Introduction to Unix

Part I: The Essentials

Frederick J Tan
Bioinformatics Research Faculty
Carnegie Institution of Washington, Department of Embryology

9 April 2013

Unix, Linux, Ubuntu, Oh My!

A Three Hour Tour

Part I: The Essentials

client-server model, command-line interface, navigation,
viewing files, searching, finding help

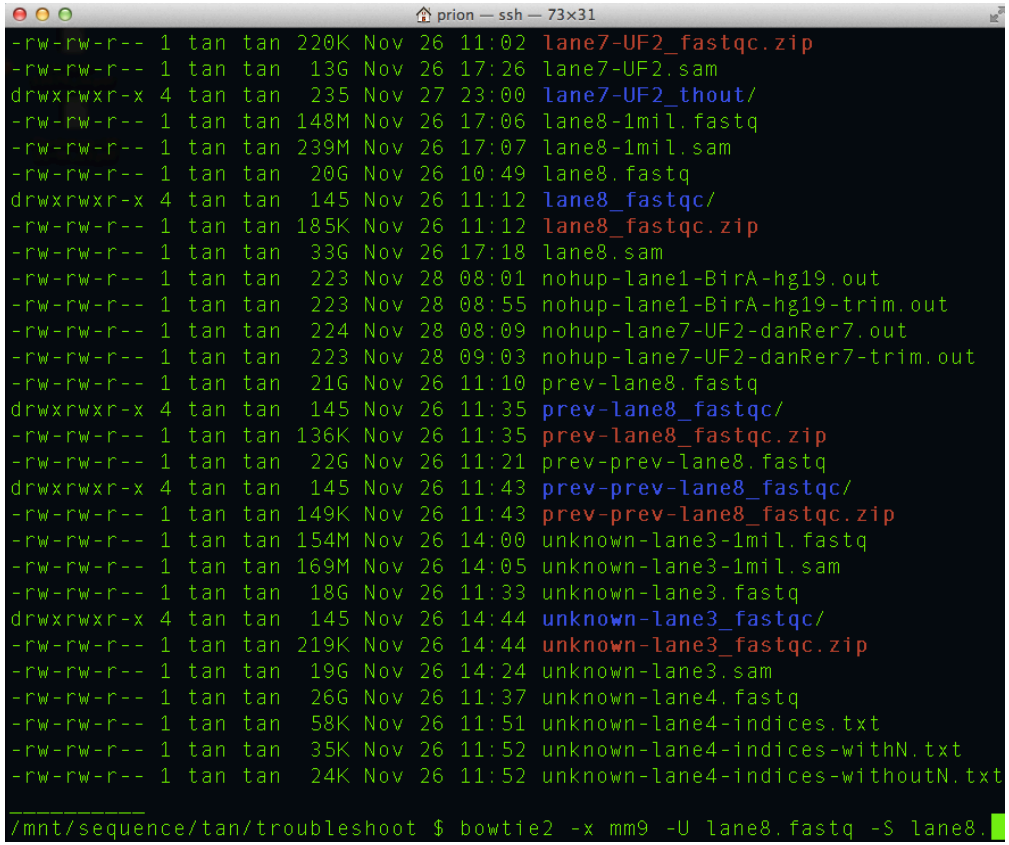
Part II: Special Topics

account settings, multi-tasking, programming,
installing programs, file systems, system administration

The Awesome Power of the Client-Server Model

A Tale of Two Interfaces

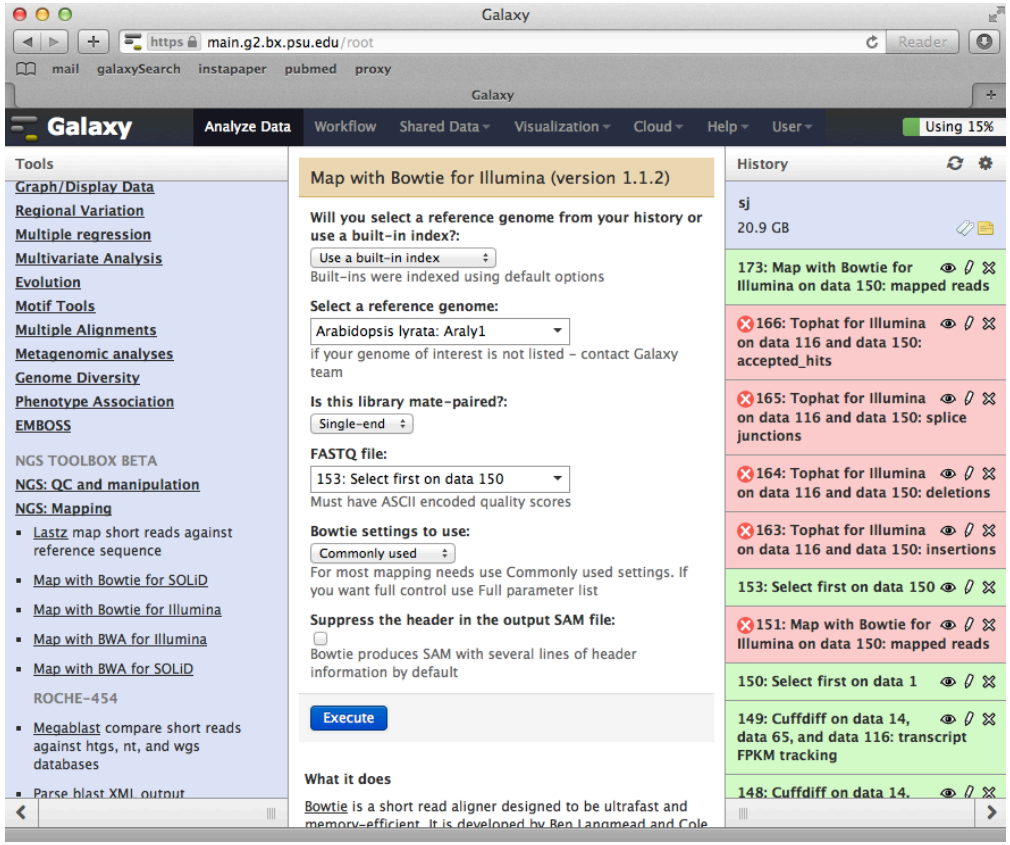
Command Line



Terminal.app, Unix shell

breadth, cutting-edge

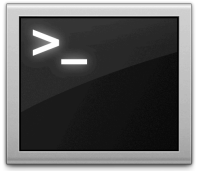
Graphical



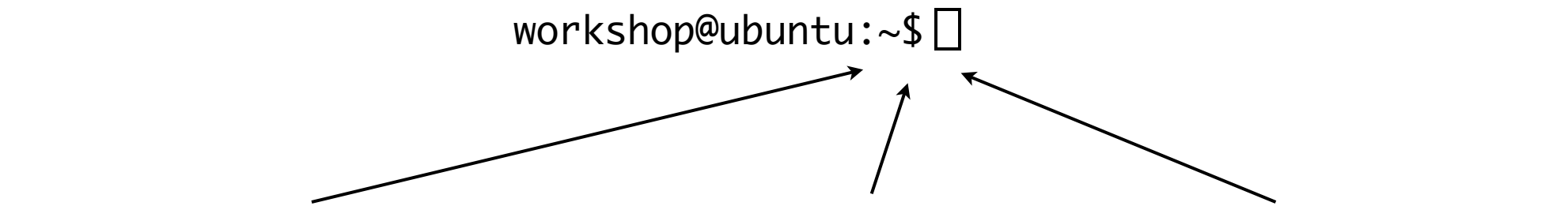
Finder.app, Galaxy

discovery, visualization

Running Programs Using the Command-Line Interface

	 command-line	 graphical
to run a program	type in the name of the program and hit <ENTER> (e.g. `bowtie2`)	double click on an icon
to modify how a program operates	type in options after the program, before hitting <ENTER> (e.g. `bowtie2 --very-sensitive`)	click on check boxes, select from pull-down menus

The Anatomy of a Shell Prompt



The text in between the colon (:) and the dollar sign ($) indicates what directory you are in.

The $ symbol indicates that the server is ready to perform a command.

The symbol indicates where what you type in will appear.

```
/home/workshop$
```

```
$ ssh workshop@192.168.56.101
```

Task 1 : Connect to your server and start `top` with `-i` option

Start VirtualBox

Start Ubuntu

Secure SHell

```
$ ssh workshop@192.168.56.101 <ENTER>
```

Shell Prompt

```
/home/workshop$
```

Start top -i

```
$ top -i
```

<SPACE>
<TAB>
<CTRL>
<UP>
<DOWN>

Task 2: Figure out what what `top` and the `-i` option does

Quit top

```
press ‘q’ to quit
```

Shell Prompt

```
/home/workshop$
```

Manual Pages

```
$ man top
```

```
(h)elp
(q)uit
(f)orward
(b)ack
```

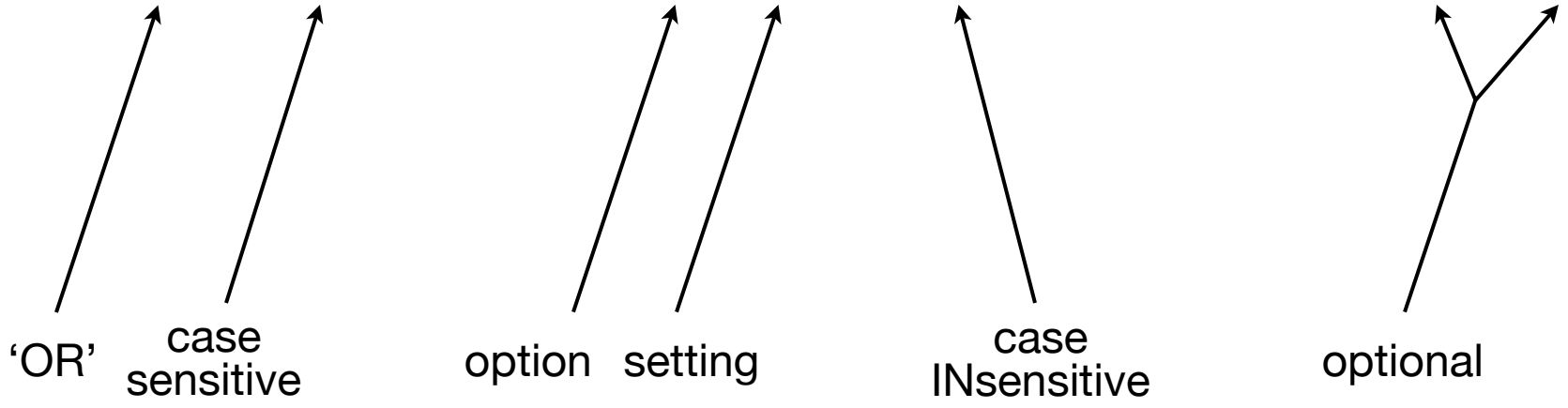
Divining Syntaxes - Example I

```
$ man top

TOP(1)                          User Commands                          TOP(1)

NAME
       top - display Linux processes

SYNOPSIS
       top -hv|-bcHisS -d delay -n limit -u|U user -p pid -w [cols]
```



Divining Syntaxes - Example II

```
$ bowtie

No index, query, or output file specified!
Usage:
  bowtie [options]* <ebwt> {-1 <m1> -2 <m2> | --12 <r> | <s>} [<hit>]
```

Document, Document, Document

130409-introToUnix-notes.rtf

Helvetica | Bold | 15 | 1.0

4/9/13

Start Ubuntu virtual machine instance

Start VirtualBox (v4.2.8)
... from http://download.virtualbox.org/virtualbox/4.2.10/VirtualBox-4.2.10-84104-OSX.dmg

Start Ubuntu 12.10 mini instance
... from https://jshare.johnshopkins.edu/Ubuntu12.10-mini.zip

Connect to Ubuntu instance

Open Terminal.app

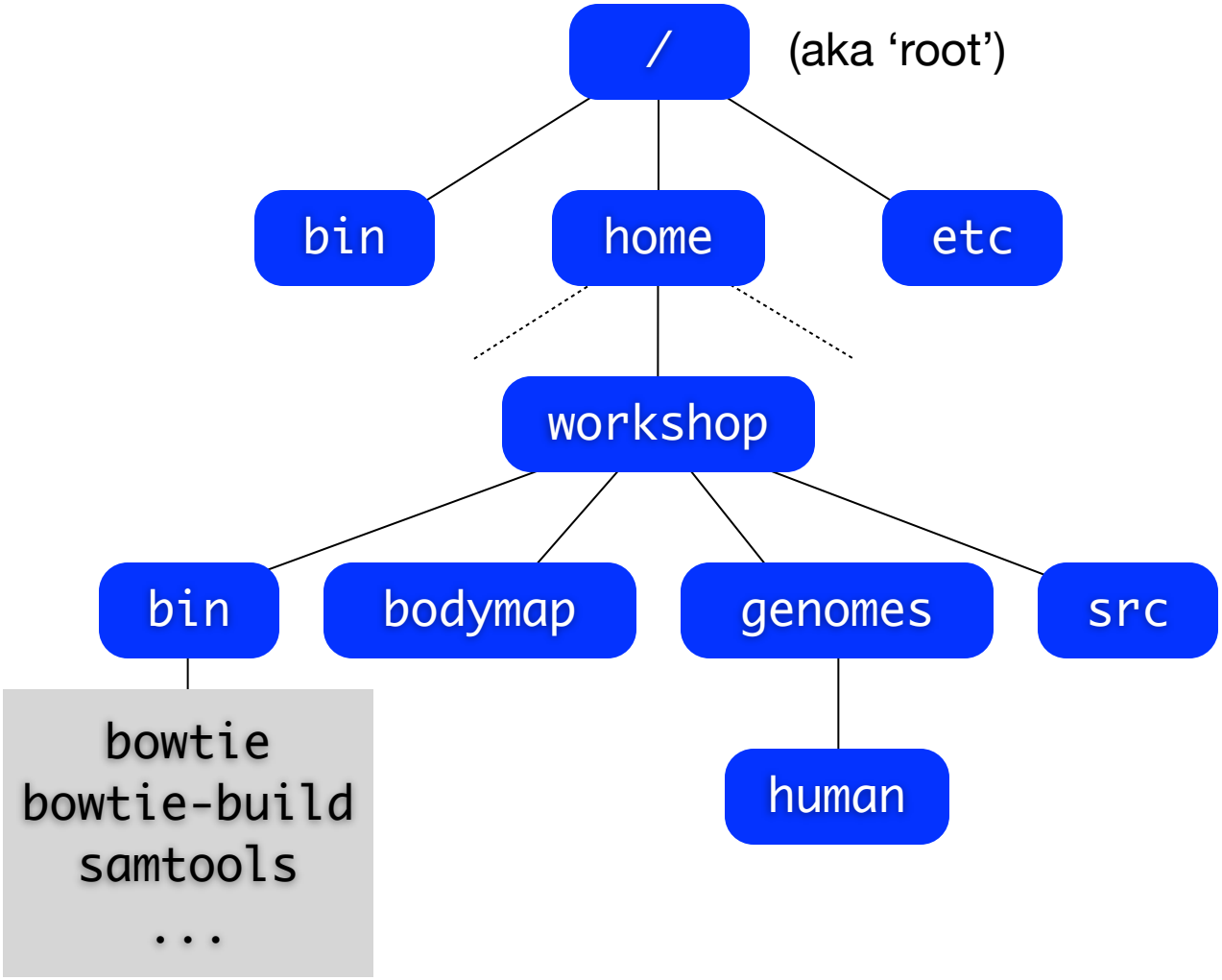
ssh workshop@192.168.56.101

Task #1 and #2: top and man pages

top -i
... press 'q' to quit

man top
... press 'f' to forward, 'b' to back
... top shows list of running programs
... -i ignores programs that are idle (i.e. not running)!

Unix Directory Structure



Task 3: List directory contents

```
$ clear
```

```
$ ls
```

```
$ ls bin
```

Color Coding

```
$ ls src
```

File Extensions

```
$ ls /home/workshop
```

```
$ ls /
```

```
$ ls bin/bowtie-b*
```

Wildcard

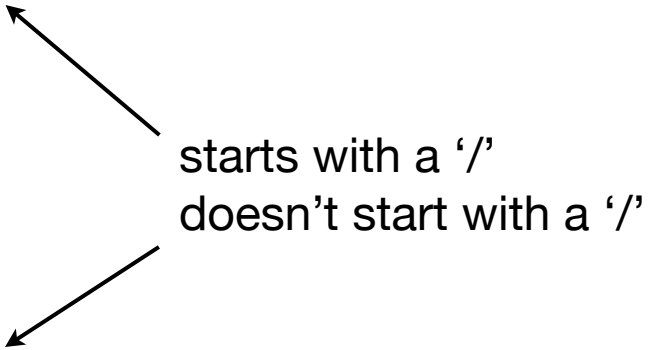
```
$ ls -l bin
```

Permissions

Task 4: Figure out how to sort contents of `/home/workshop/bin` by date

How to Get There from Here

Absolute `/home/workshop$ ls /home/workshop/bin`



Relative `/home/workshop$ ls bin`

Task 5: Change current working directory

```
$ clear
$ pwd
$ ls bodymap/data
$ cd bodymap
$ pwd
$ ls data
$ cd data
$ ls
```

```
$ ls .
$ ls ..
$ ls ~
$ pwd
$ ls
$ cd ../bin
$ cd ../../bin
$ ls
```

Path Resolution

Task 6: Figure out multiple ways to change to `/home/workshop/src`

Source, Destination; Source, Destination; Source, Destination; Source, Destination;

```
$ man cp

NAME
       cp - copy files and directories

SYNOPSIS
       cp [OPTION]... [-T] SOURCE DEST
```

```
       mv [OPTION]... [-T] SOURCE DEST
```

```
    scp [-12346BCpqrv] [-c cipher] [-F ssh_config] [-i identity_file]
        [-l limit] [-o ssh_option] [-P port] [-S program]
        [[user@]host1:]file1 ... [[user@]host2:]file2
```

WARNING

Unix assumes you know what you are doing.
Default settings will overwrite files if they exist.
There is no undo, Trash can, Recycle Bin, etc.

try using -i or -n flag

Task 7: Organize files into directories

```
$ cd
```
Defaults to $HOME

```
$ mkdir newproject
```

```
$ mkdir newProject
```
Case Sensitive

```
$ touch newfile
```

```
$ ls
```

```
$ mv newfile newproject
```

```
$ ls
```

Task 8: Work more efficiently

```
$ touch asdflhasdflbjkasdfg
```

```
$ touch ziuyxbciouyxcvbyv
```

```
$ mv a<TAB> n<TAB><TAB><TAB>P<TAB>
```

```
$ <UP><UP><LEFT><LEFT>...<LEFT>
  <BACKSPACE>{5}mv
  <CTRL>-E<SPACE>n<TAB>p<TAB>
```

Task 9: Remove files and directories

```
$ rmdir newproject
```

```
$ rm newproject/<TAB>
```

```
$ rmdir newproject
```

Task 10: Find files related to chromosome 22

```
$ man find
```

```
$ find . -name “*chr22*”
```

Task 11: Figure out how to change working directory to place with `hg19-chr22.fa` in 8 keys

Task 12: View contents of `hg19-chr22.fa`

```
$ head hg19-chr22.fa
```

```
$ cat hg19-chr22.fa
```

<CTRL>-C

```
$ wc hg19-chr22.fa
```

```
$ tail hg19-chr22.fa
```

```
$ less hg19-chr22.fa
```

(h)elp
(f)orward
(b)ack
(q)uit

/GATTACA<ENTER>

(/)search

Task 13: Figure out what is on line 321002 of `hg19-chr22.fa`

Task 14: What happens if you try and view the contents of a program?

```
$ head ~/bin/bowtie
```

```
$ less ~/bin/bowtie
```

Task 15: Find all instances of `GATTACA` in `hg19-chr22.fa`

```
$ grep GATTACA hg19-chr22.fa
```

Task 16: Figure out how to configure `grep` to tell you....

How many lines contain GATTACA?

Which lines contain GATTACA?

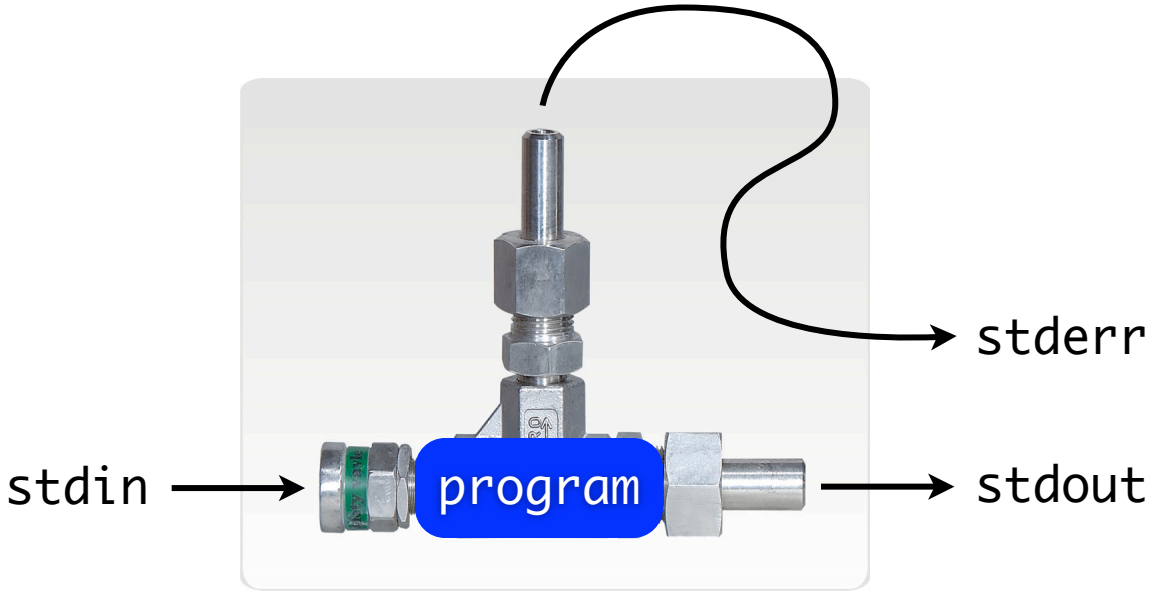
What is the line preceeding the line with GATTACA?

How many lines do **not** have GATTACA?

Match Making with Regular Expressions

0 or more	`$ grep "A*GATTACA" hg19-chr22.fa`
Wildcard	`$ grep "GATT.CA" hg19-chr22.fa`
Repetition	`$ grep "GATTA\{0,5\}CA" hg19-chr22.fa`
'OR'	`$ grep "[ACT]GATTACA" hg19-chr22.fa`
Range	`$ grep "[a-z]GATTACA" hg19-chr22.fa`
Class	`$ grep "[[:lower:]]GATTACA" hg19-chr22.fa`
'NOT'	`$ grep "[^ACT]GATTACA" hg19-chr22.fa`
Start of line	`$ grep "^GATTACA" hg19-chr22.fa`
End of line	`$ grep "GATTACA$" hg19-chr22.fa`

Streams, Pipelines, and Workflows



Redirect

```
$ head hg19-chr22.fa > first10.fa
```

```
$ less first10.fa
```

Redirect (append)

```
$ head hg19-chr22.fa >> first10.fa
```

```
$ less first10.fa
```

Pipe

```
$ head hg19-chr22.fa | less
```

Task 17: Figure out what these commands are doing

```
$ cd ~/bodymap
```

```
$ zcat data/ERR030876-chr22-00*_1.fastq.gz > reads.fastq
```

```
$ grep ^GATTACA reads.fastq | cut -dA -f1,2,3,4 \
  --output-delim=‘ ’ | sort -k4 | uniq -c > gattaca-motif.txt
```

Task 18: Copy `reads.fastq` and `gattaca-motif.txt` to your laptop

```
$ scp workshop@192.168.56.101:/home/workshop/bodymap/reads.fastq .
```

Topics Covered Today

Working with a server	connect over ssh, copy files using scp
Command line interface	prompt, running (& config) programs, <TAB>, history, streams
Navigation	paths, working directory, extensions, move/copy, wildcards
Viewing files	head, tail, less, cat, <CTRL>-C
Searching	find, less, grep, regular expressions
Finding help	man pages, syntax conventions
Best practices	electronic notebook, folder structures
System administration	top, file permissions, directory structure

Topics Covered Next Time

Account settings	passwd, .bashrc, $PATH
Programming	variables, text editors, shell scripts, chmod, AWK
Multi-tasking	ps, kill, <CTRL>-Z, nohup
Installing programs	tar, gunzip, apt-get
File systems	hidden files, whereis, df, du, system directories
System administration	ping, sudo, adduser

Useful Websites

Unix commands	linux.about.com/library/cmd/blcmdl1_export.htm
Unix utilities	en.wikipedia.org/wiki/List_of_Unix_utilities
Unix system directories	en.wikipedia.org/wiki/Unix_directory_structure
Unix habits	ibm.com/developerworks/aix/library/au-badunixhabits.html
grep	robelle.com/smugbook/regexpr.html
Unix questions	stackoverflow.com
Sys admin questions	serverfault.com
Bioinformatics questions	biostars.org
NexGenSeq questions	seqanswers.com
Workshop notes	emb.carnegiescience.edu/labs/frederick-tan