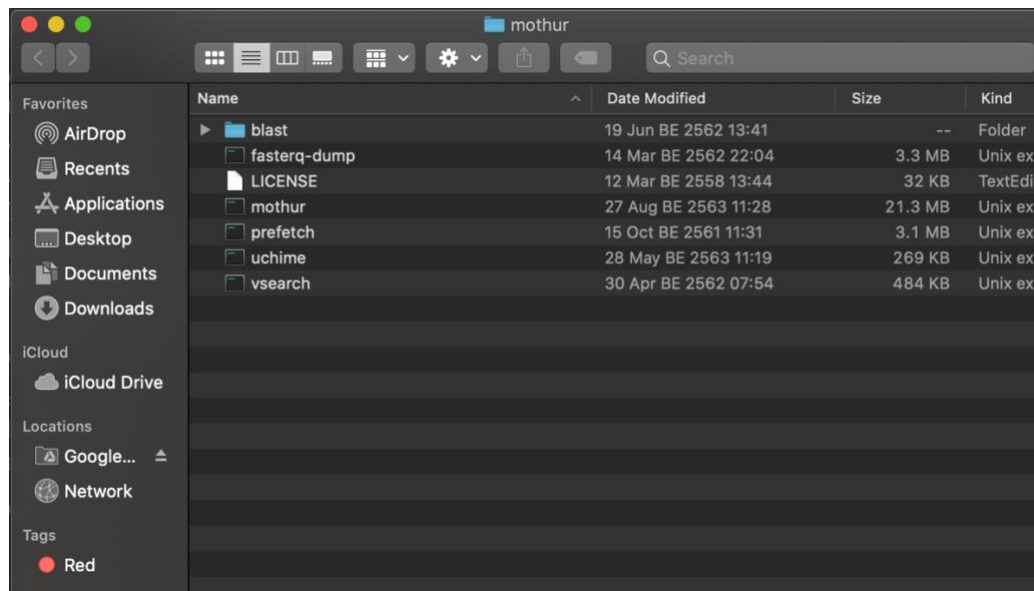

Figure 2. Decompressed mothur folder. The "mothur" program could be executed directly.

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.

2.2 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 3).

Figure 3. R installation files.

2. Download RStudio installation file and then install program based on your OS (Figure 4). The software available at https://rstudio.com/products/rstudio/download/.

Figure 4. 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 and vegan. For each package, please use a R command inside RStudio to install as:

```
install.packages("PACKAGE_NAME")
```

To activate the package, the command is:

```
library(PACKAGE_NAME)
```

2.3 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. 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 5). You can download entire repository using git clone or download compressed file(Figure 6). 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_intermidiate_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 5). You can download either entire repository or compressed file same as analysis outputs (Figure 6). Scripts can be found at "tutorial_scripts" folder.
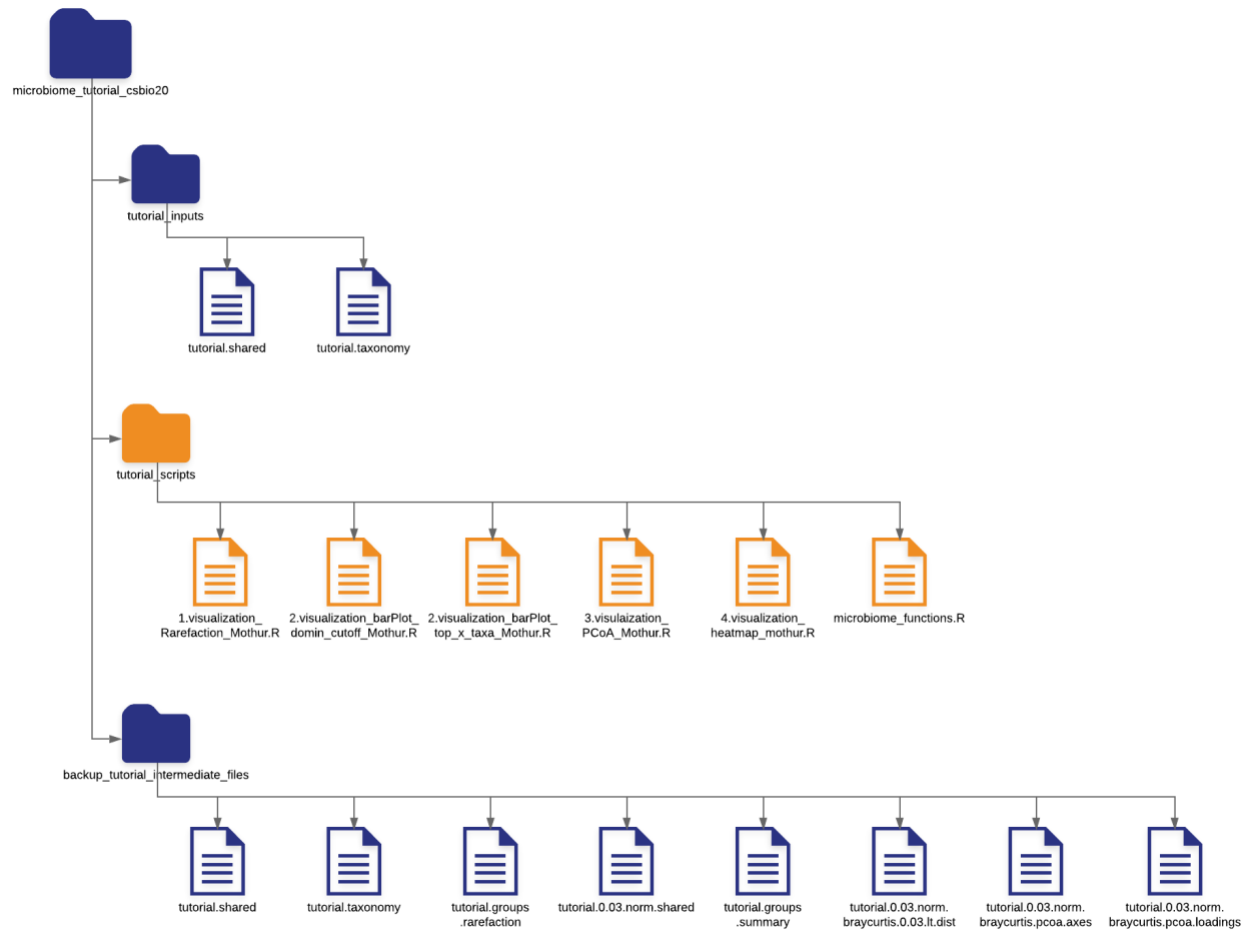
Figure 5. Tutorial files architecture. Blue icons indicate analysis files that will be produced/used during the microbiome analysis. Orange icons indicate scripts for the analysis.
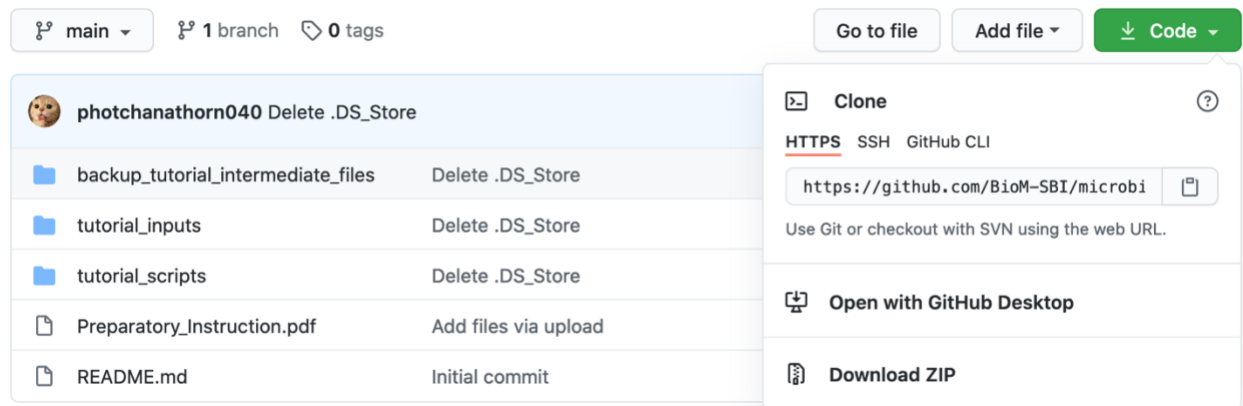

Figure 6. Github repository download location, you can either git clone or download ZIP file.

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., Prommeenate, 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.