

BIOL358 - Computer Skills for Biotechnology

Introduction to UNIX, part 2



Computer Skills for Biotechnology · <http://www.bioinformaticscourses.com/BIOL358/> · Helge Weissig, 2005 (1)

Files and Directories

- Relative and absolute paths
 - pdb_seqres.txt
 - ./pdb_data/pdb_seqres.txt
 - ../joe/files/
- Current working directory
 - Display with pwd
- Current location “.”
- Parent directory “..”
- Home directory “~”
- Hidden files
 - .ssh/ .cshrc .netscape/
- Wildcards & Command completion
 - ls pdb*
 - ls pdb_s<tab>



Computer Skills for Biotechnology · <http://www.bioinformaticscourses.com/BIOL358/> · Helge Weissig, 2005 (2)

Managing input and output

- Input/output “streams”: STDIN, STDOUT, STDERR
- Redirection operators
 - “>” send output to file
 - “<” use file as input
 - “>>” append output to file
 - “>!” force creation of file
 - “>&” send STDOUT and STDERR to file
- Examples:
 - cat file1 >! file2
 - echo “hello world” >> file2
 - (cat file1 > file2) >& stderr.txt



Computer Skills for Biotechnology · <http://www.bioinformaticscourses.com/BIOL358/> · Helge Weissig, 2005 (3)

Creating Pipes

- Output of process n becomes input of process n+1
- Pipe operator: “|”
- Examples:
 - `cat paper.txt | grep measles | wc`
 - `grep '^>' pdb_segres.txt | wc`



Computer Skills for Biotechnology · <http://www.bioinformaticscourses.com/BIOL358/> · Helge Weissig, 2005 (4)

Learning about Commands

<code>man</code>	"Man" (manual) pages for a command	<code>man man</code>
<code>apropos</code>	Keyword search of man pages	<code>apropos usage</code>
<code>info</code>	Read documentation in info format	<code>info info</code>



Computer Skills for Biotechnology · <http://www.bioinformaticscourses.com/BIOL358/> · Helge Weissig, 2005 (5)

Moving around in UNIX

<code>cd</code>	Change directory	<code>cd ../../</code>
<code>pwd</code>	Print current/working directory	<code>pwd</code>
<code>ssh</code>	Connect to a remote computer using secure shell	<code>ssh user1@login.bioinformaticscourses.com</code>
<code>ftp</code>	Connect to a remote computer using the file transfer protocol	<code>ftp ftp.rcsb.org ...</code>



Computer Skills for Biotechnology · <http://www.bioinformaticscourses.com/BIOL358/> · Helge Weissig, 2005 (6)

Downloading with FTP

- File Transfer Protocol
- Requires ftp server (anonymous or authenticated)
- Basic steps:
 - Connect
 - Change to target directory
 - Get/put the file(s)
- Limited set of commands but similar to UNIX
- Modes
 - ASCII or binary
 - Active or passive
 - Prompt or no prompt (multi transfers)
- Example: "ftp ftp.rcsb.org"



Computer Skills for Biotechnology · <http://www.bioinformaticscourses.com/BIOL358/> · Helge Weissig, 2005 (7)

Working with Files

ls	List files	ls -l -
cp	Copy a file to another file	cp oldfile newfile
mv	Rename and/or move a file	mv oldfile ..
rm	Remove a file	rm file1
mkdir	Make a directory	mkdir -p ./test/temp
rmdir	Remove a directory	rmdir -f test
chmod	Change permissions	chmod +x parser.pl
find	find and manipulate files recursively	find / -name aliases
touch	change data stamp of a file	touch test



Computer Skills for Biotechnology · <http://www.bioinformaticscourses.com/BIOL358/> · Helge Weissig, 2005 (8)

Working with File Contents

cat	Concatenate files and send them to the screen	cat file1 file2
more	Display file contents one screen at a time	more parser.pl
less	Same as more, but with more options	less parser.pl
head	Display first few lines of a file	head -15 /etc/passwd
last	Display last few lines of a file	last /var/log/messages
wc	count all characters, words and lines	wc file1 file2 file3
sort	sort file alphabetically or numerically	sort -n file1
grep	search for an expression in a file	grep measles paper.txt



Computer Skills for Biotechnology · <http://www.bioinformaticscourses.com/BIOL358/> · Helge Weissig, 2005 (9)

Working with File Contents [2]

<code>split</code>	Break up a large file into smaller ones	<code>split -l 100 pdb_seqres.txt pdbseqs.</code>
<code>csplit</code>	Similar to split, using either the number of lines or a regex	<code>csplit -f pdbseq. -n 6 pdb_seqres.txt '/^/' '{*}'</code>
<code>cut</code>	Output selected parts of a file	<code>cut -c 1,3 pdbseq.007899</code>
<code>paste</code>	Merge file contents by column	<code>paste protein*.pka.data > all.pka.data</code>
<code>join</code>	Merge file contents based on a specified join field	<code>sort mustelidae.prey join mustelidae.color - > outfile</code>
<code>diff</code>	Display the differences between files	<code>diff enolase1.list enolase2.list</code>
<code>patch</code>	Apply a "patch" to a file	<code>patch < thePatch1.txt</code>



Computer Skills for Biotechnology · <http://www.bioinformaticscourses.com/BIOL358/> · Helge Weissig, 2005 (10)

Compression and Archival

<code>tar</code>	manipulate "tape" archives	<code>tar xf archive.tar</code>
<code>[un]compress</code>	(un)compress files in different compression formats	<code>compress file1</code>
<code>[un]zip</code>		<code>unzip bigfile.zip</code>
<code>g[un]zip</code>		<code>gunzip archive.tar.gz</code>



Computer Skills for Biotechnology · <http://www.bioinformaticscourses.com/BIOL358/> · Helge Weissig, 2005 (11)

Managing Processes

<code>Ctrl-D</code>	end of file/input
<code>Ctrl-C</code>	interrupt foreground process
<code>Ctrl-Z</code>	suspend process
<code><command> &</code>	start process and return to shell immediately
<code>bg</code>	send suspended process into background
<code>fg</code>	bring background process to foreground
<code>jobs</code>	display processes in background
<code>kill</code>	kill a process
<code>ps</code>	get process status
<code>w</code>	display average load
<code>nice [command]</code>	change execution priority of a process
<code>nohup [command]</code>	do not terminate process when terminal is closed
<code>notify [jobId]</code>	report immediately when a background job finishes

Exercise:

- Use the `sleep` command to practise these



Computer Skills for Biotechnology · <http://www.bioinformaticscourses.com/BIOL358/> · Helge Weissig, 2005 (12)

The Execution Environment

- Environmental variables
 - PATH executable search path
 - HOME user home directory
 - SHELL current shell
 - TERM terminal type
- Manipulated with `setenv/printenv` commands



Computer Skills for Biotechnology · <http://www.bioinformaticscourses.com/BIOL358/> · Helge Weissig, 2005 (13)

Customizing the Shell

- Aliases, setting and unsetting
- Default options for `less`
- `.login` and `.cshrc` files
- `limits`
- `history` & short cuts
 - `!!` : repeat last command
 - `!*` : options & arguments of last command
 - `!$` : last argument of last command
 - `!-2` : penultimate command



Computer Skills for Biotechnology · <http://www.bioinformaticscourses.com/BIOL358/> · Helge Weissig, 2005 (14)

UNIX Exercise Questions

- Work through last weeks exercises completely
- How many subdirectories are there in `/sw/share` on "jaws"?
- Decompress and unarchive the file `/Users/course_data.tar.gz` in your home directory. Find out two ways in which this can be done in a single command line? What is the option for tar to display the contents of an archive?
- Look at the files included in this archive. What do they contain?
- Extract the sequence record from the PDB file and save it in a separate file.
- How many ATOMS are given listed in the PDB file?
- How can you generate the files `m3k1.[1-3]` yourself?
- Combine the contents of `M3K1_*.dta` into a single file. Do this with and without the use of a temporary file.
- Generate a tar file with all the contents of your work so far. Compress this tar file with the compress command.



Computer Skills for Biotechnology · <http://www.bioinformaticscourses.com/BIOL358/> · Helge Weissig, 2005 (15)
