

metaProbiotics manual

1 Overview

metaProbiotics is designed to identify the probiotics bins from bacterial sequences.

metaProbiotics can identify whether the given bin of certain bacterium from a single fasta file is probiotics derived, and can simultaneously identify a batch of some bacteria from several fasta files in the same folder.

metaProbiotics can be run on the **virtual machine**, via the **docker**, or via **Matlab software**. metaProbiotics takes a “fasta” format file containing sequences as input and outputs a tabular file.

2 Virtual machine version

2.1 Installing the virtual machine of metaProbiotics

Step 1: Download the “metaProbiotics .vdi.7z” file from

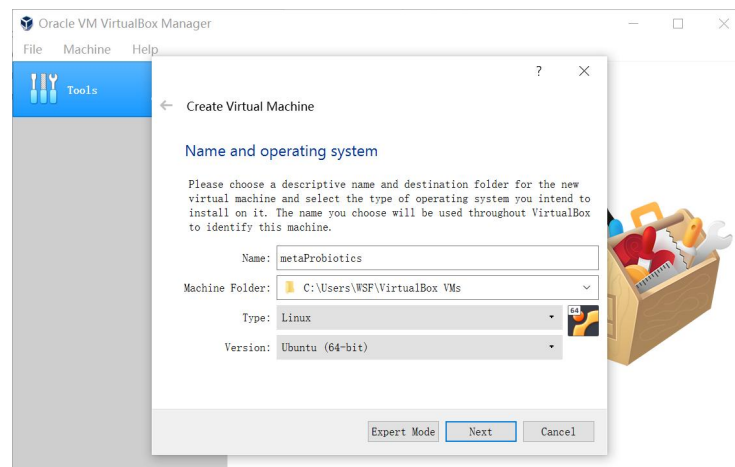
<https://zenodo.org/records/14898421/files/metaProbiotics.vdi.7z?download=1>. The “7z” file can easily be decompressed using current compressing software, such as “WinRAR”, “WinZip”, and “7-Zip”. After decompressing the 7z file, a file named “VM_Bioinfo.vdi” will occur.

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.

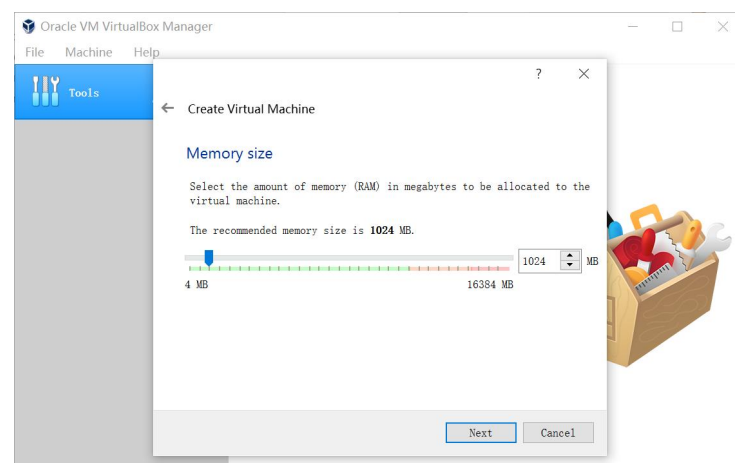
Step 3: Open VirtualBox, and click the “New” button to create a virtual machine.



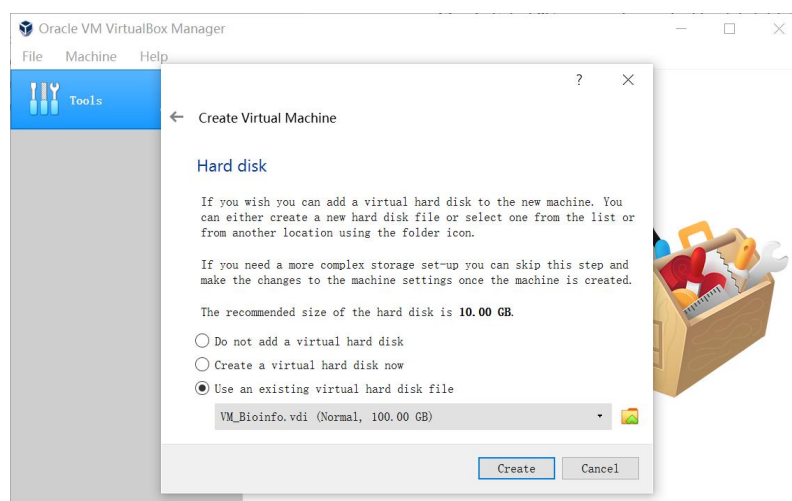
Step 4: Specify a name, select “Linux” as the operating system, and select “Ubuntu” as the version of the operating system. Then, click “Next”.



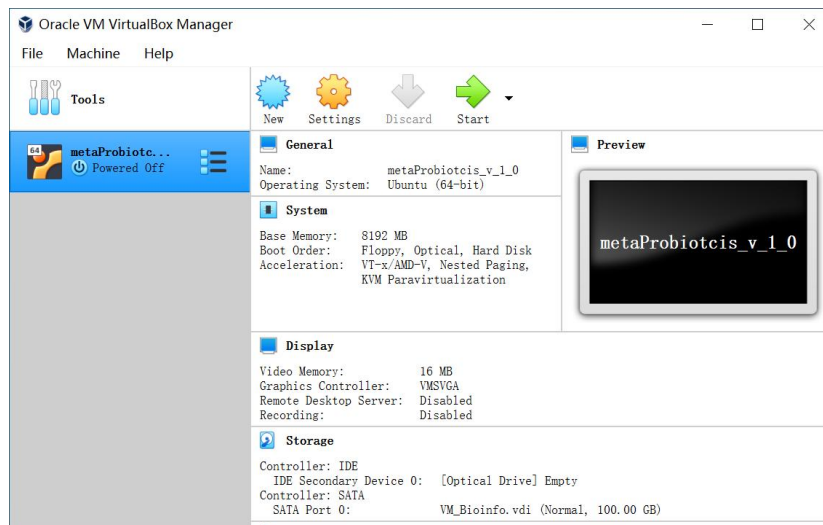
Step 5: If possible, allocate a larger amount of memory to the virtual machine. Click “next”.



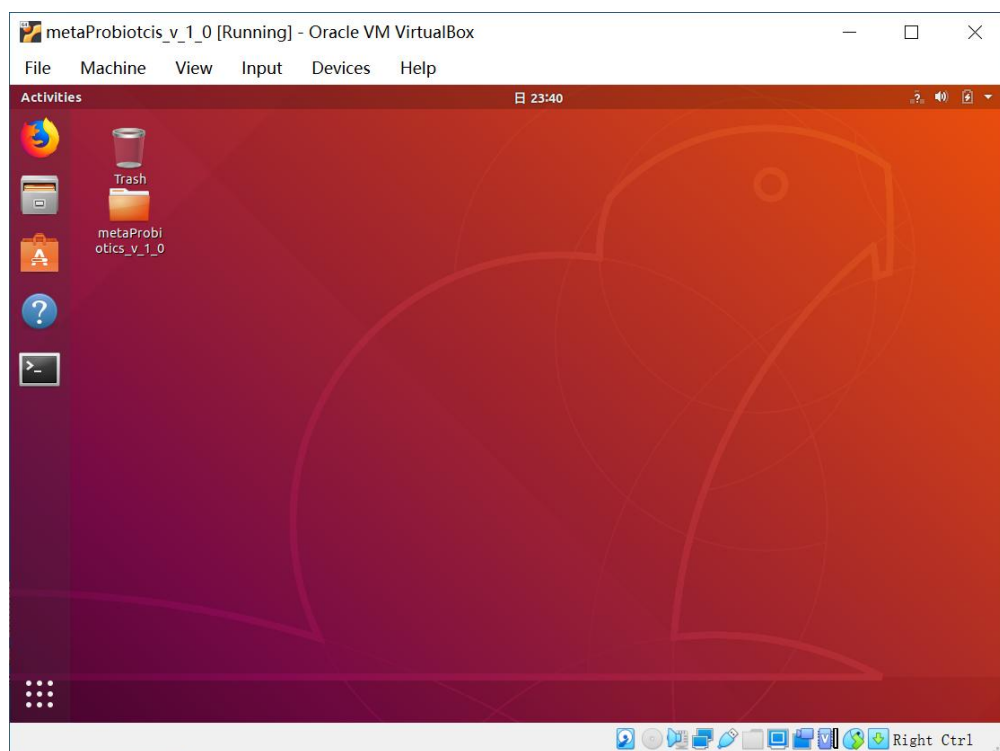
Step 6: Select “Use an existing virtual hard disk file”, and specify the “VM_Bioinfo.vdi” file.



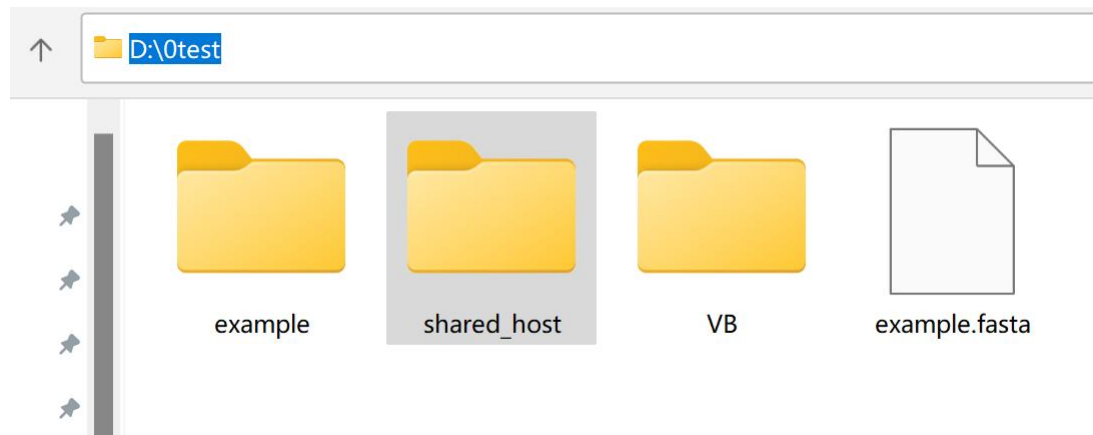
Step 7: Click “start” to open the machine.



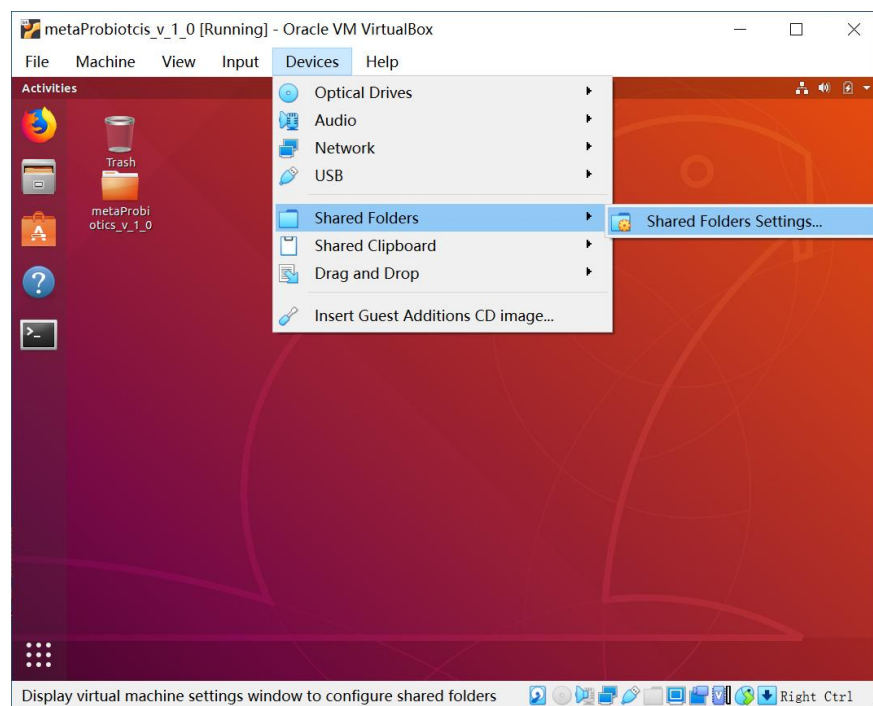
Step 8: The metaProbiotics folder is on the desktop.



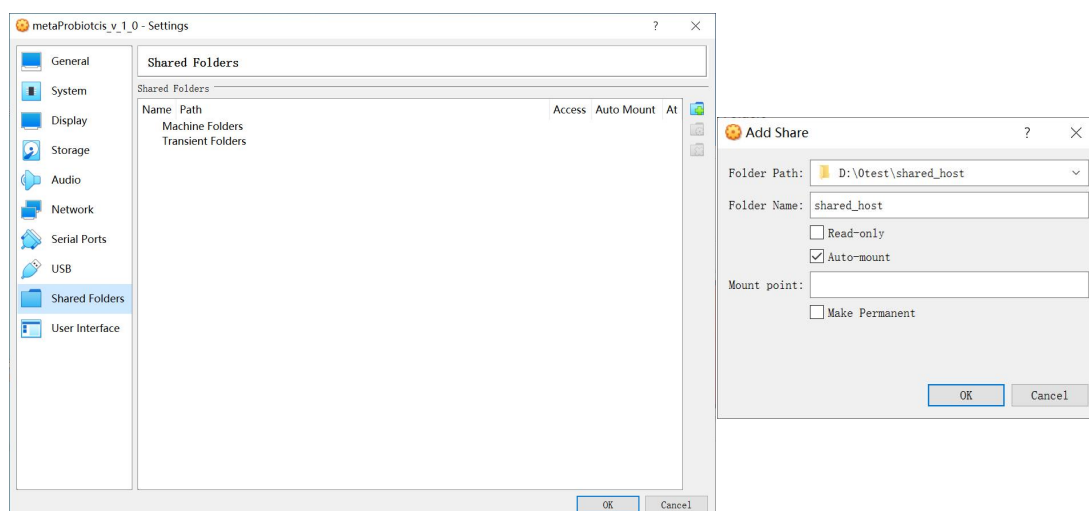
Step 9: Create a folder named “shared_host” in your physical host (not in the virtual machine, and do not change the folder name). This folder is created for the file exchange between the physical host and the virtual machine.



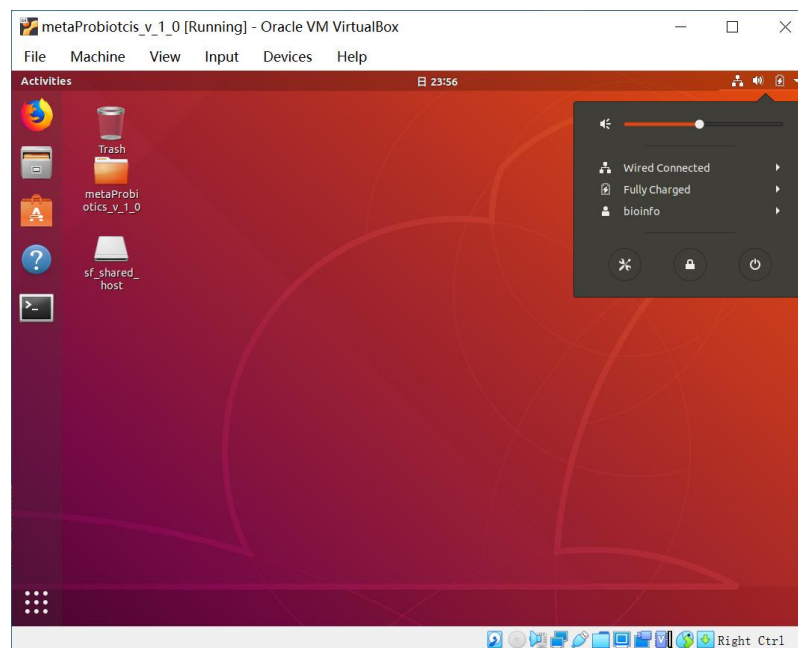
Step 10: In the window of VirtualBox, click “Devices”, “Shared Folder”, “Shared Folders Settings”



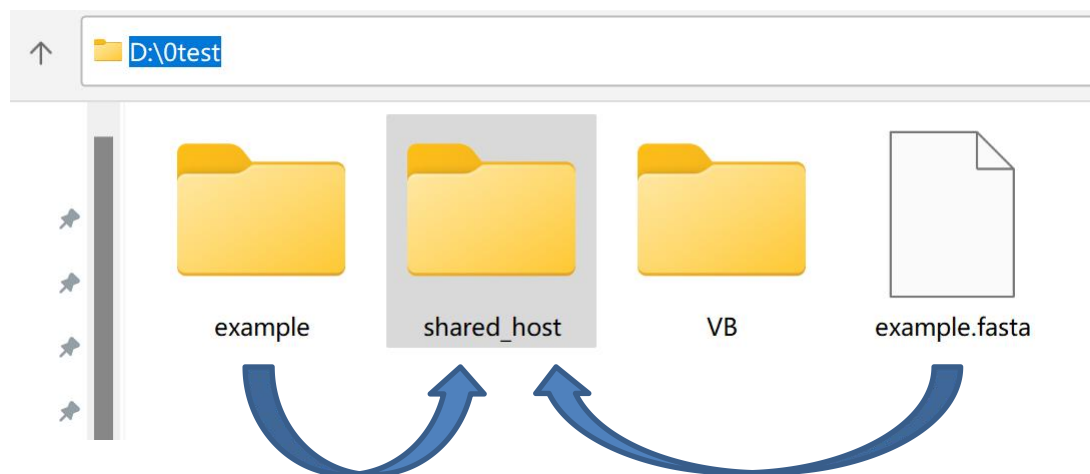
Step 11: Click ‘+’ to add the “shared_host” folder of the physical host, and then select “Auto-mount”.



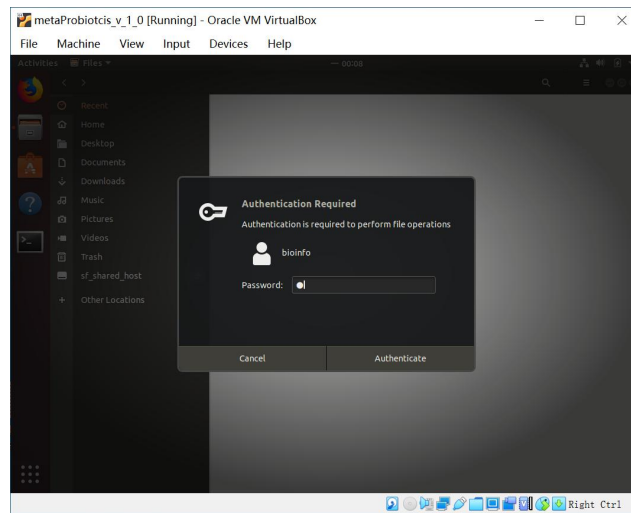
Step 12: Restart the virtual machine.



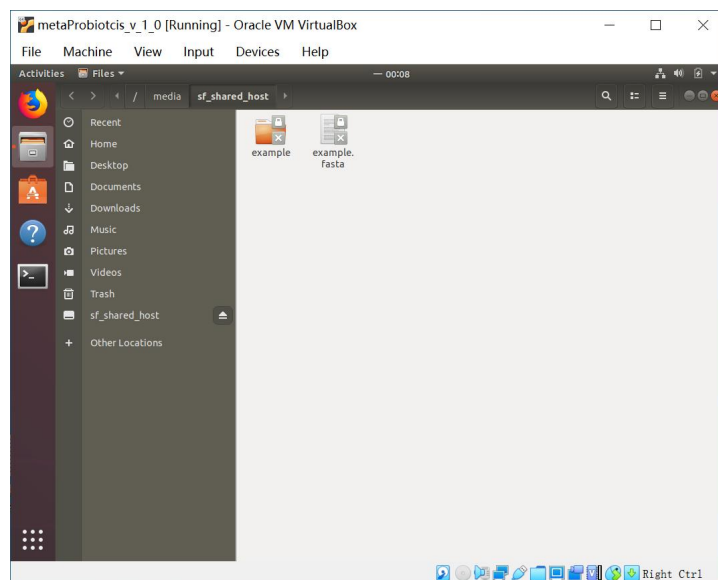
Step 13: In the physical host, copy your input file (fasta format) to the “shared_host” folder. Also, copy your input folder (fasta format file in one folder) to the “shared host” folder.



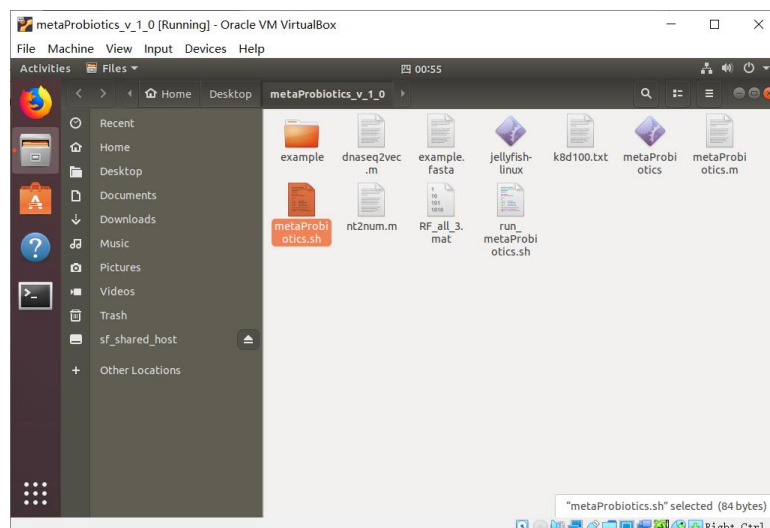
Step 14: Then the file will also occur in the “sf_shared_VM” folder on the desktop of the virtual machine. You can copy the file to other folders in the virtual machine. Before opening the “sf_shared_VM”, you need to enter "1" as the password and hit the "Authenticate" key.



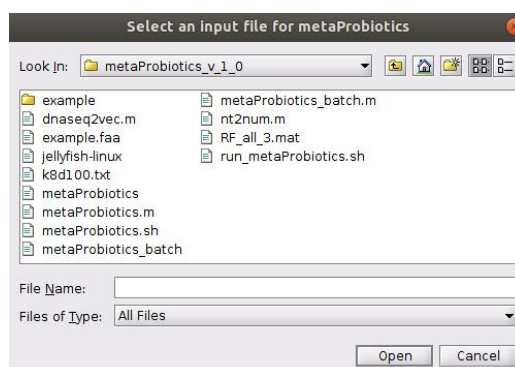
Step 15: Then you can get into the ‘sf_shared_VM” folder, select the file and folder, and “copy” them. Next, you can “paste” them to the folder “shared_VM”.



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 “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/>). Here is a video link from YouTube for installing Docker on a Linux system, which users can follow (see <https://www.youtube.com/watch?v=ynFlgkKzxOY>). 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.

3.2 Running a metaProbiotics docker

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

docker pull shufangwu/metaprobiotics: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

Step5: Prepare the fasta file that you want to predict and check its path:

ls /data/sfwu/metaProbiotics/

```
(base) sfwu@server:~/metaProbiotics$ ls /data/sfwu/metaProbiotics/  
example example1.fasta  
(base) sfwu@server:~/metaProbiotics$ █
```

Step 6: 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. “-v” parameter could help you to synchronize the folder in and out of the container of metaProbiotics. The path before “:” is the path out of docker (“/data/sfwu/metaProbiotics”), which includes a prepared fasta file and a folder of a batch of fasta files. The path after “:” is the synchronized path (“/mnt”) in the container. Then you can change the path into “/home/metaProbiotics” to use metaProbiotics.

```
(base) sfwu@server:~/metaProbiotics$ docker run -it -v /data/sfwu/metaProbiotics:/mnt shufangwu/metaprobiotics:1.0  
root@dc0d19454920:/# █
```


3.3 Running metaProbiotics script

Step 7: Predict a fasta file using metaProbiotics by running the command:

`./metaProbiotics /mnt/example1.fasta /mnt/result1`

```
root@dc0d19454920:/home/metaProbiotics# ./metaProbiotics /mnt/example1.fasta /mnt/result1
Finished.
root@dc0d19454920:/home/metaProbiotics#
```

“/mnt/example1.fasta” is the prepared fasta file synchronized into the container of metaProbiotics in step 6. The “/mnt/result1.txt” is the path and name of the output file. The predicted result could be found in the synchronized path out of the container, which is “/data/sfwu/metaProbiotics”.

Step 8(optional): Predict a batch fasta files using metaProbiotics by running the command:

`./metaProbiotics_batch /mnt/example /mnt/results`

```
root@dc0d19454920:/home/metaProbiotics# ./metaProbiotics_batch /mnt/example /mnt/results
The file example1.fasta is finished.
The file example2.fasta is finished.
The file example3.fasta is finished.
Organizing the results...
Finished.
root@dc0d19454920:/home/metaProbiotics#
```

“/mnt/example” is the prepared folder of a batch of fasta files synchronized into the container of metaProbiotics in step 6. The “/mnt/results” is the path of the output file. The predicted result could be found in the synchronized path out of the container, which is “/data/sfwu/metaProbiotics”.

Step 9: Exit and stop the current container of the container of metaProbiotics by typing “exit” in the command line and then clicking “Enter” on the keyboard.

Note: When running the program, some warning messages about the device may occur, and users can ignore the message. We have recorded a usage video executing the operation steps of sections 3.2-3.3 (see

https://metaProbiotics.obs.ap-southeast-1.myhuaweicloud.com/metaProbiotics_usage.mov).

4 Matlab version

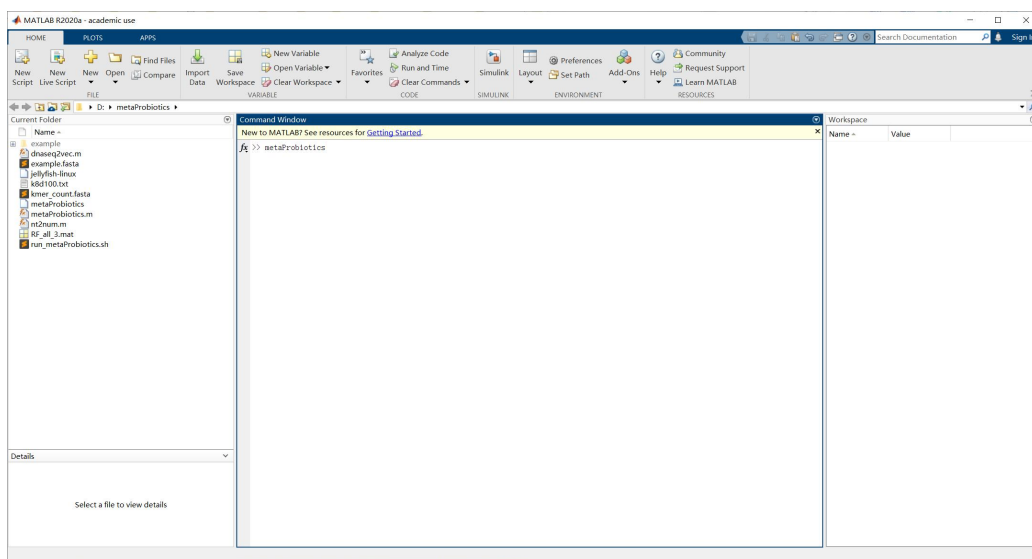
4.1 Installing the Matlab software

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>).

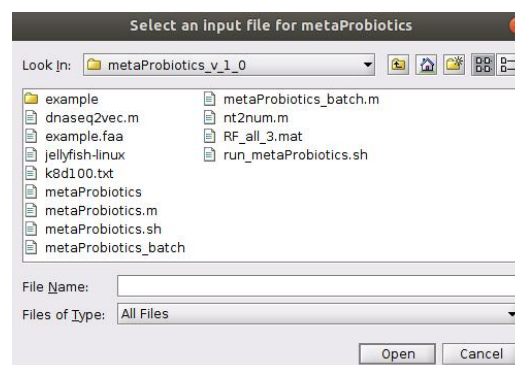
4.2 Running the script in Matlab GUI

Step 2: Use the following command directly in the MATLAB command window, please execute:

metaProbiotics



Step 3: 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.



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 the single-bin module of metaProbiotics.

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