

Linux Commands for Bioinformatics

A compact command-line reference for navigating files, inspecting biological data and combining commands.

Core commands

Task	Command	Example / note
Where am I?	pwd	Print current directory
List files	ls	ls -lh ls -lah
Change directory	cd	cd data/ cd ..
Create directory	mkdir	mkdir results
Copy	cp	cp sample.fastq backup/
Move / rename	mv	mv old.txt new.txt
Remove file	rm	rm file.txt (use carefully)
View beginning	head	head -n 20 file.fastq
View end	tail	tail -n 20 file.txt
Count lines	wc -l	wc -l samples.tsv
Search text	grep	grep -n 'BRCA1' genes.txt
Pipe output		grep '>' seq.fa wc -l
Redirect output	>	command > results.txt
Append output	>>	command >> log.txt
Compressed text	zcat	zcat reads.fastq.gz head

Bioinformatics patterns

Goal	Example
Count FASTA records	grep -c '^>' sequences.fasta
Count FASTQ reads	echo \$((\$(wc -l < reads.fastq)/4))
Preview compressed FASTQ	zcat reads.fastq.gz head -n 8
Find FASTQ files	find . -name '*.fastq.gz'
Extract FASTA headers	grep '^>' seq.fa sed 's/>/\//' head
Sort and count values	cut -f2 table.tsv sort uniq -c sort -nr

Safety tip: Verify paths before using `rm` or overwriting with `>`. Keep raw data read-only and write outputs to a separate results directory.