## June 2018

# An Introduction to Deep-Sequencing Data Analysis

## Exercise #1 Introduction to Linux and WEXAC cluster

***Dena Leshkowitz and Ester Feldmesser***

### Introduction

In this workshop we will learn how to use the cluster WEXAC and write commands in Linux through a terminal.

Supplementary information is found at the end of the document.

Commands you are requested to type appear in *italic font*.

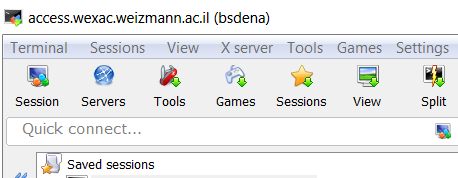
### Instructions

### Accessing the WEXAC server by opening a terminal

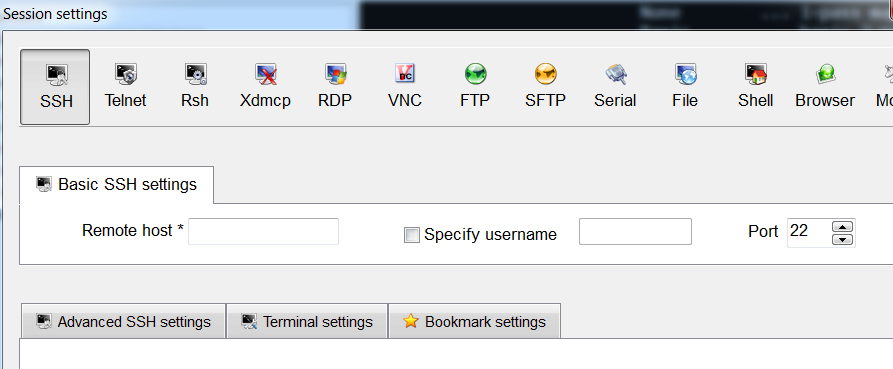
MobaXterm is an enhanced terminal for Windows with X11 server, tabbed SSH client, network tools and much more. MobaXterm is useful for working on linux servers. It is available for free download at: <http://mobaxterm.mobatek.net/download.html>

No need to do this now since it is already installed in the class.

Find the MobaXterm icon on your desktop and click to open.



Click on session (red arrow), then select SSH, the following window will open:



You are prompted to insert the host (server) name – *access.wexac.weizmann.ac.il* and the username – *please ask us for your username* and password. Then in the terminal you will be requested to enter the password.

### Learning Basic Unix/Linux commands

In order to list all the files and folders in your home directory (your current folder location), type:

*ls*

To get more information on this command type:

*man ls*

The *man* command lists all available parameters on the command of interest. To exit the manual, type: *q*.

For more commands, look at the supplementary section #2 at the end of the document.

To get more information on the files type:

*ls -l*

To move into the course\_2018 directory type:

*cd course\_2018*

### Editing a text file

Return to your home directory:

*cd ~*

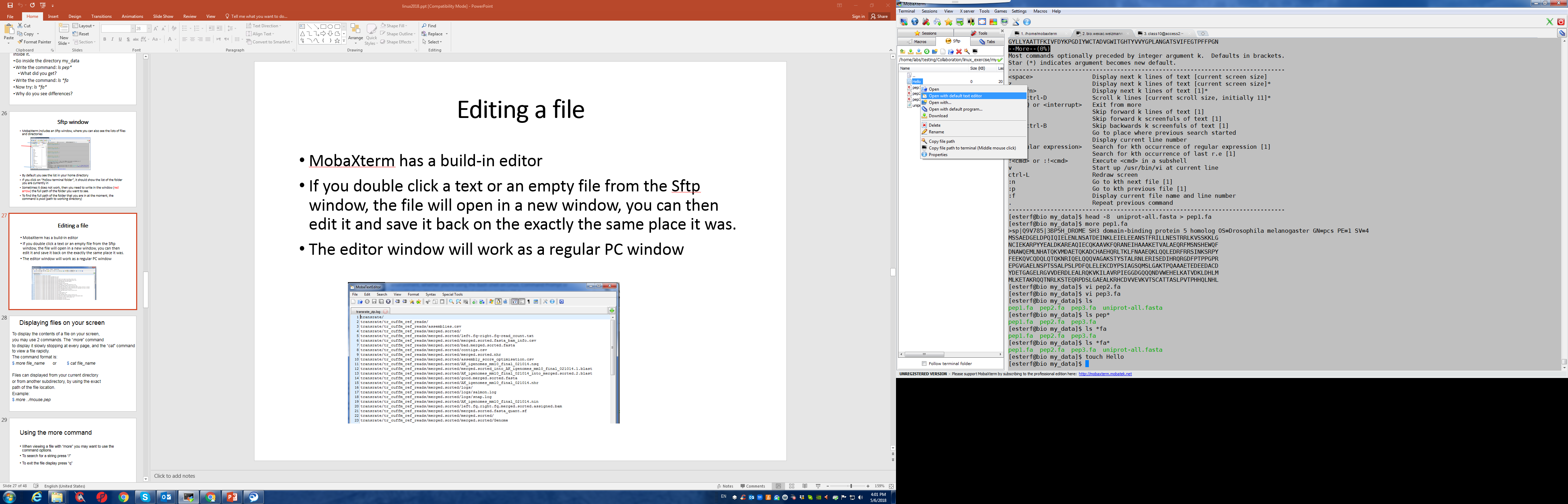
Create an empty file by typing:

*touch my\_file*

MobaXterm has a build-in editor. From the Sftp window (appears as a tab on the left pane of MobaXterm window), you can right click a text or an empty file name and open it in a new window with the default text editor (figure at the right). The editor window will work as a regular PC window.

Open the empty file that you created by double clicking on the Sftp window, write on it “This is my new file” and save it back in the same place (directory).

In the shell window (right pane), use the *more* command to view what you have written in the file.



### Loading modules to your environment

With the help of the Environment Modules package it is possible to dynamically  
modify your environment on the wexac cluster in order to be able to run particular software packages. We will load the modules that open the IGV browser.

(to learn more about IGV browser, look here: <http://software.broadinstitute.org/software/igv/home>). We will be using this tool a lot during the course.

Please type all the commands:

*module load IGV*

*module load jre*

*igv.sh*

This will take a few minutes. Please wait until the application’s window opens. Just close it for now.

### Running a job on the WEXAC cluster

The Wexac cluster is built from many host servers, the access server that we will be using, is just one of them. In order to send a job to the cluster so that it will run on one of the hosts available, use the bsub command (supplementary section #3). Following is an example for a command that will delay for 5 seconds and then will type the date in a file named date.txt:

*cd ~*

*bsub -q new-short -J mydat -o mydat\_log.txt -e mydate\_error.txt 'sleep 5;date>date.txt'*

You should get lines similar to:

Memory reservation is (MB): 2048

Memory Limit is (MB): 2048

Job <363927> is submitted to queue <new-short>

To see a job’s status type (see options in supplementary section #4):

*bjobs*

Continue to type bjobs every few minutes until you get the message “No unfinished job found”

Then to see the file that was created, type

*more date.txt*

List the files that you created by using *ls –l*. Use the *rm* (remove) command to clean your home directory from all four files you have just created with the following command (change filename with the relevant names).

*rm filename*

## Supplementary

### WEXAC introduction presentation

### <http://www.weizmann.ac.il/DIS/sites/DIS/files/uploads/it/files/wexac_introduction.pdf>

### Basic Linux commands:

man (command) ...... shows help on a specific command

ls ................. show directory, in alphabetical order

logout ............. logs off system

mkdir .............. make a directory

rmdir .............. remove directory (rm -r to delete folders with files)

rm ................. remove files

cd ................. change current directory

more .............. views a file, pausing every screenful

grep ............... search for a string in a file

head ............... show the first few lines of a file

tail ............... show the last few lines of a file

df ................. shows disk space available on the system

du ................. shows how much disk space is being used up by folders

chmod .............. changes permissions on a file

cut ............... print selected parts of lines

cp ............... copy file

mv ............... move file

wc –l .............. print the number of lines

sort ............. sort lines of text files

### bsub - Commonly Used Options

**-**q qname submits the job to the specified queue

-o file redirect stdout, stderr and resource usage information of the job to the specified output file

-e file redirect stderr to the specified error file

-oo/-eo file same as -o/-e, but overwrite file if it exists

-i filename, use the specified file as standard input for the job

-n number, specify number of job slots

-g jobgroup submit job to specified group

-J jobname assigns the specified name to the job

-R res\_req runs job on a host that meets the specified resource requirements

### bjobs - View Job Information

-a Display information about jobs in all states, including recently finished jobs

-A Displays summarized information about job arrays

-d Display information about jobs that finished recently

-l|-w Display information in long or wide format

-p Display information about pending jobs

-r Display information about running jobs

-g job\_group Display information about jobs in specified group

-J job\_name Display information about specified job or array

-q queue Display information about jobs in specified queue

-u user Display information about jobs for specified users/groups