

BIOL358 - Computer Skills for Biotechnology

HTML, WWW server &
first introduction to Unix



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

HTML

First example:

```
<html>
  <head>
    <title>My First Page</title>
  </head>
  <body>
    <h1>Hello World</h1>
  </body>
</html>
```



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

HTML Tips, Tricks & Tutorials

- Like a page? Look at the source!
- Try Netscape's "Edit Page"
 - Other alternatives: NotePad, TextEdit
- Don't forget to upload images
- Look for some online tutorials
 - Try out <http://www.w3schools.com/html/>
 - Course goal: pass the test 20/20!



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

WWW Servers

- ❑ www.apache.org
- ❑ Open Source & well maintained
- ❑ Easy installation and configuration
- ❑ Robust and secure
- ❑ Supports majority of internet sites
- ❑ Highly extensible
- ❑ Integrates well with Perl and Java



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

First introduction to Unix



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

Brief History

- ❑ Early 60's: Multics (Multiplexed Information and Computing Service), abandoned by Bell Labs in 1969
- ❑ Work on UNIX* started by Ken Thompson, Dennis Ritchie, Doug McIlroy & J.F. Ossanna
- ❑ "Space Travel" on a PDB-7 (\$75 for CPU time)
- ❑ Early on: processes, user-level utilities (copy, print, delete & edit files), shell
- ❑ 1970: Group proposes buying a PDP-11 (\$65,000)



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

Brief History [cont.]



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

Brief History [cont.]

- ❑ First customer: Bell Labs Patent Department
- ❑ Programming languages used:
 - ❑ B developed for PDP-7
 - ❑ NB on PDP-11 (B with types)
 - ❑ Adding a compiler created C
- ❑ 1972-3: re-write of UNIX in C
- ❑ 1973: addition of pipes
- ❑ 1974-: the UNIX manual (man pages) including bug sections!



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

Brief History [cont.]

- ❑ 1976-77: Ken Thompson's sabbatical at UCB (version 6)
- ❑ Continued work by UCB students and professors: Berkeley Software Distribution (BSD) Version 4.2
- ❑ Computer Systems Research Group (CSRG) at UCB develops TCP/IP
- ❑ Digital produces VAX running UNIX, affordable for academic departments



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

Brief History [cont.]

- ❑ Unix wars:
 - ❑ Other vendors: Sun's Solaris, SGI's IRIX, Hewlett-Packard's HP-UX, IBM's AIX, Compaq's Tru64 UNIX
 - ❑ Open Systems Foundation (OSF) v. UNIX International (AT&T & Sun)
- ❑ Linux, FreeBSD & NetBSD available for free
- ❑ UNIX now OS of most internet servers, businesses and universities
- ❑ Most commercial software is written in C or C++ (more recently Java)



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

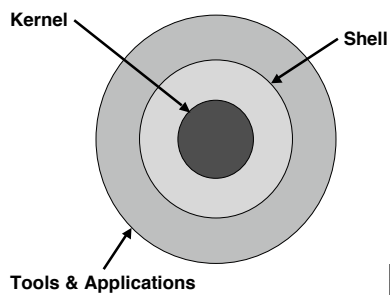
Basic Principles

- ❑ Modular structure
- ❑ The UNIX shell
- ❑ Terminals and X-Windows
- ❑ Hierarchical file systems
- ❑ Basic set of tools and commands
- ❑ Input/ Output redirection & Pipes
- ❑ Files & Processes
- ❑ Time sharing
- ❑ Multi-user environment



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

Modular Structure



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

Modular Structure [cont.]

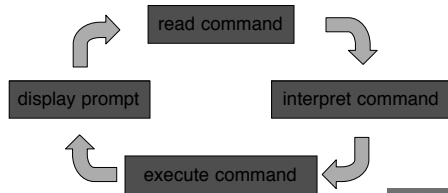
- ❑ kernel
 - ❑ First program to run when machine boots
 - ❑ Runs all other programs
 - ❑ "operating system"
- ❑ shell
 - ❑ First program to run when user logs in
 - ❑ Reads and executes user commands
 - ❑ "command line interpreter"
- ❑ tools & applications
 - ❑ Everything else



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

The UNIX shell

- ❑ Interface between user and kernel
- ❑ Interpreter loop:



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

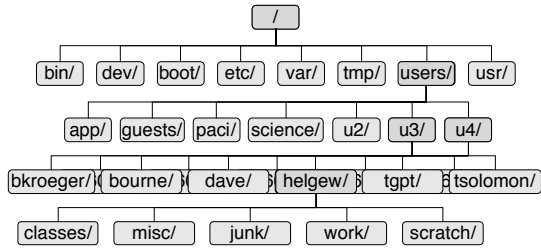
Terminals and X-Windows

- ❑ Console
 - ❑ Connected directly to machine
 - ❑ Special type of terminal recognized at start-up
 - ❑ Typically only accessible to system administrators
- ❑ Dumb terminal
 - ❑ Can receive ASCII input and output
- ❑ Smart terminals
 - ❑ Can receive graphical output
 - ❑ Can interact through "widgets" with processes
- ❑ X-Windows System
 - ❑ Graphical user interface (GUI) for programs



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

Hierarchical File System



Computer Skills for Biotechnology · <http://www.bioinformaticscourses.com/BIOL358/> · Helge Weissig, 2005 (16)

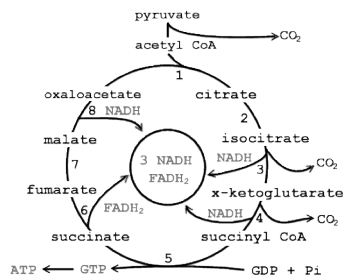
Input/Output Redirection & Pipes

- Concepts:
 - programs that do one thing and do it well
 - programs that work together
 - programs that handle text streams
- Output of step n becomes input of step n+1
- File content can become input and output for arbitrary programs
- Example:


```
cat < infile.txt | sort > outfile.txt
```

Computer Skills for Biotechnology · <http://www.bioinformaticscourses.com/BIOL358/> · Helge Weissig, 2005 (17)

Pipes, A Familiar Concept



Computer Skills for Biotechnology · <http://www.bioinformaticscourses.com/BIOL358/> · Helge Weissig, 2005 (18)

Files & Processes

- ❑ Everything is either a file or a process
- ❑ Examples of files:
 - ❑ Documents containing plain text data, reports, web pages, source code etc. ("ASCII" or "text only" files)
 - ❑ Instructions comprehensible only to the machine or programs ("binary files")
 - ❑ A directory containing information about its contents (usually a mixture of files and directory files)



Computer Skills for Biotechnology · <http://www.bioinformaticscourses.com/BIOL358/> · Helge Weissig, 2005 (19)

Files & Processes [cont.]

- ❑ A process is a program in the state of execution
- ❑ Processes are uniquely identified by a process id
- ❑ Examples of processes:
 - ❑ kernel process
 - ❑ Daemons: e.g. httpd, ftpd, sendmail, sshd
 - ❑ Other programs while running: e.g. tcsh, ls, man, more, pine



Computer Skills for Biotechnology · <http://www.bioinformaticscourses.com/BIOL358/> · Helge Weissig, 2005 (20)

Time Sharing

- ❑ UNIX machines (as most modern operating systems) run more than one process at a time
- ❑ Upper limit: "process space" or (memory + swap space)
- ❑ Process states: "sleep" or "on CPU"
- ❑ Processes can have threads
- ❑ The kernel assigns priorities based on nicety levels ("nice")



Computer Skills for Biotechnology · <http://www.bioinformaticscourses.com/BIOL358/> · Helge Weissig, 2005 (21)

UNIX commands

- ❑ Basic syntax
- ❑ Working with files and directories
- ❑ Managing input and output
- ❑ Creating pipes
- ❑ Managing processes
- ❑ The execution environment
- ❑ The dreaded vi editor
- ❑ Customizing the shell



Computer Skills for Biotechnology · <http://www.bioinformaticscourses.com/BIOL358/> · Helge Weissig, 2005 (25)

Basic Syntax

- ❑ `command [-options] [arguments]`
- ❑ Commands:
 - ❑ Built-in or programs
 - ❑ Will usually have default options
 - ❑ When in doubt use `man command`
- ❑ Options:
 - ❑ Depend on command
 - ❑ Usually given with single (old style) or double hyphens
 - ❑ Can usually be combined
- ❑ Arguments:
 - ❑ Usually one or more files
- ❑ Special characters: `/ < > ! $ % ^ & * ! { } ~ ;`
- ❑ Aborting: Ctrl-C (^C)



Computer Skills for Biotechnology · <http://www.bioinformaticscourses.com/BIOL358/> · Helge Weissig, 2005 (26)

Files and Directories

- Concepts:
- ❑ Relative and absolute paths
 - ❑ Current working directory
 - ❑ Current location `“.”`
 - ❑ Parent directory `“..”`
 - ❑ Home directory `“~”`
 - ❑ Hidden files
 - ❑ Special files (links, sockets, named pipes)



Computer Skills for Biotechnology · <http://www.bioinformaticscourses.com/BIOL358/> · Helge Weissig, 2005 (27)

File & Directory Commands

ls	List files	ls -l -
chmod	Change permissions	chmod +x parser.pl
cd	Change directory	cd ../../
pwd	Print current/working directory	pwd
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 access_log
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



Computer Skills for Biotechnology · <http://www.bioinformaticscourses.com/BIOL358/> · Helge Weissig, 2005 (28)

File & Directory Commands [cont.]

find	find and manipulate files recursively	find / -name aliases
touch	change data stamp of a file	touch test
grep	search for an expression in a file	grep meales paper.txt
wc	count all characters, words and lines	wc file1 file2 file3
sort	sort file alphabetically or numerically	sort -n file1
[un]compress	(un)compress files in different compression formats	compress file1
[un]zip		unzip bigfile.zip
g[un]zip		gunzip archive.tar.gz
tar	manipulate "tape" archives	tar xf archive.tar



Computer Skills for Biotechnology · <http://www.bioinformaticscourses.com/BIOL358/> · Helge Weissig, 2005 (29)

Exercises

- Login to account, create a directory pdb_data
- Download and uncompress
ftp://ftp.rcsb.org/pub/pdb/derived_data/pdb_seqres.txt.Z
- Move the sequence file into the pdb_data directory
- Look at the file with more or less
- Find occurrences of 'HHB' in the file
- Play with copying, deleting and renaming of your file(s)



Computer Skills for Biotechnology · <http://www.bioinformaticscourses.com/BIOL358/> · Helge Weissig, 2005 (30)
