

Bash for Bioinformatics

(MBIO 4030 T10)



Abdullah Zubaer; Hausner Lab, University of Manitoba, zubaera@myumanitoba.ca

Making files and directories

- Creating a folder

```
mkdir folder_name
```

- Creating a file

```
touch file_name
```

```
nano file_name
```

```
cat > file_name
```

Note: cat concatenates text into a file.

- Remove a folder

```
rmdir folder_name
```

Accessing files

- less (less is more!)
`less file_name`
press “q” to quit
- head
`head file_name`
- tail
`tail file_name`
- cat
`cat file_name`
press Ctrl+D to quit
- Using text editor such as **vi**, **nano**, **emacs** etc.
`nano file_name`
press Ctrl+x to quit

Finger exercise 3

- Make a folder
- Copy a file from other directory
- Create another file
- View the content of those files
- Remove the folder

Finger exercise 4

- Download a nucleotide sequence file from NCBI
- Rename it
- Make a copy of the file to your home folder
- Access the file with `less`, `head`, `tail`, `cat` and `nano`

Finger exercise 5

- Make a file and write something on it using nano
- Try cat commands with '`>`' and '`>>`' to add some text to the file
- Check the file with `less`

Note: when you name a file or folder, never use space.

Some powerful notations

- Redirect (or rewrite) with “>”
- Redirect (and append) with “>>”
- Wildcard “*”
- Pipeline “|”

Smart cat

- Combine the contents of a number of files into another file

```
cat file_1 file_2 file_n > combined_file
```


Finger exercise 6

- Download five (5) nucleotide (in FASTA format) sequence files from NCBI
- Combine all the sequences into one file

Compare files

- diff

```
diff file_1 file_2
```

- comm

```
comm file_1 file_2
```

```
comm -1 file_1 file_2
```

```
comm -2 file_1 file_2
```

```
comm -3 file_1 file_2
```

Finger exercise 7

- Download Yeast chromosome IV and V in GTF format from UCSC database.
- Copy first 10 lines from each file to new files (name them file_1 and file_2).
- Combine those files to a new file (name it file_3).
- Compare the those two files to the combined files.

Grab it with 'grep'

“You can't grep dead trees”

- **Global Regular Expression Print** (grep)
- Invoking grep

```
grep pattern file_name
```

```
grep -option pattern file_name
```

Note: option = c for count, v for invert-match

Finger exercise 8

- Create a directory
- Create 5 different files
- Write the output of `ls` into different files
- Combine the contents of all the files in the directory
- Remove the directory

Finger exercise 9

- Download “influenza.fna.gz” file from the ftp site of NCBI Influenza database
- Unzip it with “gunzip”
gunzip file-name
- How many sequences you have in total in that file?
- Make a list of the name (header-line) of those sequences.
- How many “neuraminidase” gene sequences you have?