Unix/Linux Tutorial for Beginners
Session VII

Mihaela Martis

NBIS & Faculty of Medicine and Health Sciences
Division Cell Biology, IKE

NATIONAL BIOINFORMATICS INFRASTRUCTURE SWEDEN

LINKÖPING UNIVERSITY

history

- enables the repeating of commands entered earlier in the session
- the GNU History library keeps track of all lines typed in the terminal

```
$ history
1997  find . -name '*.tex' | sort -n | wc -l
1998  exit
1999  cd myTeaching/linux_introduction/
2000  kile slides/linux_session6.tex&
2001  cd ..
...
2017  man history
2018  history
```

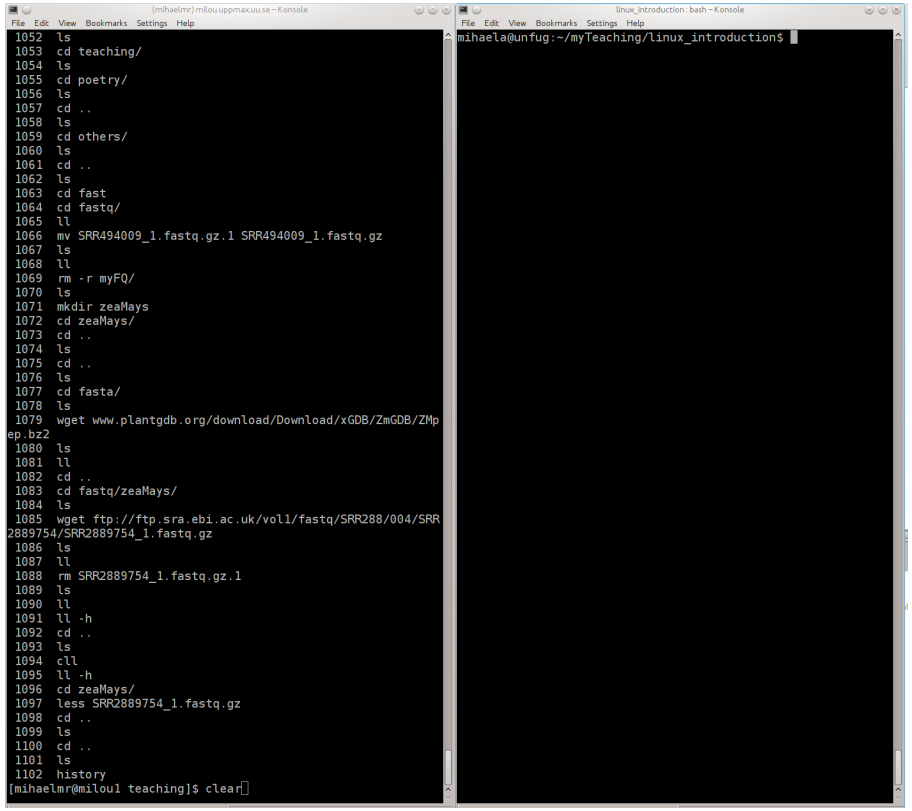
- use  and  to navigate through the history

clear

- clears the screen and shows only the prompt

```
$ clear
```

- is equivalent to Ctrl + L



Checking disk space

- du – shows how much disk space is taken by your files

```
$ cd /proj/g2015039/nobackup/nov2015/tuesday
$ du -hs .
4,2M .
```

→ displays combined size of all files in the current directory and recursively in all its subdirectories

Checking disk space

- du – shows how much disk space is taken by your files

```
$ cd /proj/g2015039/nobackup/nov2015/tuesday
$ du -hs .
4,2M .
```

→ displays combined size of all files in the current directory and recursively in all its subdirectories

→ displays combined size + the size of each subdirectory

```
$ du -h --max-depth=1 .
32K     ./TEST
224K    ./molecules
448K    ./plain_text
3,5M    ./sequences
4,2M    .
```

Checking disk space

- df – shows how much disk space is available

```
$ df -h
Filesystem            Size  Used Avail Use% Mounted on
/dev/mapper/vg-root    32G   14G   17G  44% /
tmpfs                  63G  4,0K   63G   1% /dev/shm
/dev/sda1             2,0G  277M  1,6G  15% /boot
/dev/mapper/vg-scratch 234G  6,0G  216G   3% /scratch
gulo@tcp0:/glob        1,1P  598T  452T  57% /gulo
pica1-v3:/pica/v3      228T  169T   60T  74% /pica/v3
pica1-v2:/pica/v2      228T  165T   64T  73% /pica/v2
```

column

- formats the input into multiple columns that are much easier to read
- usage: column <options> <file>

```
$ grep -v "^#" mySample_somaticMutations.vcf|cut -f 1-8|head
chr1    10043   .       T       G       .       REJECT  .
chr1    10055   .       T       G       .       REJECT  .
chr1    10067   .       T       G       .       REJECT  .
chr1    10079   .       T       G       .       REJECT  .
chr1    10157   .       T       C       .       REJECT  .
chr1    10180   rs201694901     T       C       .       REJECT  DB
chr1    10250   rs199706086     A       C       .       REJECT  DB
chr1    726859  rs139100483     A       G       .       REJECT  DB
chr1    726887  .       A       G       .       REJECT  .
chr1    726895  rs141325488     A       G       .       REJECT  DB
$ grep -v "^#" mySample_somaticMutations.vcf|cut -f 1-8|column -t|head
chr1    10043       .             T  G  .  REJECT  .
chr1    10055       .             T  G  .  REJECT  .
chr1    10067       .             T  G  .  REJECT  .
chr1    10079       .             T  G  .  REJECT  .
chr1    10157       .             T  C  .  REJECT  .
chr1    10180       rs201694901   T  C  .  REJECT  DB
chr1    10250       rs199706086   A  C  .  REJECT  DB
chr1    726859      rs139100483   A  G  .  REJECT  DB
chr1    726887      .             A  G  .  REJECT  .
chr1    726895      rs141325488   A  G  .  REJECT  DB
```

Join

- join – used to join different files together by a common column
- works only if both files are sorted by the column to be joined on

```
$ cat example.bed
chr1  26   39
chr1  53   84
chr3  32   99
chr1  9    28
chr2  10   19
$ cat example_length.txt
chr1   58352
chr2   39521
chr3   24859

$ sort -k1,1 example.bed   >   example_sorted.bed
$ sort -k1,1 example_length.txt   >  example_length_sorted.txt

$ join -1 1 -2 1 example_sorted.bed example_length_sorted.txt
chr1 26 39 58352
chr1 53 84 58352
chr1 9 28 58352
chr2 10 19 39521
chr3 32 99 24859
```

Comparing files

- diff – reports differences between files.
- usage: diff [OPTIONS] FILE1 FILE2
- useful options:
  - -b – ignore blanks
  - -w – ignore white spaces and tabs
  - -i – ignore case
  - -r – recursively compare all files (when comparing folders)
  - -y – side by side comparison of files
- normal output shows only the lines that are different between 2 files: $<$ *FROM-FILE-LINE* $>$ *TO-FILE-LINE*

diff example

```
$ diff -y genList1.txt genList2.txt
1APM:I                          1APM:I
1APM:E                          1APM:E
1AY6:I                          1AY6:I
1AY6:K                        | 1AY6:H
1AY6:L                          1AY6:L
1BH3:A                          1BH3:A
1BRR:D                        | 1BRR:A
1BRR:B                          1BRR:B
1BRR:C                          1BRR:C
1BXW:A                          1BXW:A
```

Interpreting the output

```
$ diff genList1.txt genList2.txt
gen4c4
< 1AY6:K
---
> 1AY6:H
7c7
< 1BRR:D
---
> 1BRR:A
```

- 1st line – **gen4c4**:
  - c – changed (a – added, d – deleted)
  - left number – line numbers of the original file
  - right number – line numbers of the modified file
- 2nd line – < **1AY6:K** – shows lines from the first file that are different from the second file
- 3rd line – a divider
- 4th line – > **1AY6:H** – shows lines from the second file that are different from the first one

Downloading data

- wget – is a tool for non-interactive download of files from the Web (http, https, ftp)
- wget 'http://website.url'

```
$ wget "http://www.rcsb.org/pdb/files/1ema.pdb"
--2015-08-14 17:50:59--  http://www.rcsb.org/pdb/files/1ema.pdb
Resolving www.rcsb.org (www.rcsb.org)... 128.6.70.10
Connecting to www.rcsb.org (www.rcsb.org)|128.6.70.10|:80... connected.
HTTP request sent, awaiting response... 200 OK
Length: unspecified [text/plain]
Saving to: 1ema.pdb.1
    [   <=>   ] 191 403      389KB/s   in 0,5s
2015-08-14 17:50:59 (389 KB/s) - 1ema.pdb.1 saved [191403]
```

Example curl

- curl – is a tool to transfer data from or to a server (http, https, ftp, imap, pop3)

  → the data is downloaded directly to the screen

  curl 'http://website.url'

```
$ curl "http://www.rcsb.org/pdb/files/1ema.pdb"
% Total % Received % Xferd  Average Speed Time Time Time  Current Dload Upload
      Total Spent Left Speed
0 0 0 0 0 0 0 0 --:--:-- --:--:-- --:--:--  0HEADER FLUORESCENT PROTEIN  01-AUG
     -96   1EMA
TITLE      GREEN FLUORESCENT PROTEIN FROM AEQUOREA VICTORIA
COMPND     MOL_ID: 1;
...
```

```
$ curl "http://www.rcsb.org/pdb/files/1ema.pdb" > 1ema.pdb
$ curl "http://www.rcsb.org/pdb/files/1ema.pdb" -o 1ema.pdb
$ curl "http://www.rcsb.org/pdb/files/{1ema,1gfl,1g7k,1xmz}.pdb" -o '#1'.pdb
```

Summary

- history – returns a history of used command lines
- clear – clears the screen
- join – join different files together by a common column
- column – formats the input into multiple columns
- df – shows how much disk space is available
- du – shows how much disk space is taken by your files
- diff – reports differences between files
- wget,curl – download data from an url from the shell