Linux

Saba Ghaffari
Using ClustalW to align two sequences
Step 1A: Accessing the IGB Biocluster

Open **Putty.exe**

In the **hostname** textbox type:

biologin.igb.illinois.edu

Click **Open**

If popup appears, Click **Yes**

Enter login credentials assigned to you; example, user **class00**.
You will not see any characters on screen when typing in password. Just type it.

Now you are all set!
Step 1B: Listing files and directories (`ls`)

```
$ ls
# listing files in your current directory. When you first login, your directory is your home directory.
```
Step 1C: Making Directories (mkdir)

$ mkdir ~/01_Linux

# create a subdirectory in your home directory. The tilde ~ character refers to your home directory.

$ ls

# to see the directory you just created.
Step 1D: Changing directory (cd)

The lab is located in the following directory:

/home/classroom/mayo/2020/01_Linux

$ cd /home/classroom/mayo/2020/01_Linux
# tip: use “tab” for auto-completion for path
$ ls
# to see the contents. You should see seqs.fa

Step 1E: Print working directory (pwd)

$ pwd
# to see the full pathname. You should see
"/home/classroom/mayo/2020/01_Linux"
Step 1F: Copying files (cp)

Copy `seqs.fa` from the data directory to your working directory.

```bash
$ cp /home/classroom/mayo/2020/01_Linux/seqs.fa ~/01_Linux/
# tip: use “tab” for autocompletion for path
$ cd ~/01_Linux/
```

Step 1G: Displaying the contents of a file on the screen (more)

```bash
$ more seqs.fa
# you should see two sequences on your screen
>seq1
GATCGAGCGATCGTGCAGC
GCAGAATGCGCGCTAG
>seq2
```
## Commands Summary

<table>
<thead>
<tr>
<th>Command</th>
<th>Meaning</th>
</tr>
</thead>
<tbody>
<tr>
<td>ls</td>
<td>list files and directories</td>
</tr>
<tr>
<td>mkdir directory</td>
<td>make a directory</td>
</tr>
<tr>
<td>cd directory</td>
<td>change to named directory</td>
</tr>
<tr>
<td>cd ~</td>
<td>change to home directory</td>
</tr>
<tr>
<td>cd ..</td>
<td>change to parent directory</td>
</tr>
<tr>
<td>pwd</td>
<td>display the path of the current directory</td>
</tr>
<tr>
<td>cp file1 file2</td>
<td>cp file1 and call it file2</td>
</tr>
<tr>
<td>more file</td>
<td>display the contents of a file</td>
</tr>
</tbody>
</table>
## Useful tips

<table>
<thead>
<tr>
<th>Command</th>
<th>Meaning</th>
</tr>
</thead>
<tbody>
<tr>
<td>tab</td>
<td>auto-complete path</td>
</tr>
<tr>
<td>↑</td>
<td>retrieve previous commands</td>
</tr>
</tbody>
</table>
Step 1H: Run sequence alignment program

Accessing the IGB Biocluster
Step 1H: Run sequence alignment program

```bash
$ srun -p classroom -c 2 --mem 8000 --pty bash  # SKIP IF DONE

# Open interactive session on biocluster with 2 cpus and 8G memory.

$ module load ClustalW2  # Load sequence aligner into the shell environment.

$ module list  #See loaded tools

$ clustalw2 -INFILE=seqs.fa  # Run the clustalW sequence aligner.
```
Step 1H: Run sequence alignment program

You will see this on your screen, when the program is done.

---

CLUSTAL 2.1 Multiple Sequence Alignments

Sequence format is Pearson
Sequence 1: seq1          35 bp
Sequence 2: seq2          32 bp
Start of Pairwise alignments
Aligning...

Sequences (1:2) Aligned. Score: 21
Guide tree file created: [seqs.dnd]

There are 1 groups
Start of Multiple Alignment

Aligning...
Group 1: Delayed
Alignment Score 47

CLUSTAL-Alignment file created [seqs.aln]
Step 1H: Run sequence alignment program

The alignment result is in seqs.aln. Use more command to see the result.

```
$ more seqs.aln
# You should see the following on your screen.

CLUSTAL 2.1 multiple sequence alignment

seq1  GATCGAGCGA-TCGTGCAGCGAGACAATGCAGCTAG
seq2  GGTAGGGTTATGCTACCGTCGATCGAGTA----

* * * *  * * *   * **  **  * * 
```
Exit putty by either closing the window or typing ‘exit’ in the command prompt.