# DUESSELPORE Webserver manual

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

## 1. Install and configure webserver

### 1.1. System requirement

\* CPU: 2.0 GHz 2 cores or higher  
\* System memory: 8 GB or higher  
\* Diskdrive: 100 GB or higher free space  
\* Host operating system Window 10, Linux (Ubuntu >=18.04 or Fedora) or MacOS. 64 bits

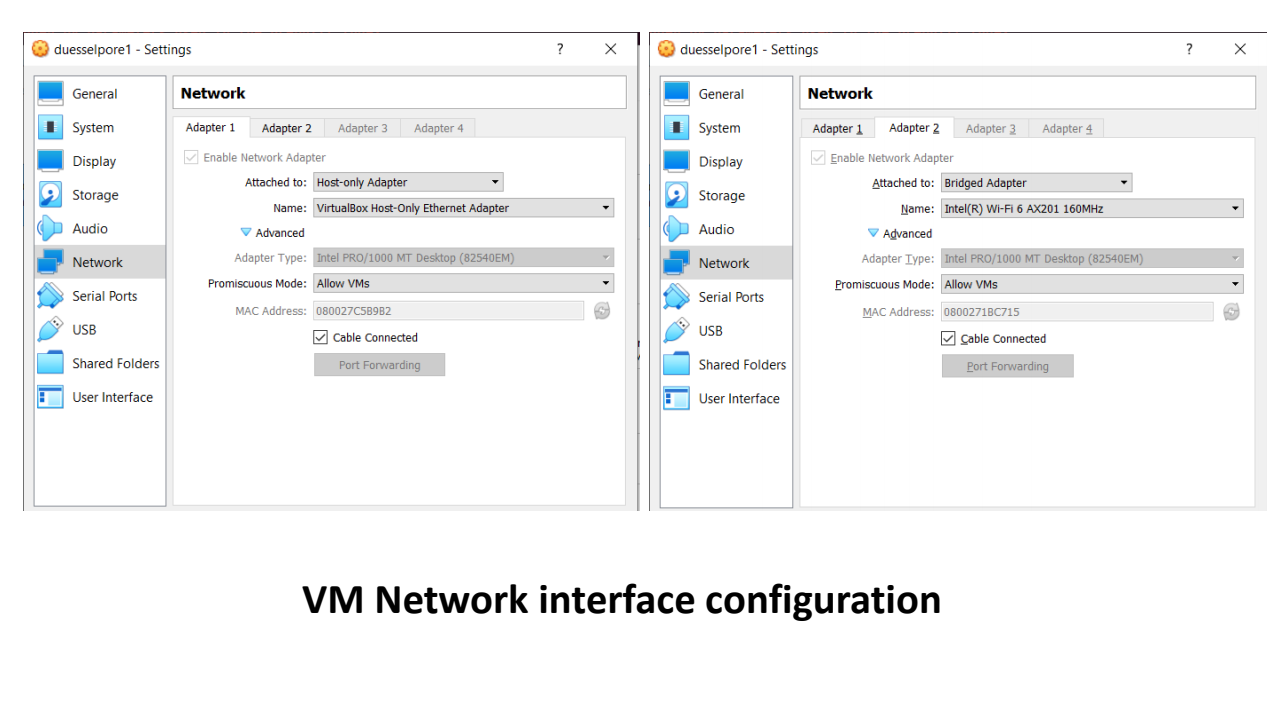
### 1.2. Installation

#### 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 VirtualBox GUI and open 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, or more. A 30 GB partition for swap, which extends your virtual memory. This configuration keeps the Minimap2 program running in a low memory machine. However, hard disk read/write speed is much slower than RAM. So to speed up the program you should use higher memory. Hard disk data is dynamically 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 host machine via VirtualBox Host-Only Ethernet adapter and the second interface the internet via one of your host machine network interface. Network configuration is critical important for our webserver.



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 default 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. The light configuration is for only the Human genome. 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. The configuration step required internet connection, therefore you should configure webserver before field work.

$setup\_webserver light  
$runserver

If you want to use RNASeq for other organisms, use this command (beta version):

$setup\_webserver full  
$runserver

### 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. The webserver can access via three ways: first way is your local network interface (normally start with 192.168.x.x), the second way your LAN IP address (depend on your LAN network), third way is on the Virtual machine (address: localhost)

#### 2.2.2. Data preparation

Users 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’. Please check the structure and the director name of your data carefully, all the name of analysis are generated by directory and file names.

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:

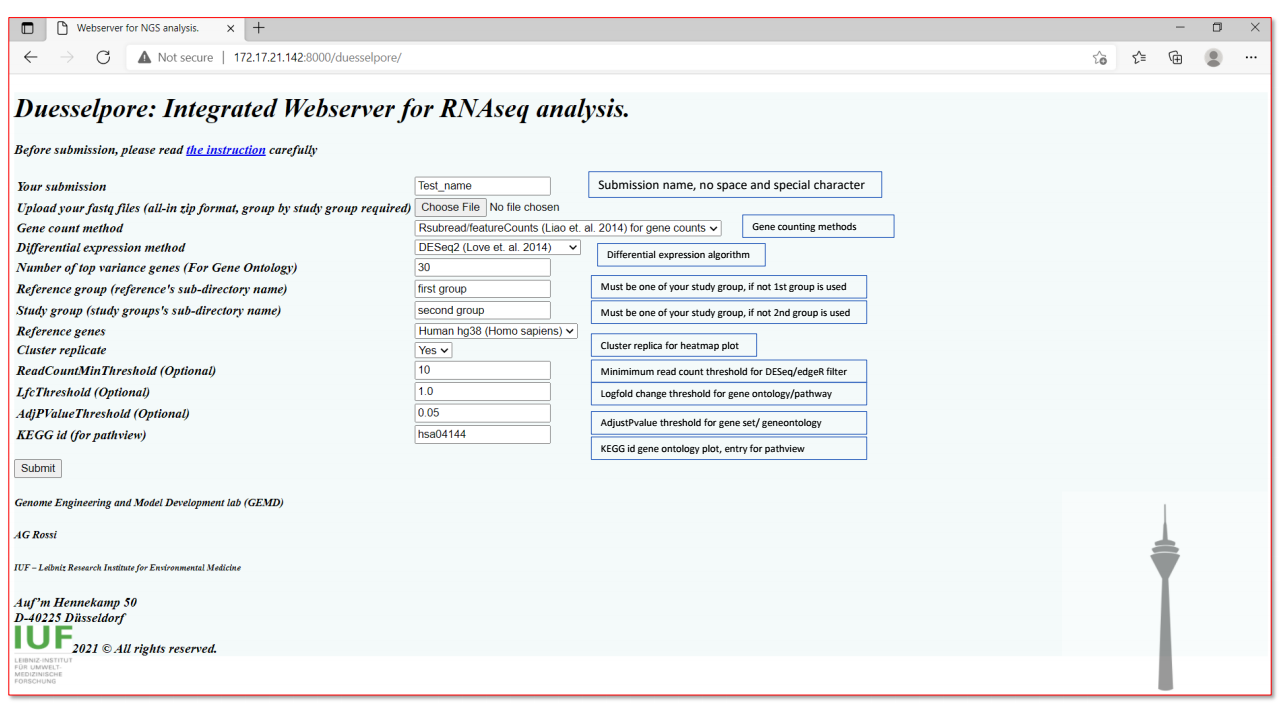
$ cat /path/to/fastq/files/\*.fastq > /your/new/location/output.fastq

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. After submit we can wait for the result. Advanced users can customize the RNA.R code to develop a new workflow. The figure bellow exlain the web input form.



Input web form explaination

#### 2.2.4. Collecting the results:

The run time depends on the your data size and the system speed. For our dataset which contains 6 replicates, approximate totals 16 million reads, the run time is 6.5 hours. After the computation is completed, all the results are downloaded from the browser. We export the interactive HTML file for some plots. Users can continue offline analysis on the Linux virtual machine directory at /home/ag-rossi/duesselpore/users\_file/{your session id}. 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. The Sample result is in the Support Information, or sample\_result.pdf.