

Handling Text in the Shell

Editing text files at the command line with 'nano'

```
lang@melanogaster:~$ cd ~/pcfb/sandbox  
lang@melanogaster:~$ nano
```

Enter some text into the blank document
Safe

Exit → nano saves a new file, call it shelltips.txt

```
lang@melanogaster:~$ nano shelltips.txt
```

Controlling the flow of data in the shell

Redirecting output to a file with '>'

```
lang@melanogaster:~$ cd ~/pcfb/sandbox
```

```
lang@melanogaster:~$ ls -l ../examples/*.seq
```

```
-rw-r--r-- 1 lang lang 526 2011-06-29 20:00 /perso/lang/Documents/Biocourse/pcfb/examples/FEC00001_1.seq  
-rw-r--r-- 1 lang lang 602 2011-06-29 20:00 /perso/lang/Documents/Biocourse/pcfb/examples/FEC00002_1.seq  
-rw-r--r-- 1 lang lang 540 2011-06-29 20:00 /perso/lang/Documents/Biocourse/pcfb/examples/FEC00003_1.seq  
-rw-r--r-- 1 lang lang 624 2011-06-29 20:00 /perso/lang/Documents/Biocourse/pcfb/examples/FEC00004_1.seq  
-rw-r--r-- 1 lang lang 492 2011-06-29 20:00 /perso/lang/Documents/Biocourse/pcfb/examples/FEC00005_1.seq  
-rw-r--r-- 1 lang lang 550 2011-06-29 20:00 /perso/lang/Documents/Biocourse/pcfb/examples/FEC00005_2.seq  
-rw-r--r-- 1 lang lang 497 2011-06-29 20:00 /perso/lang/Documents/Biocourse/pcfb/examples/FEC00006_1.seq  
-rw-r--r-- 1 lang lang 457 2011-06-29 20:00 /perso/lang/Documents/Biocourse/pcfb/examples/FEC00007_1.seq  
-rw-r--r-- 1 lang lang 503 2011-06-29 20:00 /perso/lang/Documents/Biocourse/pcfb/examples/FEC00007_2.seq  
-rw-r--r-- 1 lang lang 571 2011-06-29 20:00 /perso/lang/Documents/Biocourse/pcfb/examples/FEC00007_3.seq
```

```
lang@melanogaster:~$ ls -l ../examples/*.seq > files.txt
```

Saves list as a textfile « files.txt » in the sandbox folder

Controlling the flow of data in the shell

Displaying and joining files with 'cat'

```
lang@melanogaster:~$ cd ~/pcfb/sandbox
```

```
lang@melanogaster:~$ cat ../examples/FEC00001.seq
```

```
>Fe_MM1_01A01  
TCCTTGAAAGATGTACTACC ...
```

```
lang@melanogaster:~$ cat ../examples/FEC00001.seq ../examples/FEC00004.seq
```

```
>Fe_MM1_01A01  
TCCTTGAAAGATGTACTACC ...  
>Fe_MM1_01A04  
AAGTCAGTTTTCAAACGACT...
```

```
lang@melanogaster:~$ cat ../examples/*.seq
```

Displays content of any .seq file in the example folder

```
lang@melanogaster:~$ cat ../examples/*.seq > chaetognath.fasta
```

No display, file content directly read from examples and saved as chaetognath.fasta in sandbox

Controlling the flow of data in the shell

Displaying and joining files with 'cat'

```
lang@melanogaster:~$ cd ~/pcfb/sandbox
```

```
lang@melanogaster:~$ cat .. /examples/*.seq
```

Displays content of any .seq file in the example folder

```
lang@melanogaster:~$ cat .. /examples/*.seq > chaetognath.fasta
```

No display, file content directly read from examples and saved as chaetognath.fasta in sandbox

```
lang@melanogaster:~$ cat .. /examples/*1.seq > chaetognath_subset.fasta
```

Optional gathering of subsets of files

```
lang@melanogaster:~$ cat .. /examples/*2.seq >> chaetognath_subset.fasta
```

Adds subset *2.seq files into an existing file: chaetognath_subset.fasta

Regular expressions at the command line with 'grep'

Working with a larger dataset

Extracting particular rows from a file

```
lang@melanogaster:~$ cd ~/pcfb/sandbox
```

```
lang@melanogaster:~$ grep "Toolik Lake" .. /examples/shaver_etal.csv
```

```
lang@melanogaster:~$ grep "Toolik Lake" .. /examples/shaver_etal.csv > toolik.csv
```

```
lang@melanogaster:~$ grep " Aug.*Toolik Lake" .. /examples/shaver_etal.csv > toolik2.csv
```

```
lang@melanogaster:~$ grep -v "Toolik Lake" .. /examples/shaver_etal.csv
```

Redirecting output from one program to another with pipe '|'

Redirects a command output as input for another

```
lang@melanogaster:~$ history
```

Recalls a command line protocol

```
lang@melanogaster:~$ history | grep Toolik
```

Displays all the history commands that contain the word Toolik

```
lang@melanogaster:~$ grep "Aug"  
/perso/lang/Documents/Biocourse/pcfb/examples/shaver_etal.csv | grep "Toolik" >  
toola2.csv
```

Combines two consecutive searches

Searching across multiple files with 'grep'

```
lang@melanogaster:~$ cd ~/pcfb/examples/  
lang@melanogaster:~$ cat *.seq | grep ">"
```

```
>Fe_MM1_01A01  
>Fe_MM1_01A02  
>Fe_MM1_01A03  
>Fe_MM1_01A04  
>Fe_MM1_01A05  
>Fe_MM1_01A06  
>Fe_MM1_01A07  
>Fe_MM1_01A08  
>Fe_MM1_01A09  
>Fe_MM1_01A10
```

```
lang@melanogaster:~$ grep ">" *.seq
```

```
FEC00001_1.seq:>Fe_MM1_01A01  
FEC00002_1.seq:>Fe_MM1_01A02  
FEC00003_1.seq:>Fe_MM1_01A03  
FEC00004_1.seq:>Fe_MM1_01A04  
FEC00005_1.seq:>Fe_MM1_01A05  
FEC00005_2.seq:>Fe_MM1_01A06  
FEC00006_1.seq:>Fe_MM1_01A07  
FEC00007_1.seq:>Fe_MM1_01A08  
FEC00007_2.seq:>Fe_MM1_01A09  
FEC00007_3.seq:>Fe_MM1_01A10
```

Default behaviour of grep : indicates file name

Searching across multiple files with 'grep'

```
lang@melanogaster:~$ grep -l "GAATTC" *.seq
```

```
FEC00002_1.seq  
FEC00004_1.seq  
FEC00005_1.seq  
FEC00005_2.seq  
FEC00006_1.seq
```

Searches the sequences that contain an EcoRI site

Grep specifications

- c show only a count of the results in the file
- v invert search and show only lines that do not match
- i match without regard to case
- E use regular expression syntax
- l list only files that contain matches
- n show line number of match
- h hide filename in output

Retrieving Web content using 'curl'

```
lang@melanogaster:~$ curl "www.rcsb.org/pdb/files/iema.pdb"
```

```
lang@melanogaster:~$ curl "www.rcsb.org/pdb/files/iema.pdb" > iema.pdb
```

or

```
lang@melanogaster:~$ curl "www.rcsb.org/pdb/files/iema.pdb" -o iema.pdb
```

```
lang@melanogaster:~$ curl "www.wunderground.com/history/airport/MIA/1979/[01-30]/01/DailyHistory.html?&format=1" >> miamiweather.txt
```

Downloads 30 files, [days put in brackets], adds them into: miamiweather.txt

```
lang@melanogaster:~$ curl "www.wunderground.com/history/airport/MIA/1979/[01-12]/01/DailyHistory.html?&format=1" -o Miami_1979_#1.txt
```

Saves 12 files, separately (-o) with a consecutive integer for each file (#1)

Retrieving Web content using 'curl'

Downloads 30 files, [days put in brackets], adds them into: miamiweather.txt

```
lang@melanogaster:~$ curl "www.wunderground.com/history/airport/MIA/1979/[01-12]/01/DailyHistory.html?&format=1" -o Miami_1979_#1.txt
```

Saves 12 files, separately (-o) with a consecutive integer for each file (#1)

```
lang@melanogaster:~$ ls Mi*  
Lists the downloaded files
```

Particular, rather than sequential downloads :

```
lang@melanogaster:~$ curl "www.rcsb.org/pdb/files/{iema,igfl,1g7k,1xmz}.pdb"
```

Will download just with the query put in brackets {...}