

metaProbiotics manual

1. Overview

metaProbiotics is designed to identify the probiotics bins from metagenomic data. The tool takes one or more fasta files as input, where each fasta file contains sequences from the same bin, and identifies bins originating from probiotic bacteria.

metaProbiotics can be run on the **virtual machine**, via the **docker**, or via **Matlab interface**. Please refer to the following step-by-step guide for the usage of metaProbiotics. Any questions, please contact fangzc@smu.edu.cn.

2. Virtual machine version

2.1. Installing the virtual machine of metaProbiotics

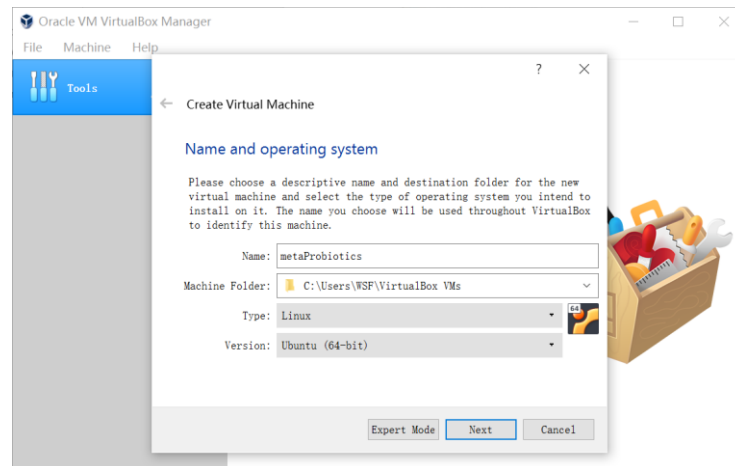
Step 1: Download the “metaProbiotics.vdi.7z” file from <https://metaprobiotics.obs.ap-southeast-1.myhuaweicloud.com/metaProbiotics.vdi.7z>. The “7z” file can easily be decompressed using current compressing software, such as “WinRAR”, “WinZip”, and “7-Zip”. After decompressing the 7z file, a file named “VM_Bioinfo.vdi” will occur.

Step 2: Download the VirtualBox software from <https://www.virtualbox.org> and install the VirtualBox. The VirtualBox is easy to install, you just need to select an installation folder and click the “next” button in each step.

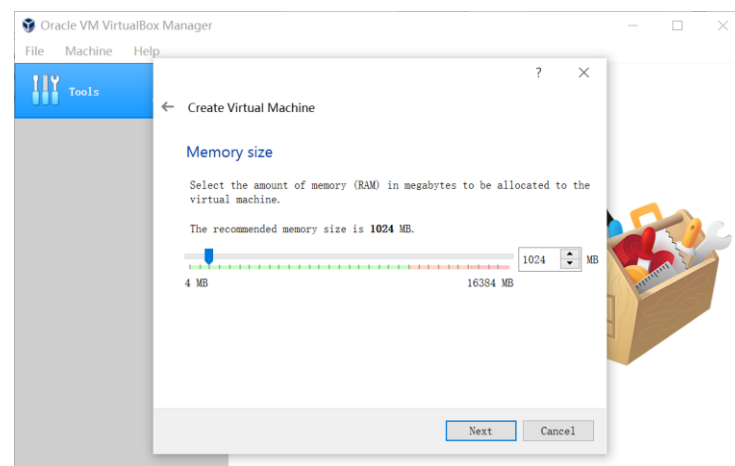
Step 3: Open VirtualBox, and click the “New” button to create a virtual machine.



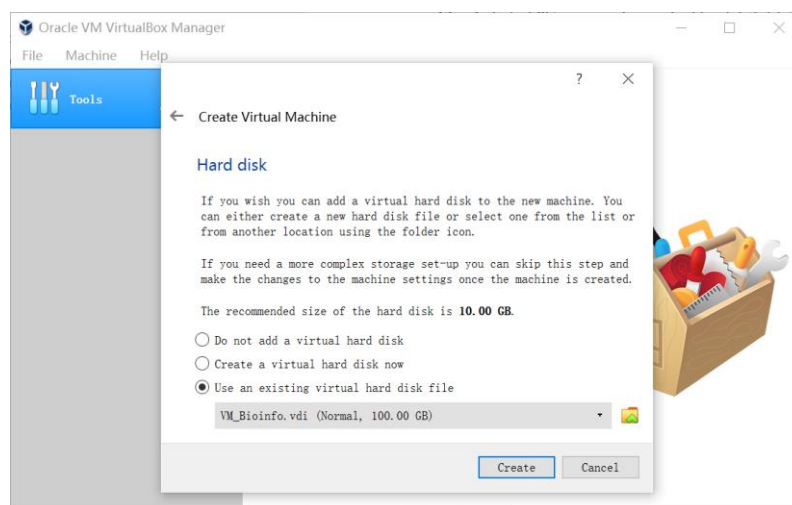
Step 4: Specify a name, select “Linux” as the operating system, and select “Ubuntu” as the version of the operating system. Then, click “Next”.



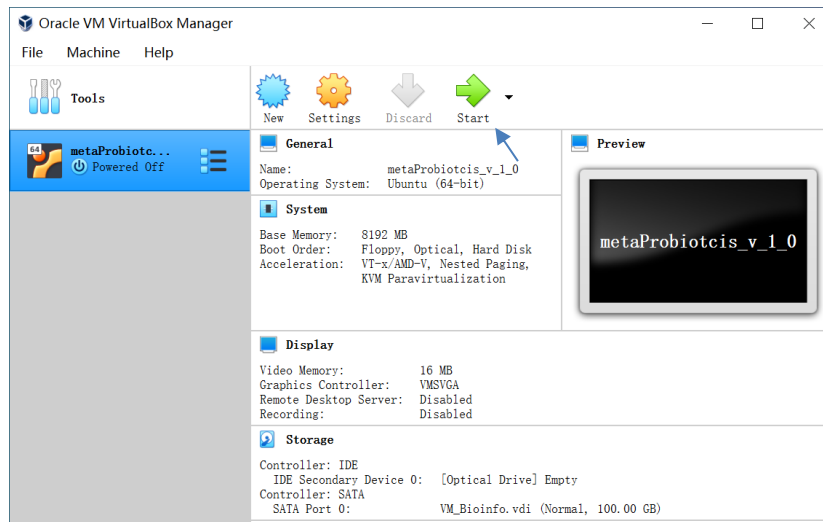
Step 5: If possible, allocate a larger amount of memory to the virtual machine (at least 6GB is recommend). Click “next”.



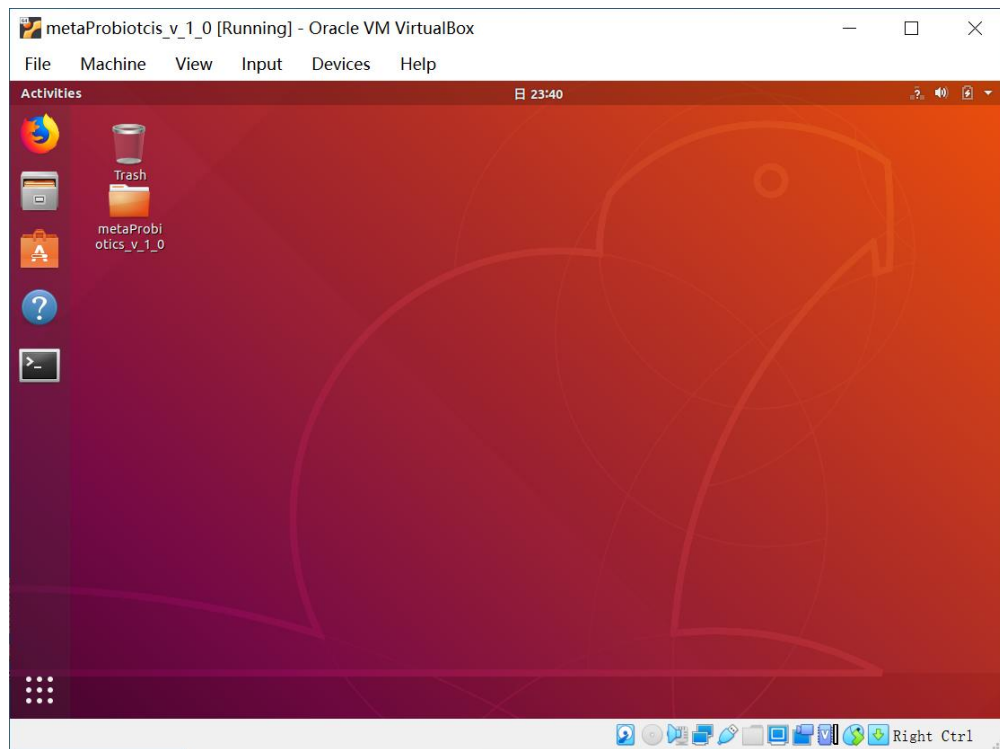
Step 6: Select “Use an existing virtual hard disk file”, and specify the “metaProbiotics.vdi” file.



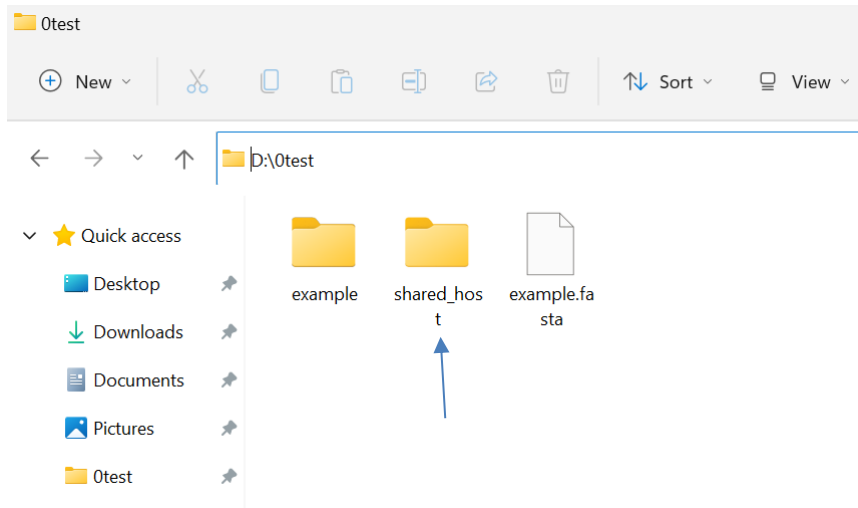
Step 7: Click “start” to open the machine.



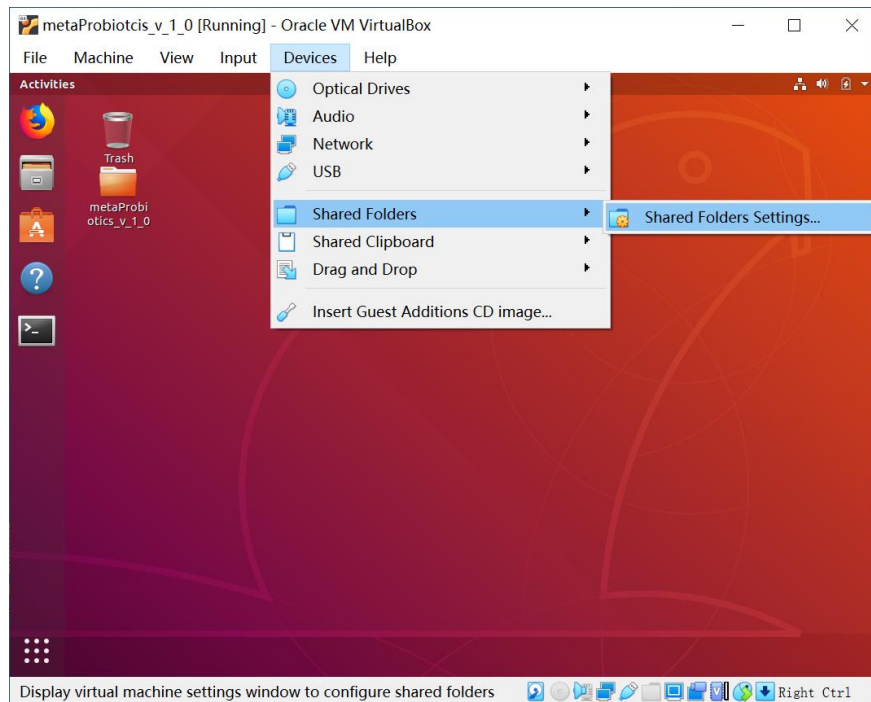
Step 8: The metaProbiotics folder is on the desktop.



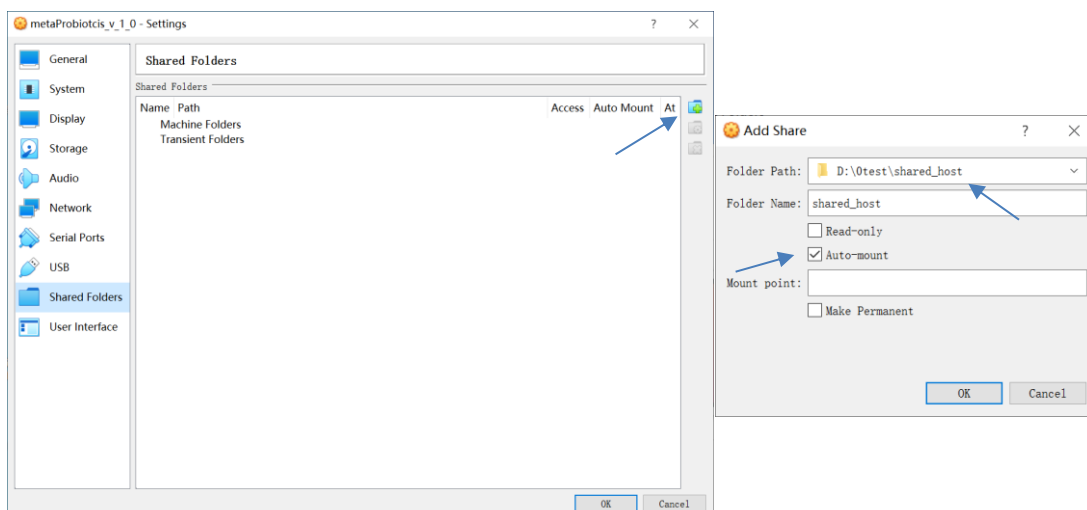
Step 9: Create a folder named “shared_host” in your physical host (not in the virtual machine). This folder is created for the file exchange between the physical host and the virtual machine.



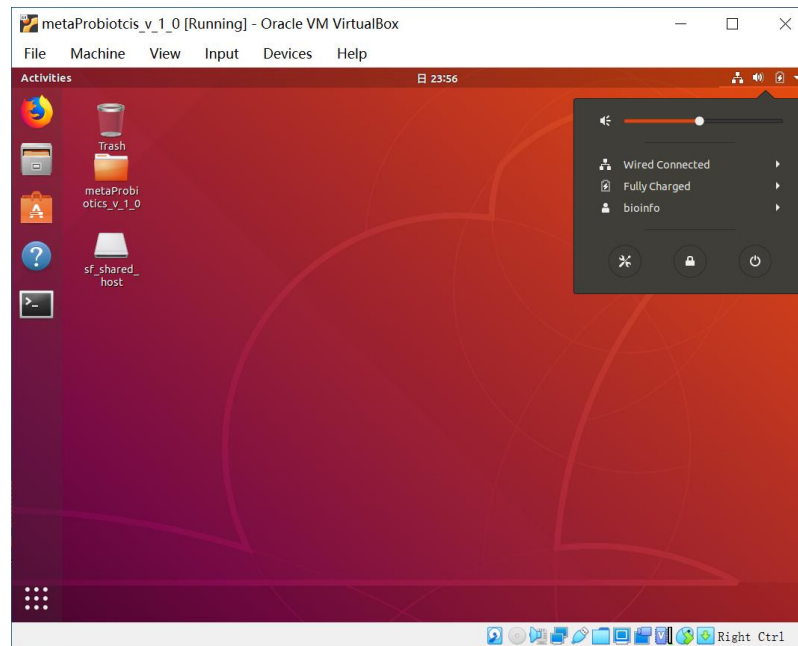
Step 10: In the window of VirtualBox, click “Devices”, “Shared Folder”, “Shared Folders Settings”



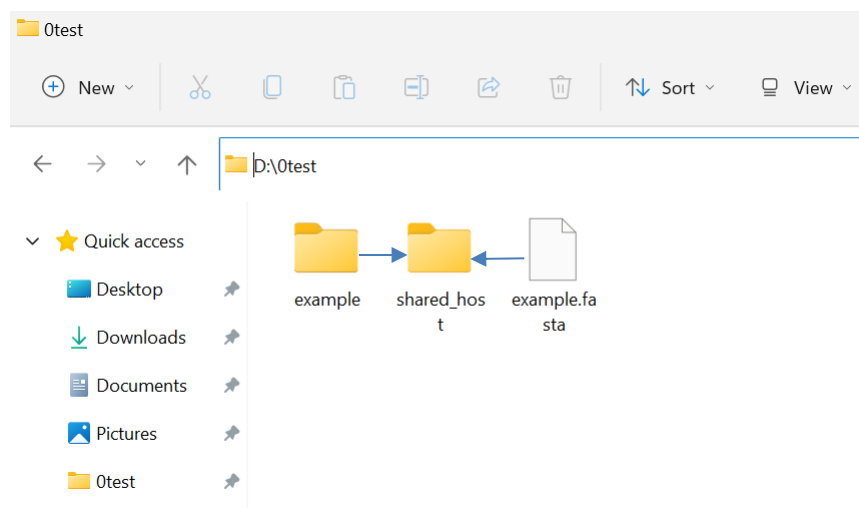
Step 11: Click ‘+’ to add the “shared_host” folder of the physical host, and then select “Auto-mount”.



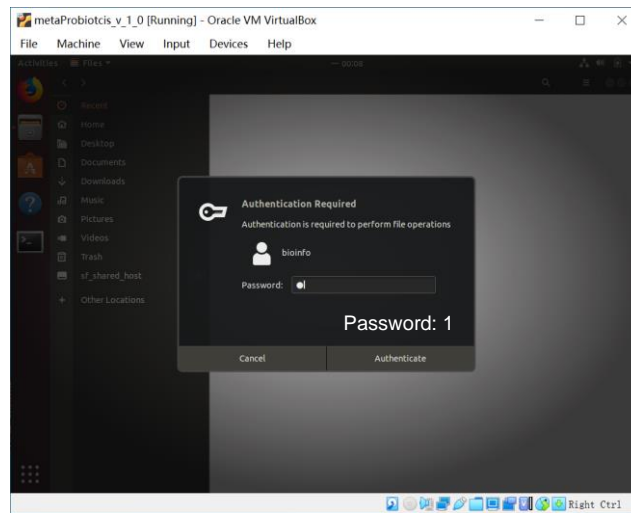
Step 12: Restart the virtual machine.



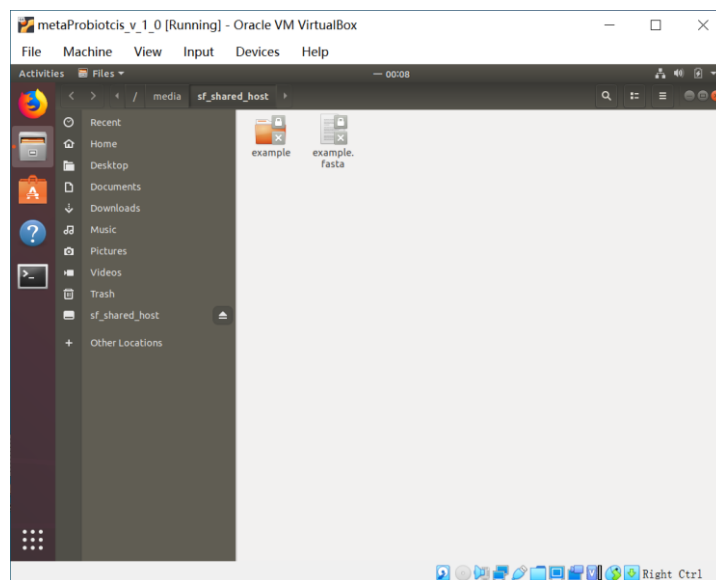
Step 13: In the physical host, copy your input file (fasta format) to the “shared_host” folder. Also, you can copy input folder containing multiple fasta file to the “shared host” folder.



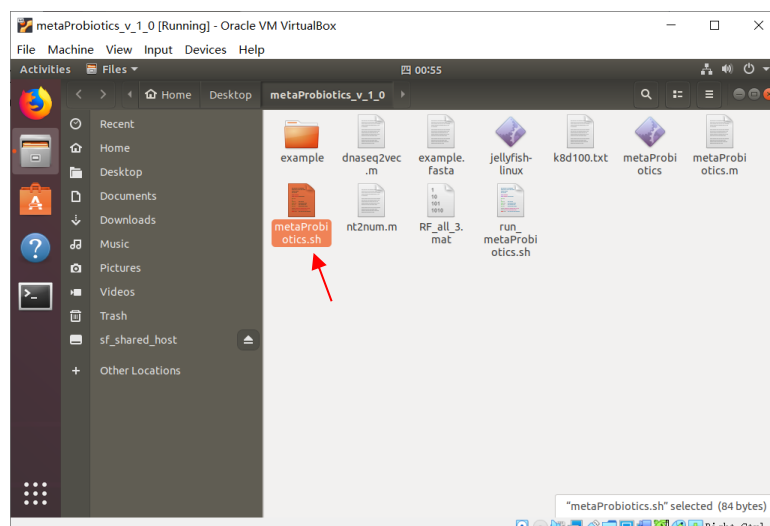
Step 14: Then the file will also occur in the “sf_shared_VM” folder on the desktop of the virtual machine. You can copy the file to other folders in the virtual machine. Before opening the “sf_shared_VM”, you need to enter "1" as the password and hit the "Authenticate" key.



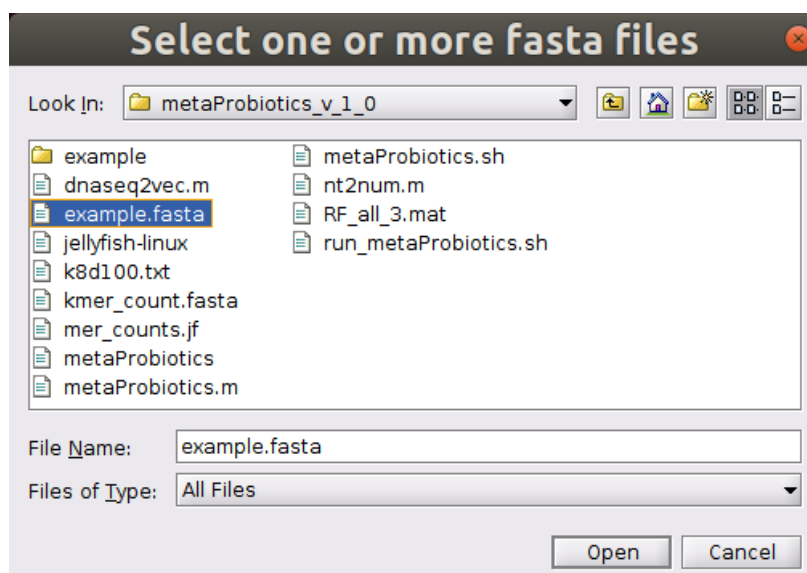
Step 15: Then you can get into the ‘sf_shared_VM” folder, and copy the input file or folder to other folders in the virtual machine.



Step 16: To run the metaProbiotics, go into the metaProbiotics_v_1_0 folder on the desktop, and double-click the “metaProbiotics.sh” file.



Step 17: Please wait for a few seconds. Then, select your input file and click the “Open” button, and the program will start running. You could select one fasta file or a list of fasta files as the input of metaProbiotics. Especially, if you want to select a series of files in the same folder, you could press the ‘shift’ first clicking the first and then the last file of this series, or press ‘ctrl’ and ‘A’ to select all the files in the current folder.



Step 18: When the terminal displays disappearance, the output file will occur under the same folder as the input file. The format of the output file is like “metaProbiotics-14-Sep-2023 08:00:02.txt”, which is named using the current time. You can copy the output file to the “sf_shared_VM” folder and the file will occur under the “shared_host” folder of the physical host.

3. Docker version

3.1. Installing the docker

Step 1: Users can download and install Docker on Windows, Mac, and Linux platforms from <https://docs.docker.com/get-docker/>. You can see the following options and click a corresponding platform for you. After switching to detailed installing instructions, you can install a docker. Especially, if you are a non-root user on a Linux platform, you should add yourself to a docker group so that you can have access to running docker (see <https://docs.docker.com/engine/install/linux-postinstall/>). To test whether Docker has been installed correctly, you could open your terminal and run:

docker run hello-world

If no error messages, the docker has been successfully installed.

Docker Desktop

Install Docker Desktop – the fastest way to containerize applications.



Step 2: After installing the docker successfully, you can start to use the docker version of metaProbiotics in a terminal. Open a new terminal on your computer. Then change the path to the location that you want.

Step 3: Pull down the image from Docker Hub by running the command:

`docker pull shufangwumetaprobiotics:1.0`

Step 4: Obtain all the local docker images that you have by running the command:

`docker images`

```
(base) sfwu@server:~$ docker images
REPOSITORY          TAG          IMAGE ID          CREATED          SIZE
shufangwu/metaprobiotics  1.0         a7e37b8c15ab     8 hours ago     5.23GB
```

Step 5: Start the docker container based on downloaded images by running the command:

`docker run -it -v /data/sfwu/metaProbiotics:/mnt shufangwu/metaprobiotics:1.0`

Then you can enter the docker container and begin to use metaprobiotics. Especially, “-v” parameter could help you to synchronize the folder in and out of the docker container. The first path is from the host machine, such as “/data/sfwu/metaProbiotics:/home/metaProbiotics”. The second path is the path in the docker, such as ‘/mnt’. Note, it is not recommended to set the docker path as “/home”, which could cover the inner ‘metaProbiotics’ folders in the docker.

```
(base) sfwu@server:~$ docker run -it -v /data/sfwu/metaProbiotics:/mnt shufangwu/metaprobiotics:1.0
root@204de5284f03:/#
```

3.2. Reenter the container

Step 6: If you want to exit from a current using container and stop this container, you can just type ‘exit’ in the command line and then click ‘Enter’ on the keyboard. While you just

want to exit and keep the container running, you can press Ctrl+P+Q. Then you can come back to your host.

Step 7: Obtain the information about all the available containers:

docker ps -a

```
(base) sfwu@server:~$ docker ps -a
```

CONTAINER ID	IMAGE	COMMAND	CREATED	STATUS	PORTS	NAMES
204de5284f03	shufangwu/metaprobiotics:1.0	"bash"	About a minute ago	Up About a minute		hardcore_pare2
6729275a0cdf	shufangwu/metaprobiotics:1.0	"bash"	2 weeks ago	Up 2 weeks		dreamy_franklin
99d869216741	shufangwu/metaprobiotics:1.0	"bash"	2 weeks ago	Exited (0) 2 weeks ago		pensive_shockley

Step 8: You need to pay attention to ‘CONTAINER ID’ and ‘STATUS’ columns. For example, the container (CONTAINER ID: 99d869216741)’s STATUS is ‘Exited’, which means this container is stopped when using the ‘exit’ command to exit. However, the container (CONTAINER ID: 204de5284f03)’s STATUS is ‘Up’, which means this container is running when using ‘Ctrl+P+Q’ to exit. The running containers could transfer files between themselves and the host while the stopped containers could not. You can also rerun a stopped container by:

docker start 204de5284f03

(‘204de5284f03’ is the CONTAINER ID of the stopped container)

3.3. Running metaProbiotics script

Step 9: Under the metaProbiotics folder, the user can directly type the following command to run the integrated pipeline for a signal fasta file:

./metaProbiotics example.fasta result

```
root@204de5284f03:/home/metaProbiotics# ./metaProbiotics example.fasta result
Finished.
```

The “result” is the output file’s name specified by the user.

To predict a batch fasta files using metaProbiotics, the user can type the following command:

./metaProbiotics_batch example results

```
root@204de5284f03:/home/metaProbiotics# ./metaProbiotics_batch example results
The file example1.fasta is finished.
The file example2.fasta is finished.
The file example3.fasta is finished.
Organizing the results...
Finished.
```

The “example” is a folder that contains all fasta files (other non-fasta files should not be included). The “results” is the output file’s name specified by the user, which contains the prediction results for each fasta file in the “example” folder.

Note: When running the program, some warning messages about the device may occur, and users can ignore the message.

4. Matlab version (Linux system is required)

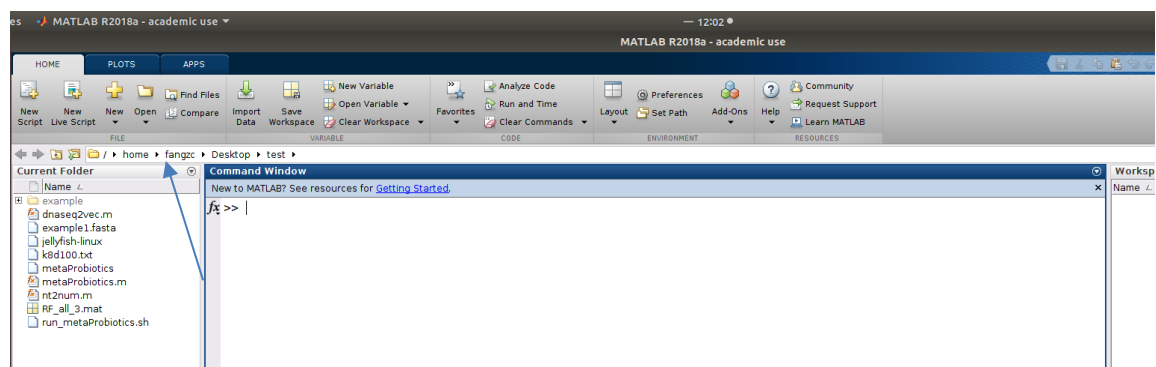
4.1. Installing the Matlab software

Step 1: Download the Matlab software and install it. You can download and install Matlab according to the instructions from the matlab website (<https://www.mathworks.com/products/matlab.html>).

4.2. Running the script in Matlab GUI

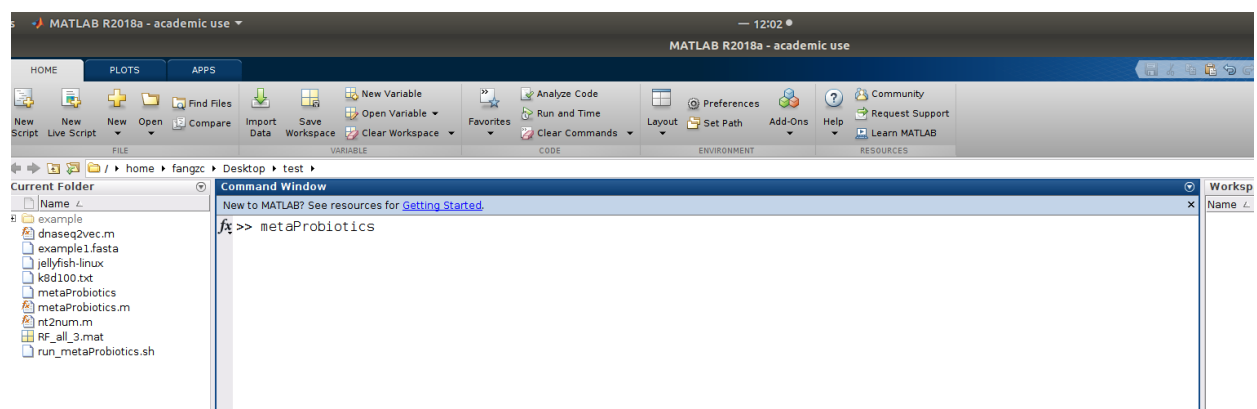
Step 2: Download the metaProbiotics package from <https://github.com/zhenchengfang/metaProbiotics>.

Step 3: Open Matlab, and change the working path to the metaProbiotic folder downloaded from step 2.



Step 4: In the Matlab interface, directly excute:

metaProbiotics



Step 5: The following steps are the same as step 17 and 18 of the virtual machine version.

5. Output

When running the metaProbiotics, the output file contains 3 columns:

The 1st column contains the name of the fasta file.

The 2nd column contains the probiotics scores (between 0 and 1) calculated by metaProbiotics.

The 3rd column contains the prediction results. By default, the bins with probiotics scores higher than 0.5 are regarded as probiotics, while the others are regarded as non-probiotics.