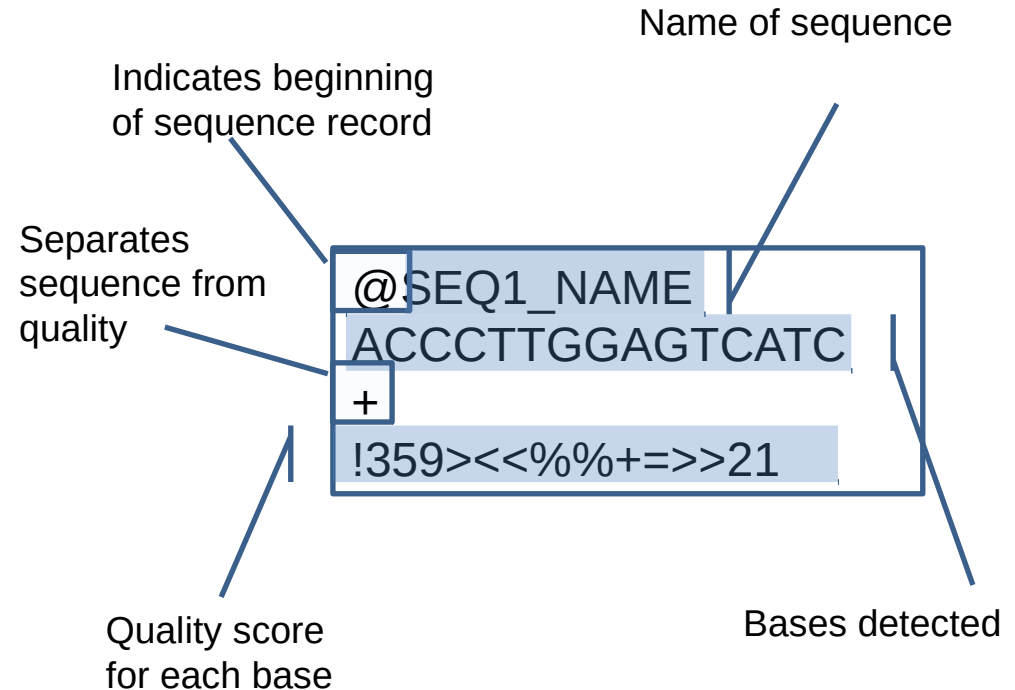


Assembly overview

Fastq files: store sequence and quality in 4 lines

- For paired end sequencing, the insert (original sequence from the sample), is sequenced twice, starting from either end.
- This results in two files for each sample called: read1 and read2.
- These are in .fastq format.
- Compressed with GNU zip to save space (.gz).

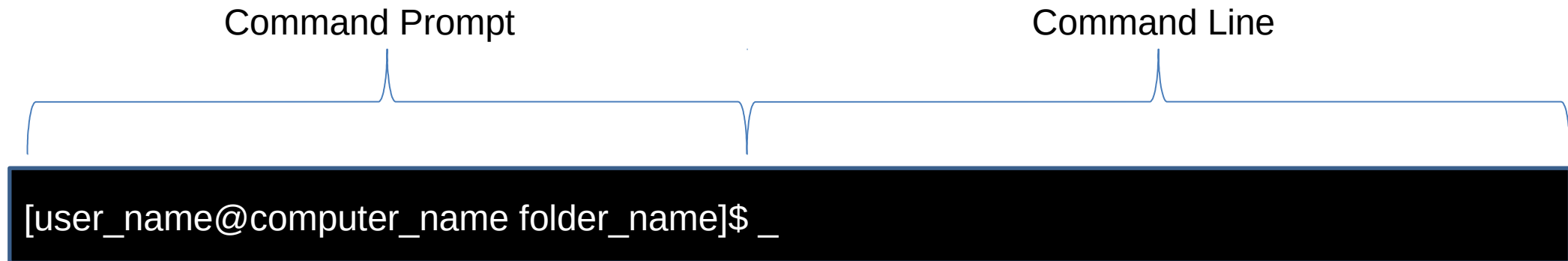


File types

.gz	Appended to files that are compressed
.fasta	Simple format for storing sequence information.
.fastq	Stores sequence and quality information

The Shell

- A Shell is a program that provides a text only user interface for interacting with the computer.
- The shell consists of a command prompt, showing the user name and location, and the command line, where commands are entered.



The Command Line

- The command prompt shows basic information.

Name of
logged in user.

Name of computer or server
where the user is logged in.

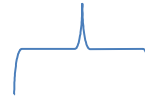
```
[user_name@computer_name folder_name]$ _
```

Current location in the
computer's file structure.

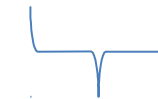
The Command Line

- Run programs and navigate files by typing commands into the command line, next to the command prompt.

Program



```
[user_name@computer_name folder_name]$ fastqc --help
```



Option

Command line caveats

- Commands are case sensitive. Enter commands exactly as written!
- Spacing and ordering of arguments are important, and can change the output of the command, so enter the commands exactly as written!
- Pressing enter runs the command as it's written. There is no warning/confirmation!
- Shells are text only interfaces. You cannot click on a space in the

Benefits of using the command line.

- Finer control of program parameters.
- Can string together multiple programs into analysis pipelines.
- Record of exactly what commands have been run.