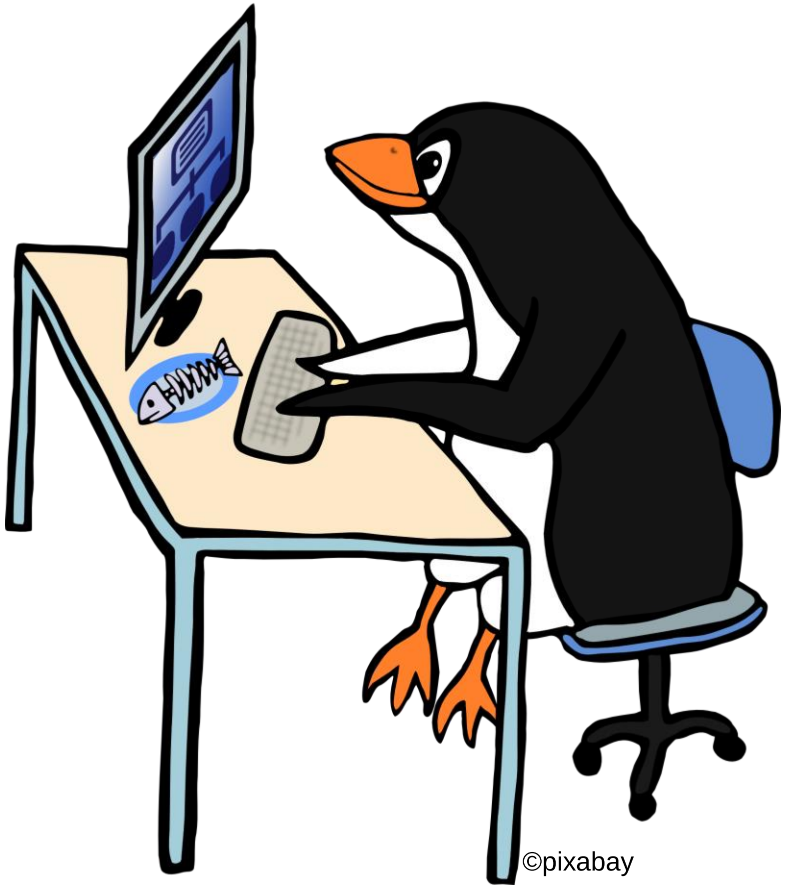


©pixabay

Fun with Linux!

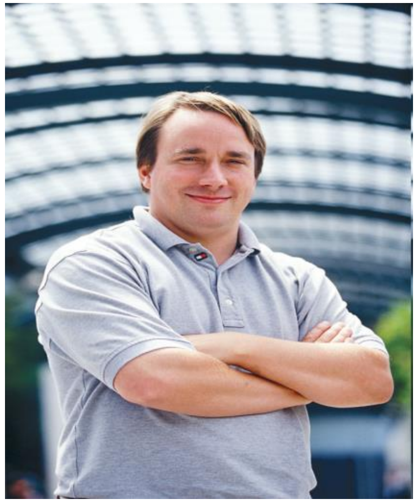
Rahadian Pratama
Department of Biochemistry
IPB University

in collaboration with

Institut Bioinformatika Indonesia
(INBIO - Indonesia)

Linux from Linus

Linux is a free operating system (OS) based on **UNIX** OS. It was developed by **Linus Torvald** in 1991



It is free as you can distribute with no cost!!!

Linux OS

- A fully-networked 32/64-Bit Unix-like Operating System
- Multi-user, Multitasking, Multiprocessor
- Has the X Windows GUI
- Coexists with other Operating Systems
- Runs on multiple platforms
- Includes the Source Code

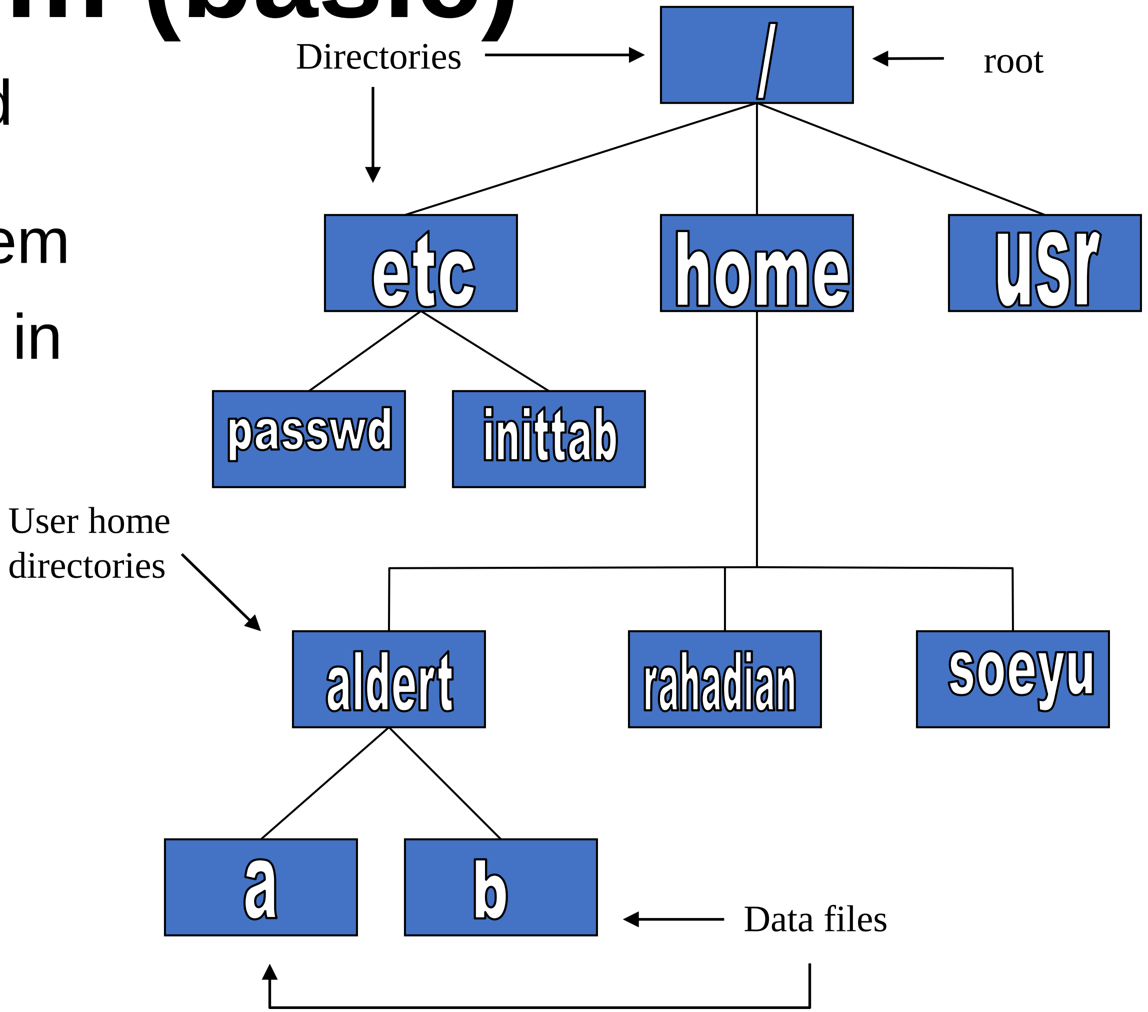
Linux distribution

Today there are hundreds of different distributions available popular Linux distributions include

- SUSE Linux
- Fedora Linux
- Red Hat Enterprise Linux
- Debian Linux
- Linspire
- Gentoo Linux
- Slackware Linux
- TurboLinux
- Mandrake Linux
- Lycoris Linux
- CentOS
- ALT Linux
- Ubuntu

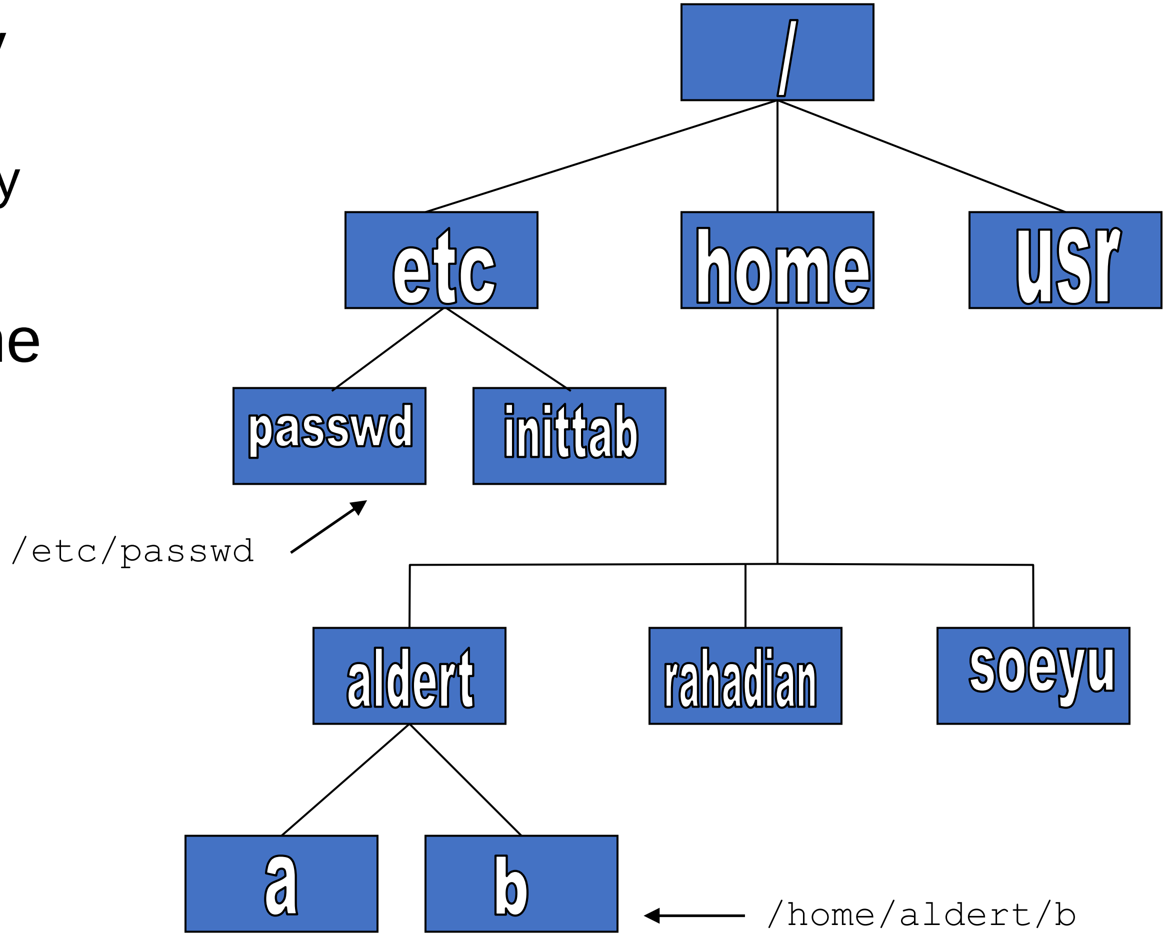
Linux file system (basic)

- Linux files are stored in a single rooted, hierarchical file system
- Data files are stored in directories (folders)
- Directories may be nested as deep as needed



Naming files in Linux

- Files are named by
  - naming each containing directory
  - starting at the root
- This is known as the *pathname*



Absolute path:
/home/aldert/a

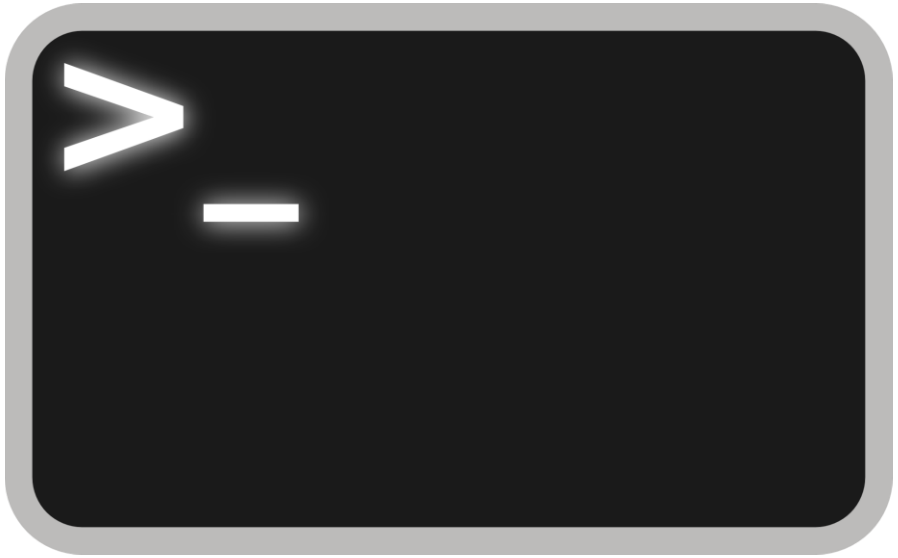
Relative path:
~/a

Some special file names

- Some file names are special:
  - / The root directory (not to be confused with the root user)
  - . The current directory
  - .. The parent (previous) directory
  - ~ My home directory
- Examples:
  - ./a same as a
  - ../jane/x go up one level then look in directory jane for x

Linux terminal

Terminal or often calles shell / command prompt, is where the user able to interact directly with the OS



Commands in terminal

Some basic commands in linux terminal often used:

- **cd** - to change directory
- **ls** - list the files in the directory
- **mkdir** - create directory
- **rm** - remove file(s)
- **cp** - copy a file from one location to other
- **mv** - move file to other location. Can rename file

COINCIDE Utrecht University The Fleming Fund

Nanopore Microbial Genomics Course

Advanced (Day 3-5)

Soe Yu Naing
Rahadian Pratama
Aldert Zomer

Learning objectives

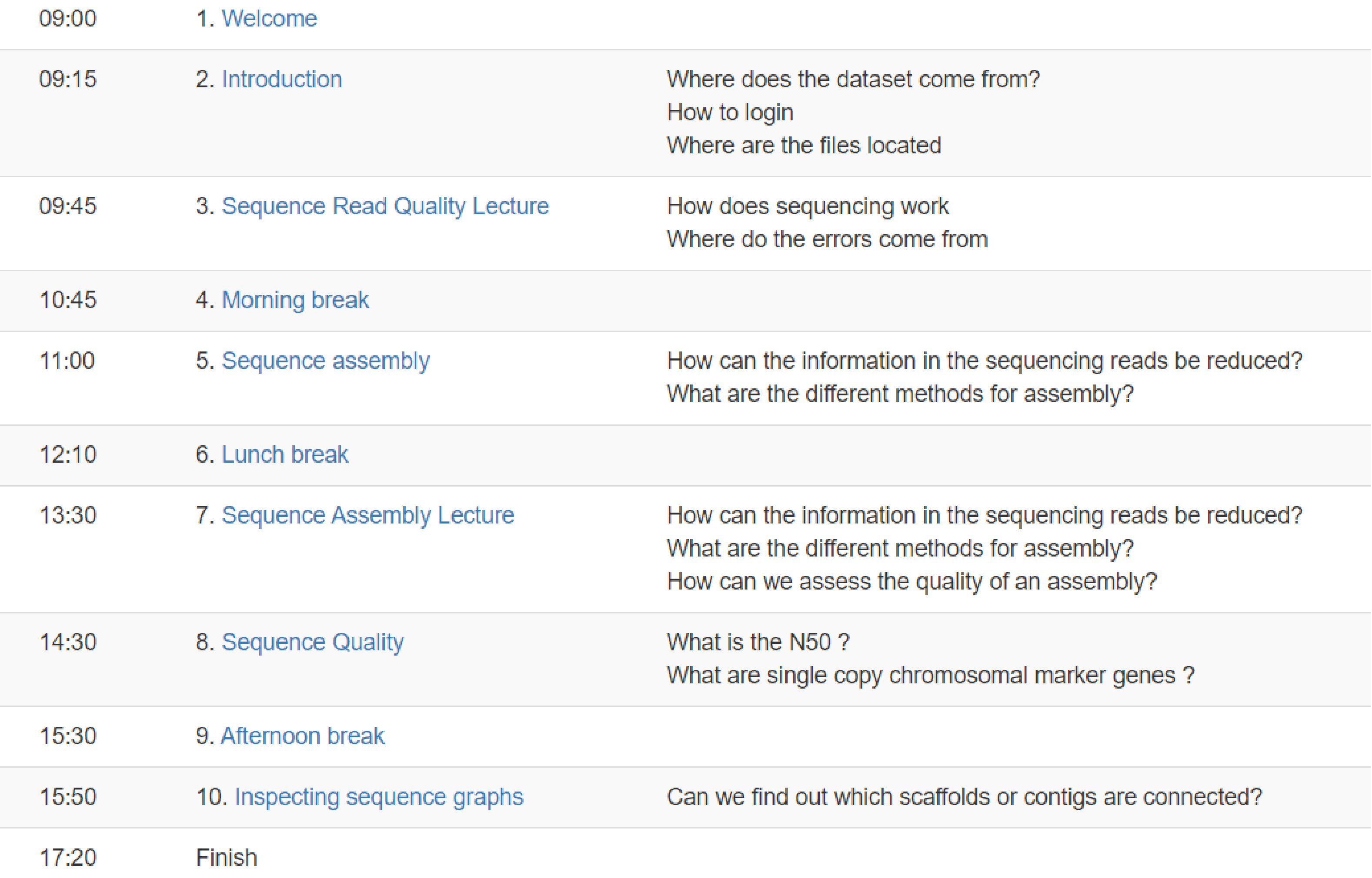
- How to assemble genomes and check them
- Determine MLST, detected resistance genes, plasmids for genomic epidemiology
- How to annotate genomes and work with annotations
- How to make phylogenetic trees of genome sequences
- Understand the concept of the pan genome and the core genome and associate gene presence/absence with a phenotype (e.g. resistance) using contingency testing (“GWAS”)

COINCIDE

Utrecht University

The Fleming Fund

Part 1 – Genome assembly



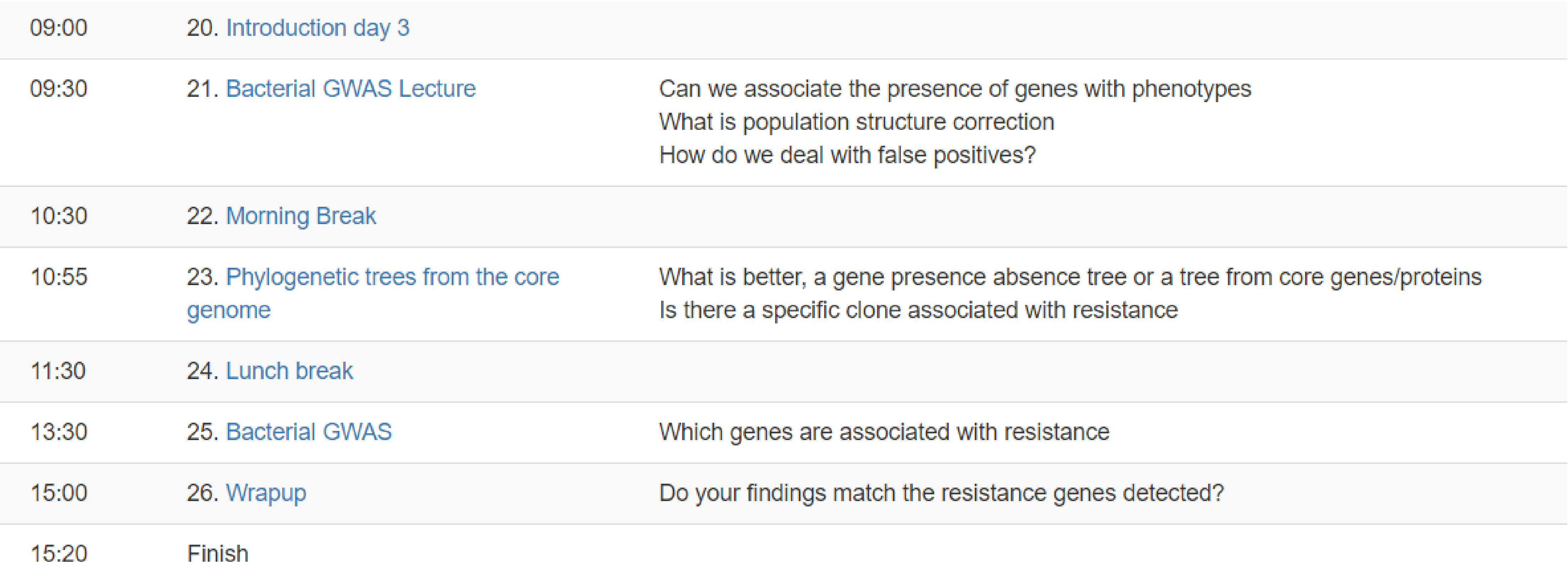
09:00	1. Welcome	
09:15	2. Introduction	Where does the dataset come from? How to login Where are the files located
09:45	3. Sequence Read Quality Lecture	How does sequencing work Where do the errors come from
10:45	4. Morning break	
11:00	5. Sequence assembly	How can the information in the sequencing reads be reduced? What are the different methods for assembly?
12:10	6. Lunch break	
13:30	7. Sequence Assembly Lecture	How can the information in the sequencing reads be reduced? What are the different methods for assembly? How can we assess the quality of an assembly?
14:30	8. Sequence Quality	What is the N50 ? What are single copy chromosomal marker genes ?
15:30	9. Afternoon break	
15:50	10. Inspecting sequence graphs	Can we find out which scaffolds or contigs are connected?
17:20	Finish	

Part 2 – Genomic Epidemiology

09:00	11. Introduction day 2	
09:30	12. MLST	How can we detect the MLST type
10:10	13. Morning Break	
10:35	14. Resistance genes	How can we detect the resistance genes and point mutations
11:35	15. Plasmid typing and classification	How can we detect which contig is a plasmid and which is from the chromosome
12:35	16. Lunch break	
13:55	17. Annotation	How are proteins predicted from a DNA sequence?
14:55	18. Afternoon break	
15:20	19. Pangenome analysis	How to determine a pangenome from a collection of isolate genome sequences?
16:20	Finish	

COINCIDE Utrecht University BPMSPH The Fleming Fund

Part 3 Pan genomes and GWAS



09:00	20. Introduction day 3	
09:30	21. Bacterial GWAS Lecture	Can we associate the presence of genes with phenotypes What is population structure correction How do we deal with false positives?
10:30	22. Morning Break	
10:55	23. Phylogenetic trees from the core genome	What is better, a gene presence absence tree or a tree from core genes/proteins Is there a specific clone associated with resistance
11:30	24. Lunch break	
13:30	25. Bacterial GWAS	Which genes are associated with resistance
15:00	26. Wrapup	Do your findings match the resistance genes detected?
15:20	Finish	

Course outline

- Working in groups of two
- Web browser needed (Chrome, Edge, Firefox, etc)
- https://klif2.uu.nl/

Dataset:

- 35 E. coli isolates
- Colistin resistant or sensitive
- Nanopore sequenced at BPMSPH

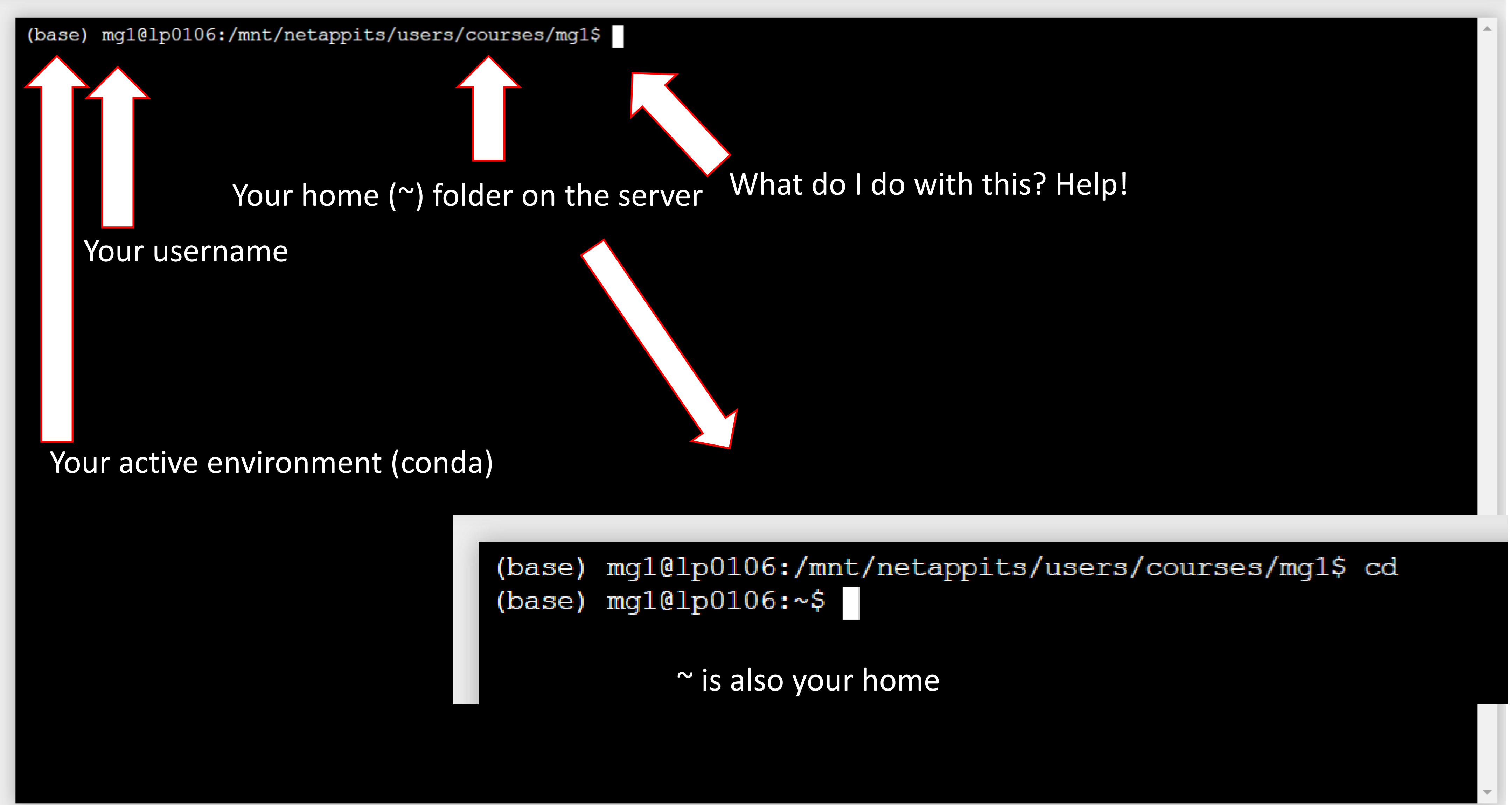
COINCIDE Utrecht University BPMSPH The Fleming Fund

Commandline/shell/terminal/CLI

Select items to perform actions on them.

☐ 0 /

☐ annotation
☐ assembly
☐ Desktop
☐ old
☐ reads
☐ barcode02.fasta 18 dagen geleden 0 B
☐ barcode03.fasta 18 dagen geleden 0 B



```
(base) mg1@lp0106:/mnt/netappits/users/courses/mg1$
```

Your active environment (conda)

Your username

Your home (~) folder on the server

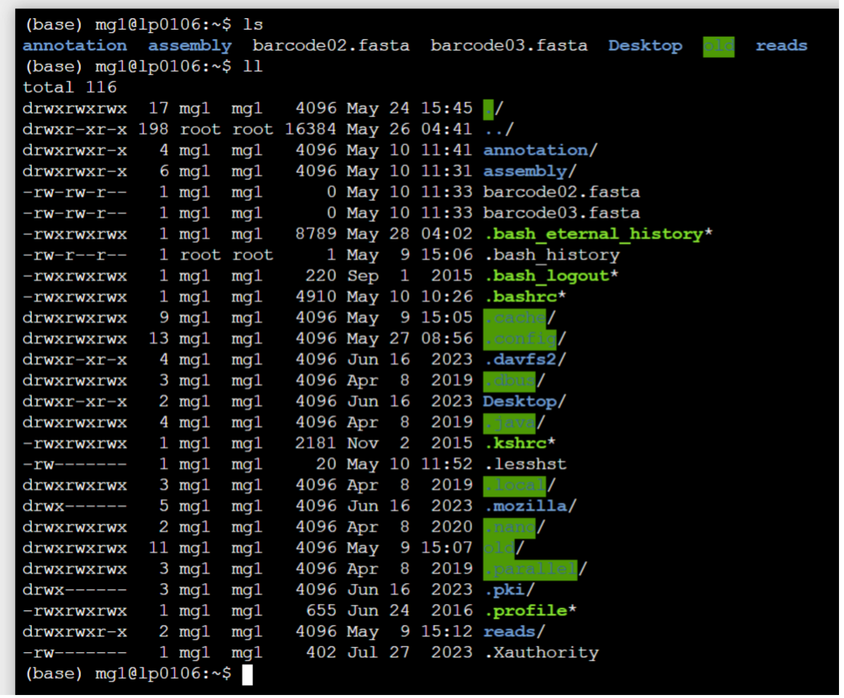
What do I do with this? Help!

```
(base) mg1@lp0106:/mnt/netappits/users/courses/mg1$ cd
(base) mg1@lp0106:~$
```

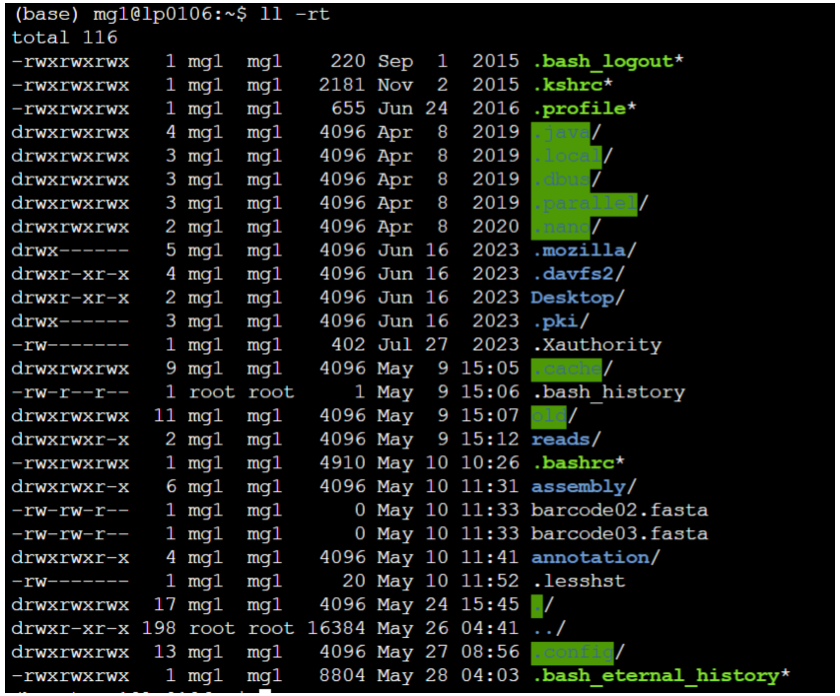
~ is also your home

Listing files

```
ls          #list filenames. Do not list hidden files (files starting with .)
ls -laF     #list your files with details
ll          # shortcut to list your files with details
ll -rt      # list your files sorted with oldest at top, newest at bottom
```



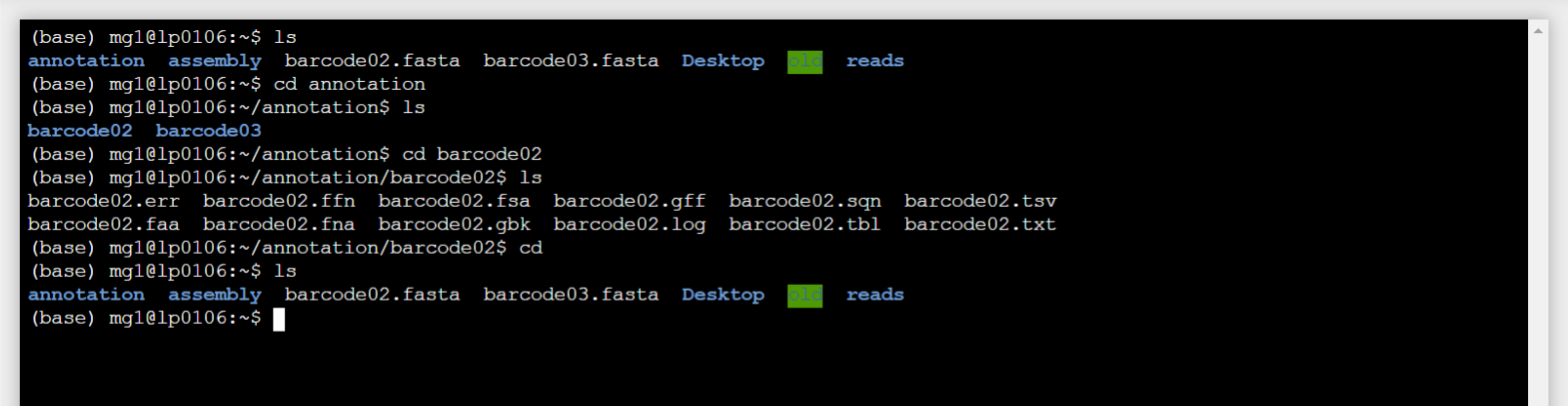
```
(base) mg1@lp0106:~$ ls
annotation  assembly  barcode02.fasta  barcode03.fasta  Desktop  old  reads
(base) mg1@lp0106:~$ ll
total 116
drwxrwxrwx  17 mg1  mg1   4096 May 24 15:45 ./
drwxr-xr-x 198 root root 16384 May 26 04:41 ../
drwxrwxr-x   4 mg1  mg1   4096 May 10 11:41 annotation/
drwxrwxr-x   6 mg1  mg1   4096 May 10 11:31 assembly/
-rw-rw-r--   1 mg1  mg1      0 May 10 11:33 barcode02.fasta
-rw-rw-r--   1 mg1  mg1      0 May 10 11:33 barcode03.fasta
-rwxrwxrwx   1 mg1  mg1   8789 May 28 04:02 .bash_eternal_history*
-rw-r--r--   1 root root     1 May  9 15:06 .bash_history
-rwxrwxrwx   1 mg1  mg1    220 Sep  1  2015 .bash_logout*
-rwxrwxrwx   1 mg1  mg1   4910 May 10 10:26 .bashrc*
drwxrwxrwx   9 mg1  mg1   4096 May  9 15:05 .cache/
drwxrwxrwx  13 mg1  mg1   4096 May 27 08:56 .config/
drwxr-xr-x   4 mg1  mg1   4096 Jun 16  2023 .davfs2/
drwxrwxrwx   3 mg1  mg1   4096 Apr  8  2019 .dbus/
drwxr-xr-x   2 mg1  mg1   4096 Jun 16  2023 Desktop/
drwxrwxrwx   4 mg1  mg1   4096 Apr  8  2019 .java/
-rwxrwxrwx   1 mg1  mg1   2181 Nov  2  2015 .kshrc*
-rw-------   1 mg1  mg1     20 May 10 11:52 .lesshst
drwxrwxrwx   3 mg1  mg1   4096 Apr  8  2019 .local/
drwx------   5 mg1  mg1   4096 Jun 16  2023 .mozilla/
drwxrwxrwx   2 mg1  mg1   4096 Apr  8  2020 .nano/
drwxrwxrwx  11 mg1  mg1   4096 May  9 15:07 old/
drwxrwxrwx   3 mg1  mg1   4096 Apr  8  2019 .parallel/
drwx------   3 mg1  mg1   4096 Jun 16  2023 .pki/
-rwxrwxrwx   1 mg1  mg1    655 Jun 24  2016 .profile*
drwxrwxr-x   2 mg1  mg1   4096 May  9 15:12 reads/
-rw-------   1 mg1  mg1    402 Jul 27  2023 .Xauthority
(base) mg1@lp0106:~$
```



```
(base) mg1@lp0106:~$ ll -rt
total 116
-rwxrwxrwx   1 mg1  mg1    220 Sep  1  2015 .bash_logout*
-rwxrwxrwx   1 mg1  mg1   2181 Nov  2  2015 .kshrc*
-rwxrwxrwx   1 mg1  mg1    655 Jun 24  2016 .profile*
drwxrwxrwx   4 mg1  mg1   4096 Apr  8  2019 .java/
drwxrwxrwx   3 mg1  mg1   4096 Apr  8  2019 .local/
drwxrwxrwx   3 mg1  mg1   4096 Apr  8  2019 .dbus/
drwxrwxrwx   3 mg1  mg1   4096 Apr  8  2019 .parallel/
drwxrwxrwx   2 mg1  mg1   4096 Apr  8  2020 .nano/
drwx------   5 mg1  mg1   4096 Jun 16  2023 .mozilla/
drwxr-xr-x   4 mg1  mg1   4096 Jun 16  2023 .davfs2/
drwxr-xr-x   2 mg1  mg1   4096 Jun 16  2023 Desktop/
drwx------   3 mg1  mg1   4096 Jun 16  2023 .pki/
-rw-------   1 mg1  mg1    402 Jul 27  2023 .Xauthority
drwxrwxrwx   9 mg1  mg1   4096 May  9 15:05 .cache/
-rw-r--r--   1 root root     1 May  9 15:06 .bash_history
drwxrwxrwx  11 mg1  mg1   4096 May  9 15:07 old/
drwxrwxr-x   2 mg1  mg1   4096 May  9 15:12 reads/
-rwxrwxrwx   1 mg1  mg1   4910 May 10 10:26 .bashrc*
drwxrwxr-x   6 mg1  mg1   4096 May 10 11:31 assembly/
-rw-rw-r--   1 mg1  mg1      0 May 10 11:33 barcode02.fasta
-rw-rw-r--   1 mg1  mg1      0 May 10 11:33 barcode03.fasta
drwxrwxr-x   4 mg1  mg1   4096 May 10 11:41 annotation/
-rw-------   1 mg1  mg1     20 May 10 11:52 .lesshst
drwxrwxrwx  17 mg1  mg1   4096 May 24 15:45 ./
drwxr-xr-x 198 root root 16384 May 26 04:41 ../
drwxrwxrwx  13 mg1  mg1   4096 May 27 08:56 .config/
-rwxrwxrwx   1 mg1  mg1   8804 May 28 04:03 .bash_eternal_history*
```

Go to a different folder

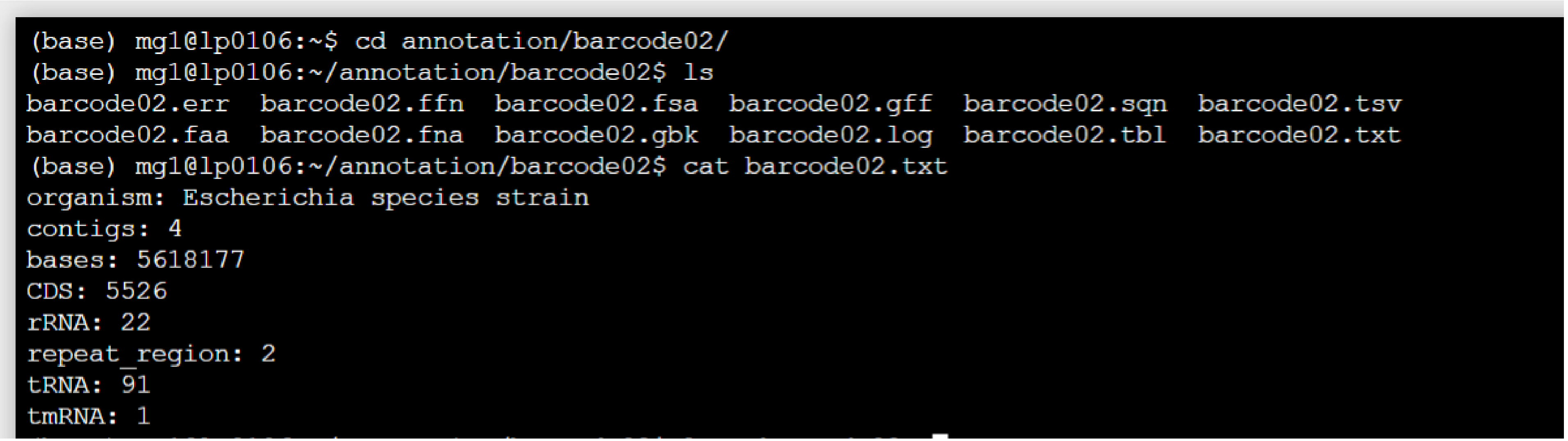
- cd # go back to your home
- cd folder # go into the folder
- cd .. # go one folder up



```
(base) mg1@lp0106:~$ ls
annotation  assembly  barcode02.fasta  barcode03.fasta  Desktop  old  reads
(base) mg1@lp0106:~$ cd annotation
(base) mg1@lp0106:~/annotation$ ls
barcode02  barcode03
(base) mg1@lp0106:~/annotation$ cd barcode02
(base) mg1@lp0106:~/annotation/barcode02$ ls
barcode02.err  barcode02.ffn  barcode02.fsa  barcode02.gff  barcode02.sqn  barcode02.tsv
barcode02.faa  barcode02.fna  barcode02.gbk  barcode02.log  barcode02.tbl  barcode02.txt
(base) mg1@lp0106:~/annotation/barcode02$ cd
(base) mg1@lp0106:~$ ls
annotation  assembly  barcode02.fasta  barcode03.fasta  Desktop  old  reads
(base) mg1@lp0106:~$ 
```

Showing a file on the screen

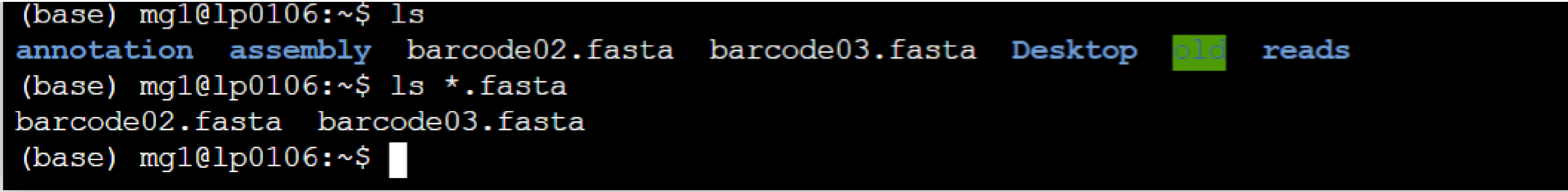
- cat file #display the whole file on the screen at once
- less file #display the file per page
- less -S file #display the file per page with scroll left/right



```
(base) mg1@lp0106:~$ cd annotation/barcode02/
(base) mg1@lp0106:~/annotation/barcode02$ ls
barcode02.err  barcode02.ffn  barcode02.fsa  barcode02.gff  barcode02.sqn  barcode02.tsv
barcode02.faa  barcode02.fna  barcode02.gbk  barcode02.log  barcode02.tbl  barcode02.txt
(base) mg1@lp0106:~/annotation/barcode02$ cat barcode02.txt
organism: Escherichia species strain
contigs: 4
bases: 5618177
CDS: 5526
rRNA: 22
repeat_region: 2
tRNA: 91
tmRNA: 1
```

“wildcards”

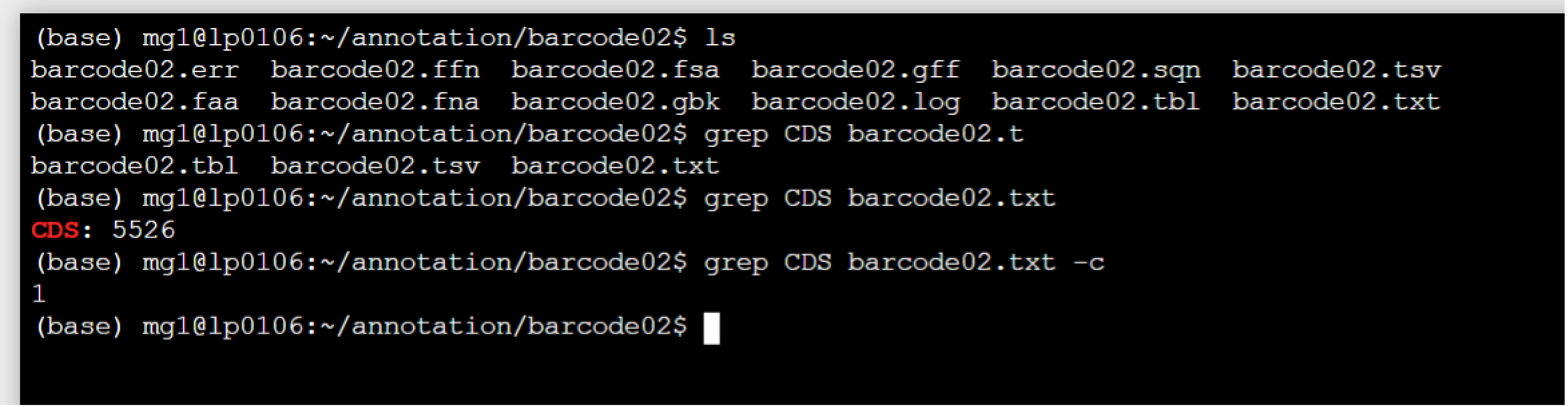
- ls *.fasta # lists all files matching .fasta . The * is a wildcard



```
(base) mg1@lp0106:~$ ls
annotation  assembly  barcode02.fasta  barcode03.fasta  Desktop  old  reads
(base) mg1@lp0106:~$ ls *.fasta
barcode02.fasta  barcode03.fasta
(base) mg1@lp0106:~$ 
```

Finding a word in a file

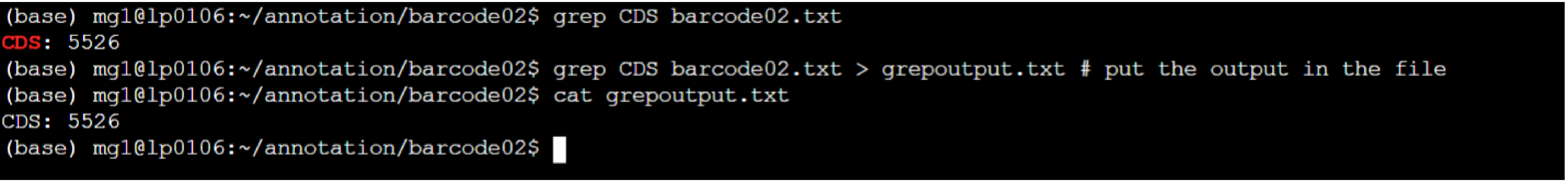
- grep word file #find the word “word” in file
- grep word file -c #find the word “word” in file and count



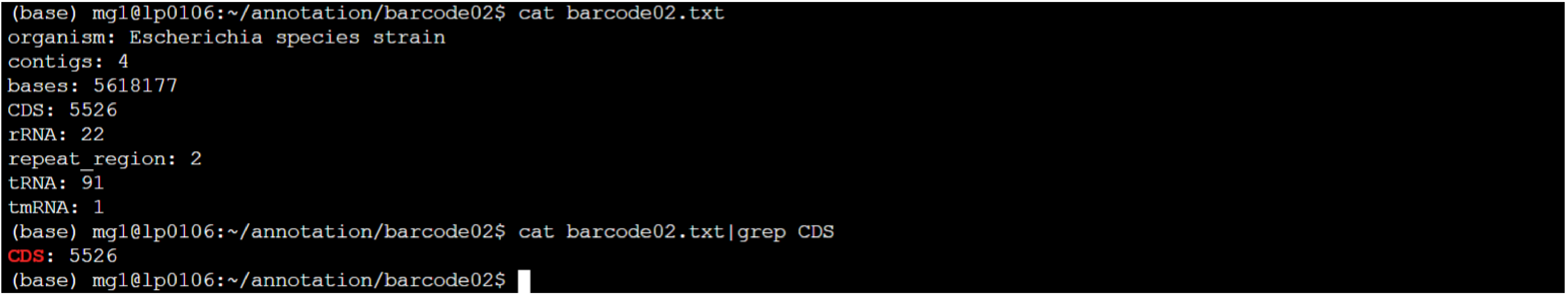
```
(base) mg1@lp0106:~/annotation/barcode02$ ls
barcode02.err  barcode02.ffn  barcode02.fsa  barcode02.gff  barcode02.sqn  barcode02.tsv
barcode02.faa  barcode02.fna  barcode02.gbk  barcode02.log  barcode02.tbl  barcode02.txt
(base) mg1@lp0106:~/annotation/barcode02$ grep CDS barcode02.t
barcode02.tbl  barcode02.tsv  barcode02.txt
(base) mg1@lp0106:~/annotation/barcode02$ grep CDS barcode02.txt
CDS: 5526
(base) mg1@lp0106:~/annotation/barcode02$ grep CDS barcode02.txt -c
1
(base) mg1@lp0106:~/annotation/barcode02$ 
```

Redirecting output to file or to another program

This is a **crucial** part. The commandline allows us to send screen output to a file with ‘>’ or to another program with ‘|’



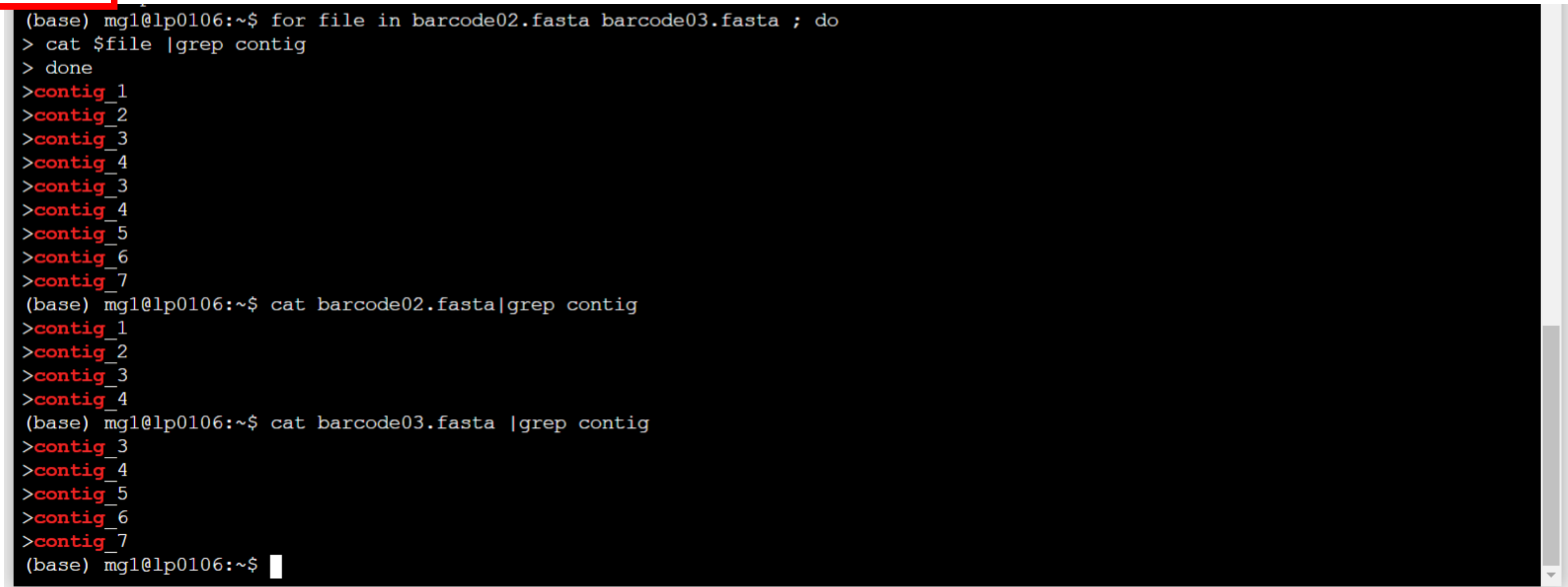
```
(base) mg1@lp0106:~/annotation/barcode02$ grep CDS barcode02.txt
CDS: 5526
(base) mg1@lp0106:~/annotation/barcode02$ grep CDS barcode02.txt > grepoutput.txt # put the output in the file
(base) mg1@lp0106:~/annotation/barcode02$ cat grepoutput.txt
CDS: 5526
(base) mg1@lp0106:~/annotation/barcode02$ 
```



```
(base) mg1@lp0106:~/annotation/barcode02$ cat barcode02.txt
organism: Escherichia species strain
contigs: 4
bases: 5618177
CDS: 5526
rRNA: 22
repeat_region: 2
tRNA: 91
tmRNA: 1
(base) mg1@lp0106:~/annotation/barcode02$ cat barcode02.txt|grep CDS
CDS: 5526
(base) mg1@lp0106:~/annotation/barcode02$ 
```

Loops – how to process many files

```
for name in sample1 sample2 sample2 ; do
        cat $name
done
```



```
(base) mg1@lp0106:~$ for file in barcode02.fasta barcode03.fasta ; do
> cat $file |grep contig
> done
>contig_1
>contig_2
>contig_3
>contig_4
>contig_3
>contig_4
>contig_5
>contig_6
>contig_7
(base) mg1@lp0106:~$ cat barcode02.fasta|grep contig
>contig_1
>contig_2
>contig_3
>contig_4
(base) mg1@lp0106:~$ cat barcode03.fasta |grep contig
>contig_3
>contig_4
>contig_5
>contig_6
>contig_7
(base) mg1@lp0106:~$ 
```

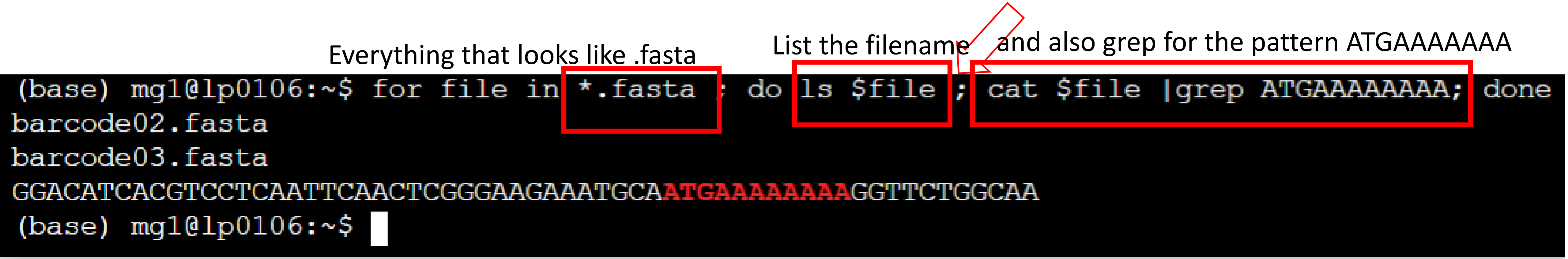
Very fancy loops

“;” takes the place of pressing enter

Everything that looks like .fasta

List the filename

and also grep for the pattern ATGAAAAAAA



```
(base) mg1@lp0106:~$ for file in *.fasta ; do ls $file ; cat $file |grep ATGAAAAAAAA; done
barcode02.fasta
barcode03.fasta
GGACATCACGTCCTCAATTCAACTCGGGAAGAAATGCAATGAAAAAAAAGGTTCTGGCAA
(base) mg1@lp0106:~$ 
```

Other commands.

- tr “a” “b” # replace a letter in an output (a with b)
- sed ‘s/contig/c/’ # replace words in an output (contig with c)
- cut -f 1 # select a column from the data (here : 1)
- echo “word” # print a word or words on a screen
- head -n 10 # print the first 10 lines
- tail -f 10 # print the last 10 lines

- conda # activate or deactivate a different software environment

https://www.codecademy.com/learn/learn-the-command-line/modules/learn-the-command-line-navigation/cheatsheet

Questions, help

- Place red post-it note on top your laptop if you need help
- Place green post-it on laptop when you have finished the exercise