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

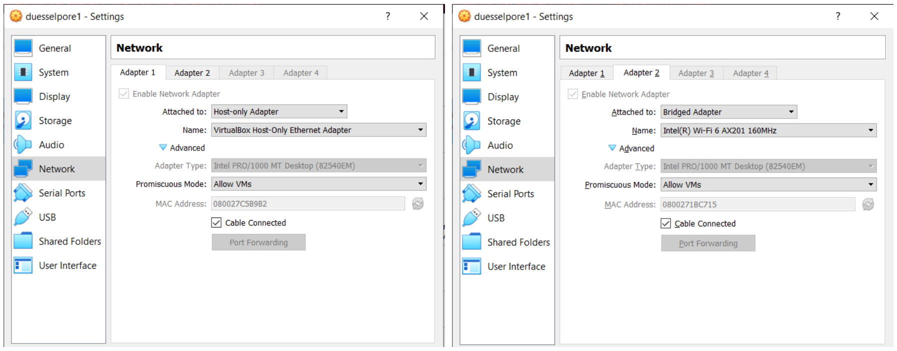
```
* CPU: 2.5 GHz 8 cores or higher
* System memory: 8 GB or higher
* Diskdrive: 200 GB free space
* Host operating system Window 10, Linux (Ubuntu >=18.04 or Fedora) or MacOS
```

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 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

Adapter 1 | Adapter 2 | Adapter 3 | Adapter 4

☑ Enable Network Adapter
Attached to: Host-only Adapter
Name: VirtualBox Host-Only Ethernet Adapter
Advanced
Adapter Type: Intel PRO/1000 MT Desktop (82540EM)
Promiscuous Mode: Allow VMs
MAC Address: 080027C5B9B2
☑ Cable Connected
Port Forwarding

Network

Adapter 1 | Adapter 2 | Adapter 3 | Adapter 4

☑ Enable Network Adapter
Attached to: Bridged Adapter
Name: Intel(R) Wi-Fi 6 AX201 160MHz
Advanced
Adapter Type: Intel PRO/1000 MT Desktop (82540EM)
Promiscuous Mode: Allow VMs
MAC Address: 0800271BC715
☑ Cable Connected
Port Forwarding

VM Network interface configuration

Figure 1: 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:

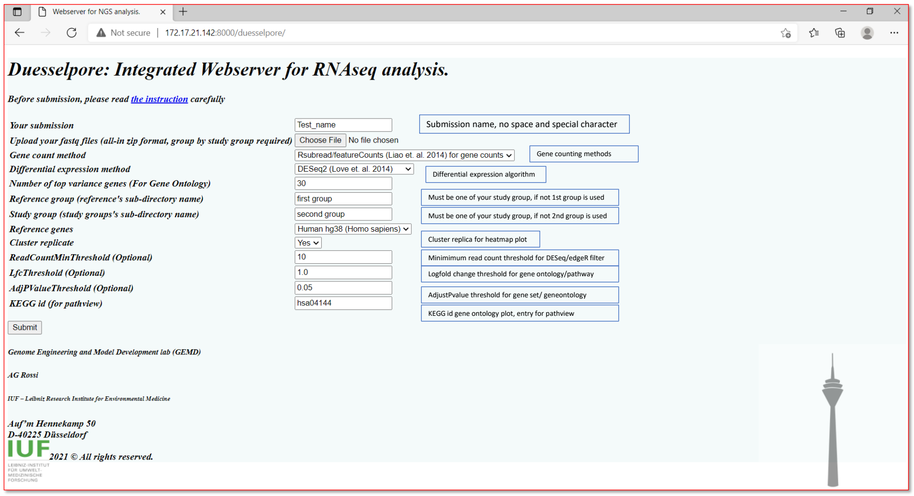
```
$ cat /path/to/fastq/files/*.fastq > /your/new/location/output.fastq
```

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.

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



Duesselpore: Integrated Webserver for RNAseq analysis.

Before submission, please read the instruction carefully

Your submission	Test_name	Submission name, no space and special character
Upload your fastq files (all-in zip format, group by study group required)	Choose File No file chosen	
Gene count method	Rsubread/featureCounts (Liao et. al. 2014) for gene counts	Gene counting methods
Differential expression method	DESeq2 (Love et. al. 2014)	Differential expression algorithm
Number of top variance genes (For Gene Ontology)	30	
Reference group (reference's sub-directory name)	first group	Must be one of your study group, if not 1st group is used
Study group (study groups's sub-directory name)	second group	Must be one of your study group, if not 2nd group is used
Reference genes	Human hg38 (Homo sapiens)	
Cluster replicate	Yes	Cluster replica for heatmap plot
ReadCountMinThreshold (Optional)	10	Minimimum read count threshold for DESeq/edgeR filter
LfcThreshold (Optional)	1.0	Logfold change threshold for gene ontology/pathway
AdjPValueThreshold (Optional)	0.05	AdjustPvalue threshold for gene set/ geneontology
KEGG id (for pathview)	hsa04144	KEGG id gene ontology plot, entry for pathview

Genome Engineering and Model Development lab (GEMD)

AG Rossi

IUF – Leibniz Research Institute for Environmental Medicine

Auf'm Hennekamp 50
D-40225 Düsseldorf
IUF *2021 © All rights reserved.*

Figure 2: Input web form explaination

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.