#

DUESSELPORE Webserver manual

This is instruction of using Duesselpore webserver. Video of instruction can be found at:

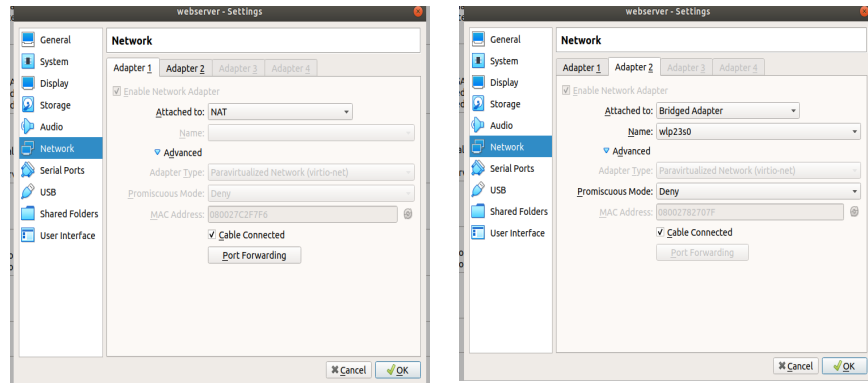# 1. Install and configure webserver

1.1. System requirement

```
* CPU: 2.5 GHz 8 cores or higher
* System memory: 8 GB or higher
* Diskdrive: 200 GB free space
* Host operating system Window 10, Linux (Ubuntu >=18.04 or Fedora) or MacOS
```

1.2. Installation

1.2.1 Download and install VMWare   Note: For inexperienced Linux user our software are tested with current version pipeline. We do not recommend upgrading the version on Linux Virtual machine. The webserver may crash when new software is updated

- Download and install Virtualbox (VB) installation and VirtualBox 6.1.22 Oracle VM VirtualBox Extension Pack from https://www.virtualbox.org/wiki/Downloads. Already tested Virtualbox version 6.1.22 on Ubuntu 18.04 and Window 10.
- Download the webserver.ova image file from this address https://iufduesseldorf-my.sharepoint.com/:u:/g/personal/thach_nguyen_iuf-duesseldorf_de/ET7zomuFVRBBheV-S3TZ6soBH7GiduAEWkp_XF0foxYI3A

After installing VB and its Extension Pack, open VB >File> Import Appliance to select webserver.ova downloaded file, then set up configuration based on your machine configuration. By default, our web server uses 4 cores CPU, 8 GB RAM. We recommend using 8 CPUs, 16 GB RAM, HDD is auto allocated. Therefore when your data increases, the image file size also increases. We recommend deploying a VB image in the partition with at least 200 GB (depends on the number of users and data size, TB volume is highly recommended). Configure the network interface on your host site (your primary OS): Before we start the Virtual machine in the Virtual box configuration panel, we configure two network interfaces as in the figure below. The first network interface to the internet (NAT) and the second interface to our host machine.

VM Network interface configuration

1.2.2. Login and configure webserver   After booting up our guest OS, log in to your Virtual Machine (VM) with this credential:

```
* user name: ag-rossi (preset)
* password 123456
```

Open the terminal, and we can get our web server IP address by this command on the guest terminal. When you want to use only the Human genome.

```
$setup_webserver light
$runserver
```

If you want to use RNASeq for other organisms:

```
$setup_webserver full
$runserver
```

The program will download all reference genomes, genome annotation, and other required packages. It also sets your IP address into the allowed IP list of the webserver then the IP address is printed out from the printout messages.

2.2. Using webserver

2.2.1. Access webserver   Now you can use your webserver within your Local Area Network (LAN) with a regular web browser (e.g., Firefox or Google Chrome port: 8000) http://{Your IP address}:8000/duesselpore.

2.2.2. Data preparation   User can upload fastq files as ONE compressed zip file: each subfolder contains several replicas with one experimental condition. NOTE: files and folders' names must contain only alphabetic and numeric characters. Below is an example of data separated into two conditions, 'condition1' and 'condition2'.

```
fastq/(folder)
   condition1 (subfolder)
      condition1_replica1.fastq (single fastq file)
      condition1_replica2.fastq (single fastq file)
   condition2 (subfolder)
      condition2_replica1.fastq (single fastq file)
      condition2_replica2.fastq (single fastq file)
      condition2_replica3.fastq (single fastq file)
```

How to merge multiple fastq files into a single file: On Linux terminal:

```
$ cat /path/to/fastq/files/*.fastq > /your/new/location/output.fastq
```

On Window command prompt (path syntax is different):

```
$ type \path\to\fastq\files\*.fastq> \your\new\location\output.fastq
```

2.2.3. Setup running parameter: First, select one group among your groups as the reference group. Select the gene (transcriptome) counting method, then select the differential expression algorithm you want to analyze. To run the analysis, we have to set up other parameters of the analysis function. There are some optional parameters, e.g., ReadCountMinThreshold, Logfold change threshold, adjPValueThreshold. Submit and wait for the result. Advanced users can customize the RNA.R code to develop a new workflow.

2.2.4. Get the results: After the computation is completed, all the results are downloaded from the browser. We export the interactive Html file for some plots. Offline analysis can be continued on the Linux virtual machine directory. Experienced users can continue the further analysis by editing the R script. NGS data is high volume, therefore we recommend erasing the data on the virtual machine regularly.