Introduction to the Unix Shell for biologists

THIS IS WORK IN PROGRESS!

Motivation and Background

Dating back to the early 1970s.

- Possible to deal with large data sets
- Avoid repetitive work
- Reproducible
- Geeky

The basic anatonomy of a command line call

- program name + parameters + attributes

The most important tool

- Please, RTFM
- man
- apropos

Files, folder, location

- ls
- pwd
- cd
- mkdir
- type
- ~/
- Relative vs. absolute path
- Globbing

Manipulating files and folder

- mv
- cp
- rm

File content

- less / more
- cat
- head
- tail
- echo
- editors vi, emacs, nano

- wc
- sort
- grep
- cut
- uniq
- tr

File archives and compression

- tar
- gzip/gunzip
- bzip2/bunzip2
- zcat

Connecting tools - pipes

- Philosophy - build small tools that do one thing well - then connect them.

Examples analysis

XXX

```
wget ftp://ftp.ncbi.nih.gov/genbank/genomes/Bacteria/Campylobacter_jejuni_81116_uid17953/CP(
wget ftp://ftp.ncbi.nih.gov/genomes/Bacteria/Salmonella_enterica_serovar_Typhimurium_SL1344_
```

Calculate the GC content of a genome

```
wget ftp://ftp.ncbi.nih.gov/genomes/Bacteria/Salmonella_enterica_serovar_Typhimurium_SL1344_
```

```
grep -v ">" NC_016810.fna | grep -o "A" | wc -l

grep -v ">" NC_016810.fna | grep -o "C" | wc -l

grep -v ">" NC_016810.fna | grep -o "G" | wc -l

grep -v ">" NC_016810.fna | grep -o "T" | wc -l
```

or

```
grep -v ">" NC_016810.fna | grep -Po "A|T" | wc -l

grep -v ">" NC_016810.fna | grep -Po "C|G" | wc -l

echo "scale=5; 2332503/(2332503+2545509)*100" | bc
```

Run RNAfold