

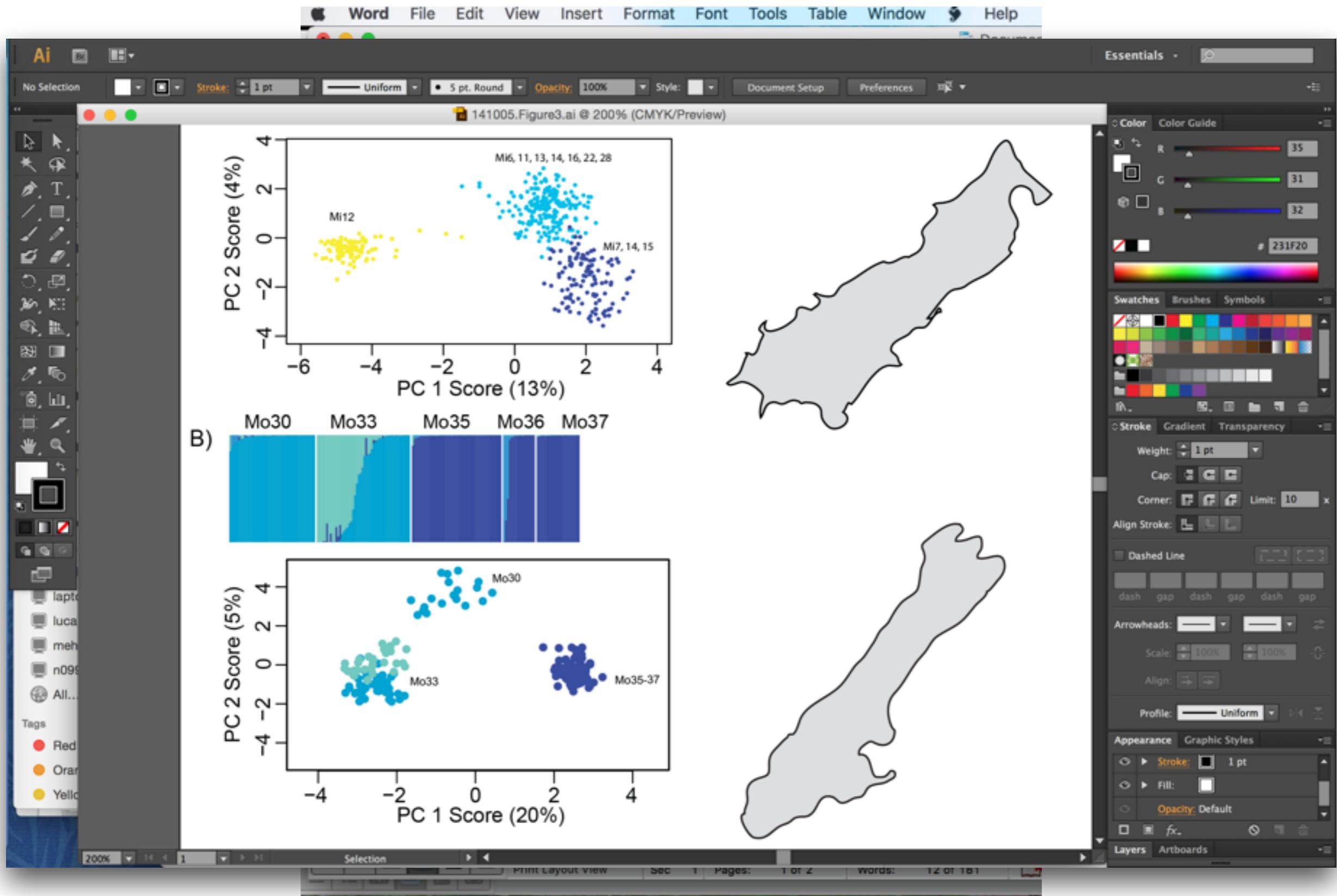
**digital™**

**LIVE FREE OR DIE**

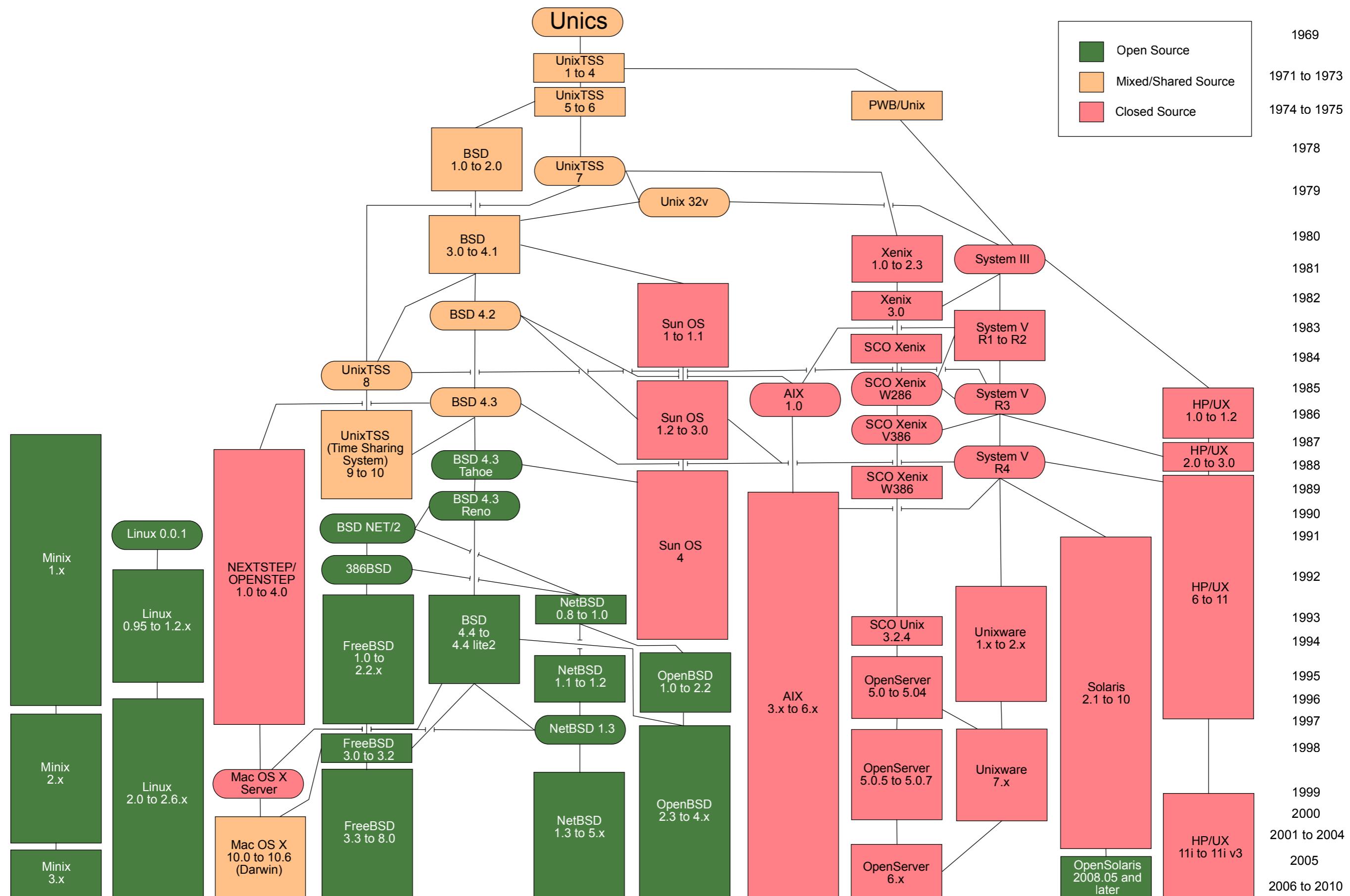
**UNIX**

Registered TM of UNIX System Laboratories,  
A wholly owned subsidiary of Novell, Inc.

948-0034



# Unix History



# What computers can run Unix?



Apple OS X Macs



Google's Android phones

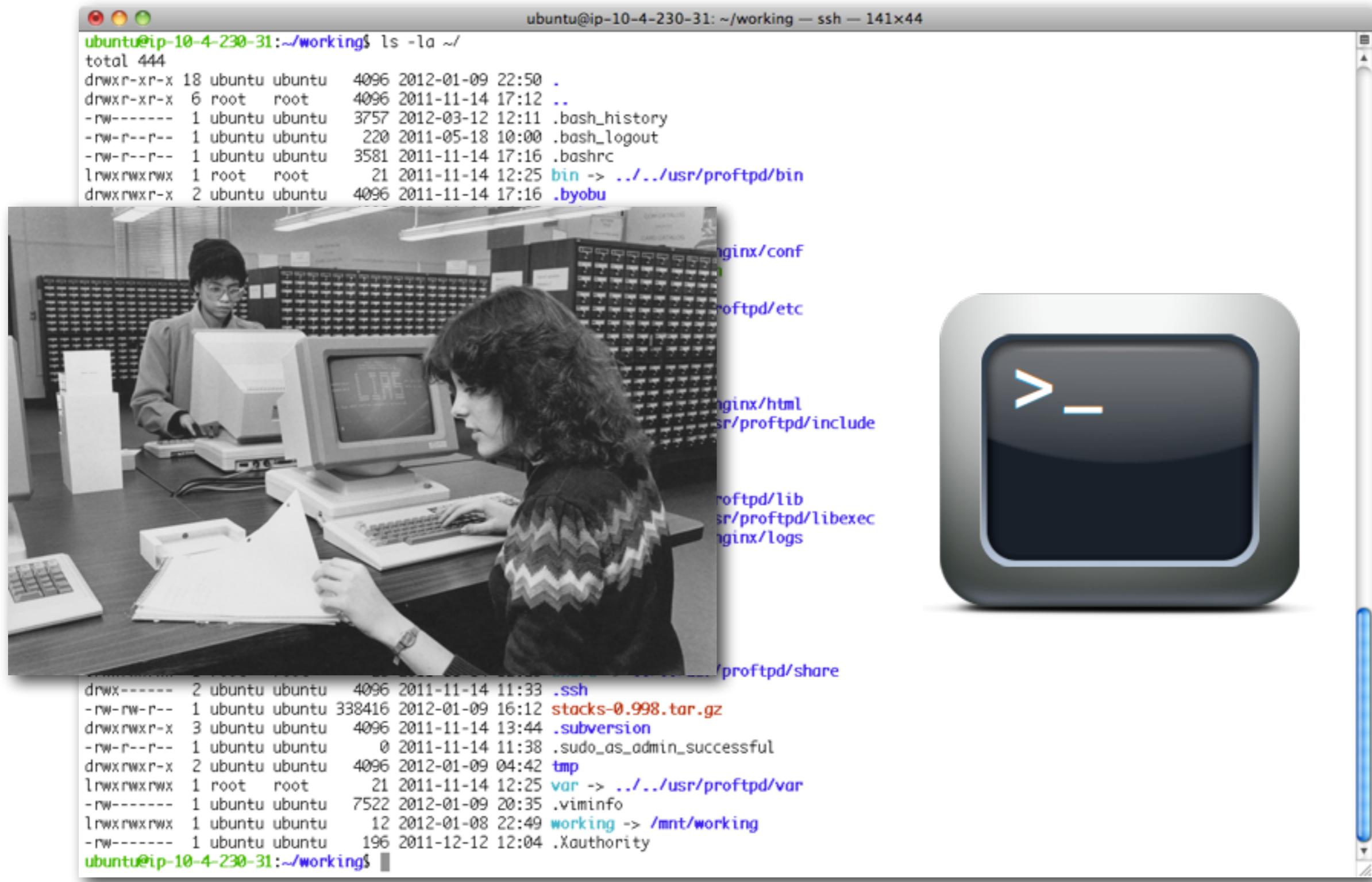


Wireless internet routers



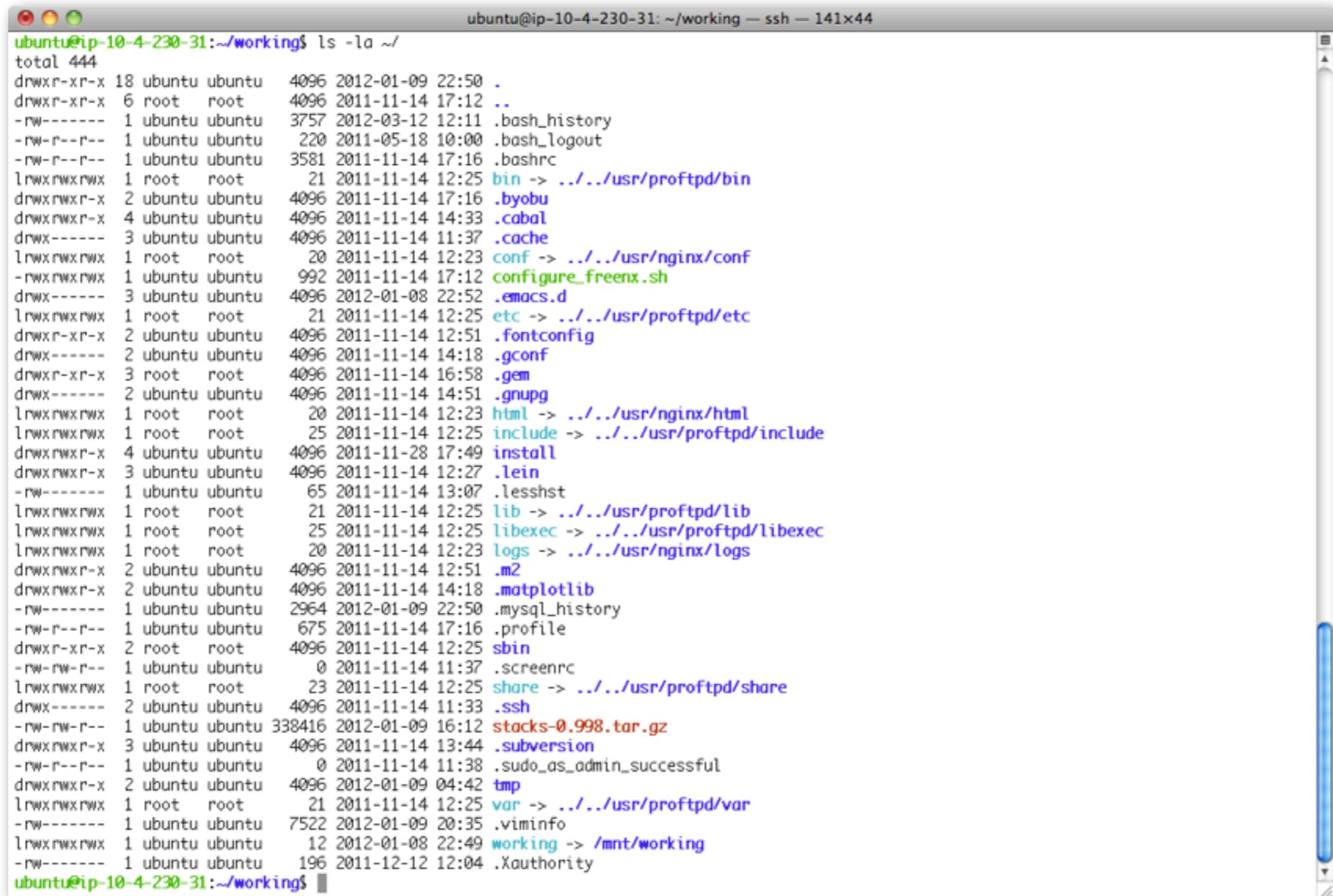
Airplane  
entertainment  
systems

# The Terminal Window



the shell, the prompt, the command line

# The Terminal Window



A screenshot of a terminal window titled "ubuntu@ip-10-4-230-31: ~/working — ssh — 141x44". The window displays the output of the command "ls -la ~/" which lists numerous files and directories in the user's home directory. The output includes file names, permissions, ownership, sizes, and modification dates. A vertical scroll bar is visible on the right side of the terminal window.

```
ubuntu@ip-10-4-230-31:~/working$ ls -la ~/
total 444
drwxr-xr-x 18 ubuntu ubuntu 4096 2012-01-09 22:50 .
drwxr-xr-x  6 root   root   4096 2011-11-14 17:12 ..
-rw-----  1 ubuntu ubuntu 3757 2012-03-12 12:11 .bash_history
-rw-r--r--  1 ubuntu ubuntu  220 2011-05-18 10:00 .bash_logout
-rw-r--r--  1 ubuntu ubuntu 3581 2011-11-14 17:16 .bashrc
lrwxrwxrwx  1 root   root   21 2011-11-14 12:25 bin -> ../../usr/proftpd/bin
drwxrwxr-x  2 ubuntu ubuntu 4096 2011-11-14 17:16 .byobu
drwxrwxr-x  4 ubuntu ubuntu 4096 2011-11-14 14:33 .cabal
drwx-----  3 ubuntu ubuntu 4096 2011-11-14 11:37 .cache
lrwxrwxrwx  1 root   root   20 2011-11-14 12:23 conf -> ../../usr/nginx/conf
-rwxrwxrwx  1 ubuntu ubuntu 992 2011-11-14 17:12 configure_freenx.sh
drwx-----  3 ubuntu ubuntu 4096 2012-01-08 22:52 .emacs.d
lrwxrwxrwx  1 root   root   21 2011-11-14 12:25 etc -> ../../usr/proftpd/etc
drwxr-xr-x  2 ubuntu ubuntu 4096 2011-11-14 12:51 .fontconfig
drwx-----  2 ubuntu ubuntu 4096 2011-11-14 14:18 .gconf
drwxr-xr-x  3 root   root   4096 2011-11-14 16:58 .gem
drwx-----  2 ubuntu ubuntu 4096 2011-11-14 14:51 .gnupg
lrwxrwxrwx  1 root   root   20 2011-11-14 12:23 html -> ../../usr/nginx/html
lrwxrwxrwx  1 root   root   25 2011-11-14 12:25 include -> ../../usr/proftpd/include
drwxrwxr-x  4 ubuntu ubuntu 4096 2011-11-28 17:49 install
drwxrwxr-x  3 ubuntu ubuntu 4096 2011-11-14 12:27 .lein
-rw-----  1 ubuntu ubuntu 65 2011-11-14 13:07 .lessht
lrwxrwxrwx  1 root   root   21 2011-11-14 12:25 lib -> ../../usr/proftpd/lib
lrwxrwxrwx  1 root   root   25 2011-11-14 12:25 libexec -> ../../usr/proftpd/libexec
lrwxrwxrwx  1 root   root   20 2011-11-14 12:23 logs -> ../../usr/nginx/logs
drwxrwxr-x  2 ubuntu ubuntu 4096 2011-11-14 12:51 .m2
drwxrwxr-x  2 ubuntu ubuntu 4096 2011-11-14 14:18 .matplotlib
-rw-----  1 ubuntu ubuntu 2964 2012-01-09 22:50 .mysql_history
-rw-r--r--  1 ubuntu ubuntu 675 2011-11-14 17:16 .profile
drwxr-xr-x  2 root   root   4096 2011-11-14 12:25 sbin
-rw-rw-r--  1 ubuntu ubuntu     0 2011-11-14 11:37 .screenrc
lrwxrwxrwx  1 root   root   23 2011-11-14 12:25 share -> ../../usr/proftpd/share
drwx-----  2 ubuntu ubuntu 4096 2011-11-14 11:33 .ssh
-rw-rw-r--  1 ubuntu ubuntu 338416 2012-01-09 16:12 stacks-0.998.tar.gz
drwxrwxr-x  3 ubuntu ubuntu 4096 2011-11-14 13:44 .subversion
-rw-r--r--  1 ubuntu ubuntu     0 2011-11-14 11:38 .sudo_as_admin_successful
drwxrwxr-x  2 ubuntu ubuntu 4096 2012-01-09 04:42 tmp
lrwxrwxrwx  1 root   root   21 2011-11-14 12:25 var -> ../../usr/proftpd/var
-rw-----  1 ubuntu ubuntu 7522 2012-01-09 20:35 .viminfo
lrwxrwxrwx  1 ubuntu ubuntu  12 2012-01-08 22:49 working -> /mnt/working
-rw-----  1 ubuntu ubuntu 196 2011-12-12 12:04 .Xauthority
ubuntu@ip-10-4-230-31:~/working$
```

Make it comfortable to work in:

- Resize the window
- Change your font size
- Open multiple terminal windows

Grab File Edit Capture Window Help

catchen@Whitaker.local:/Users/catchen/hg/stacks — bash — 181x39

Processes: 868  
Usage of /: 85% of 917G  
Memory usage: 78%  
Users: 2  
IP address for eth0 is 128.223.42.24

```
[catchen@genome]:~% ls
oarl_test/          clone/
abyss-1.3.4/        cl_test/
abyss-1.3.4.tar.gz coord_test/
adapter/           corr/
aj/                count-genome-len.pl*
allpathslg-41370.tar.gz courses/
allpathslg-44683.tar.gz danio_sexdet_radtags-sep_29_2011.tsv.gz
allpathslg-44837.tar.gz dave/
ALLPATHS-LG.test_genome/ denovo_map.log
ash.py*             Desktop/
ash.py-              Documents/
ash.sam             Downloads/
batch_1.fst_9-12.tsv emily/
batch_1.fst_9-12.tsv.new erica/
batch_1.fst_9-12.tsv.old fancc_radtags.genotype_corrections.tsv.gz
batch_1_ids.txt     fastphase_finallikelihoods
batch_1.sumstats_1.csv fastPHASE_Linux*
ben/               faststructure/
bentley/           gac_gen_broads1_e64.coords
bi610/              gac_map_test/
blair/              Gasterosteus_aculeatus_assembly.ogp
bpitest/            Gasterosteus_aculeatus.BROADS1.68.gtf
broad/              genepop/
build_tags.sh*      gmap-gsnap-2013-11-27.tar.gz
build_tags.sh-*    gnuplot-4.6.4/
catalog_test/       gnuplot-4.6.4.tar.gz
catchen_ec2.pem     groupI_contigs.tsv
[catchen@genome]:~% emacs /var/www/creskolab/stacks/index.php
[catchen@genome]:~% exit
logout
Connection to genome.uoregon.edu closed.
[catchen@Whitaker]:~/hg/stacks%
```

```
[catchen@Whitaker]:~/hg/stacks% hg pull
warning: bitbucket.org certificate with fingerprint 24:9c:45:8b:9c:aa:ba:55:4e:01:6d:58:ff:e4:28:7d:2a:14:ae:3b not verified (check hostfingerprints or web.cacerts config setting)
http authorization required
realm: Bitbucket.org HTTP
user: jcatchen
password:
warning: bitbucket.org certificate with fingerprint 24:9c:45:8b:9c:aa:ba:55:4e:01:6d:58:ff:e4:28:7d:2a:14:ae:3b not verified (check hostfingerprints or web.cacerts config setting)
pulling from https://jcatchen@bitbucket.org/jcatchen/stacks
warning: bitbucket.org certificate with fingerprint 24:9c:45:8b:9c:aa:ba:55:4e:01:6d:58:ff:e4:28:7d:2a:14:ae:3b not verified (check hostfingerprints or web.cacerts config setting)
warning: bitbucket.org certificate with fingerprint 24:9c:45:8b:9c:aa:ba:55:4e:01:6d:58:ff:e4:28:7d:2a:14:ae:3b not verified (check hostfingerprints or web.cacerts config setting)
searching for changes
warning: bitbucket.org certificate with fingerprint 24:9c:45:8b:9c:aa:ba:55:4e:01:6d:58:ff:e4:28:7d:2a:14:ae:3b not verified (check hostfingerprints or web.cacerts config setting)
adding changesets
adding manifests
adding file changes
added 1 changesets with 1 changes to 1 files
warning: bitbucket.org certificate with fingerprint 24:9c:45:8b:9c:aa:ba:55:4e:01:6d:58:ff:e4:28:7d:2a:14:ae:3b not verified (check hostfingerprints or web.cacerts config setting)
(run 'hg update' to get a working copy)
[catchen@Whitaker]:~/hg/stacks% hg update
1 files updated, 0 files merged, 0 files removed, 0 files unresolved
[catchen@Whitaker]:~/hg/stacks%
```

Macintosh HD  
499.42 GB, 332.84 GB free

- awk\_test  
2 items
- metapop\_commands.txt
- scan.stacks
- vStacks.app.zip  
285 KB
- scan\_stacks  
77 items
- 10.tar.gz ustacks\*
- 11.tar.gz vStacks/
- 12.tar.gz
- 13.tar.gz
- 10.tar.gz ustacks\*
- 11.tar.gz vStacks/
- 12.tar.gz
- 13.tar.gz



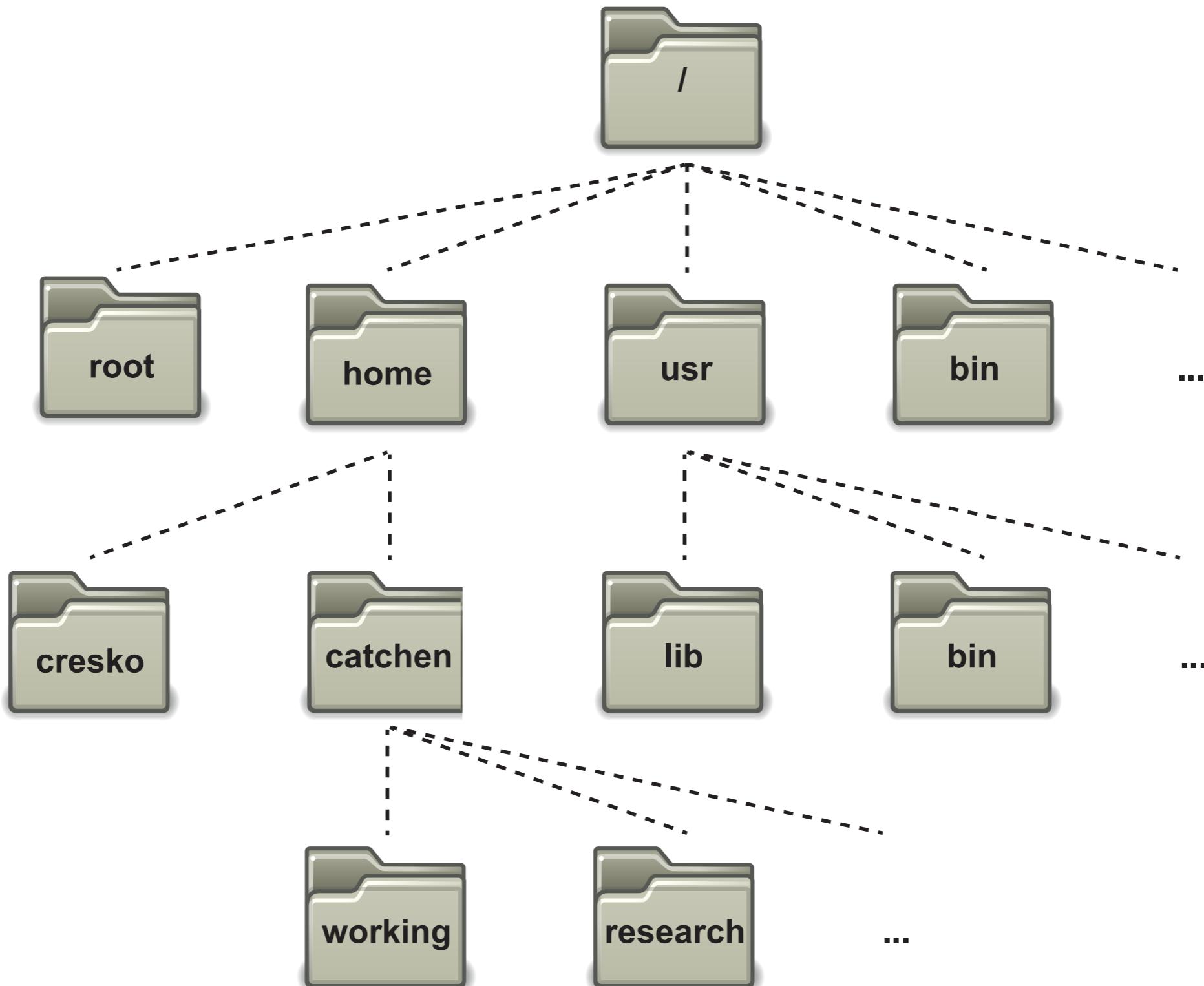
# Obtain a cheat sheet

google “unix commands”

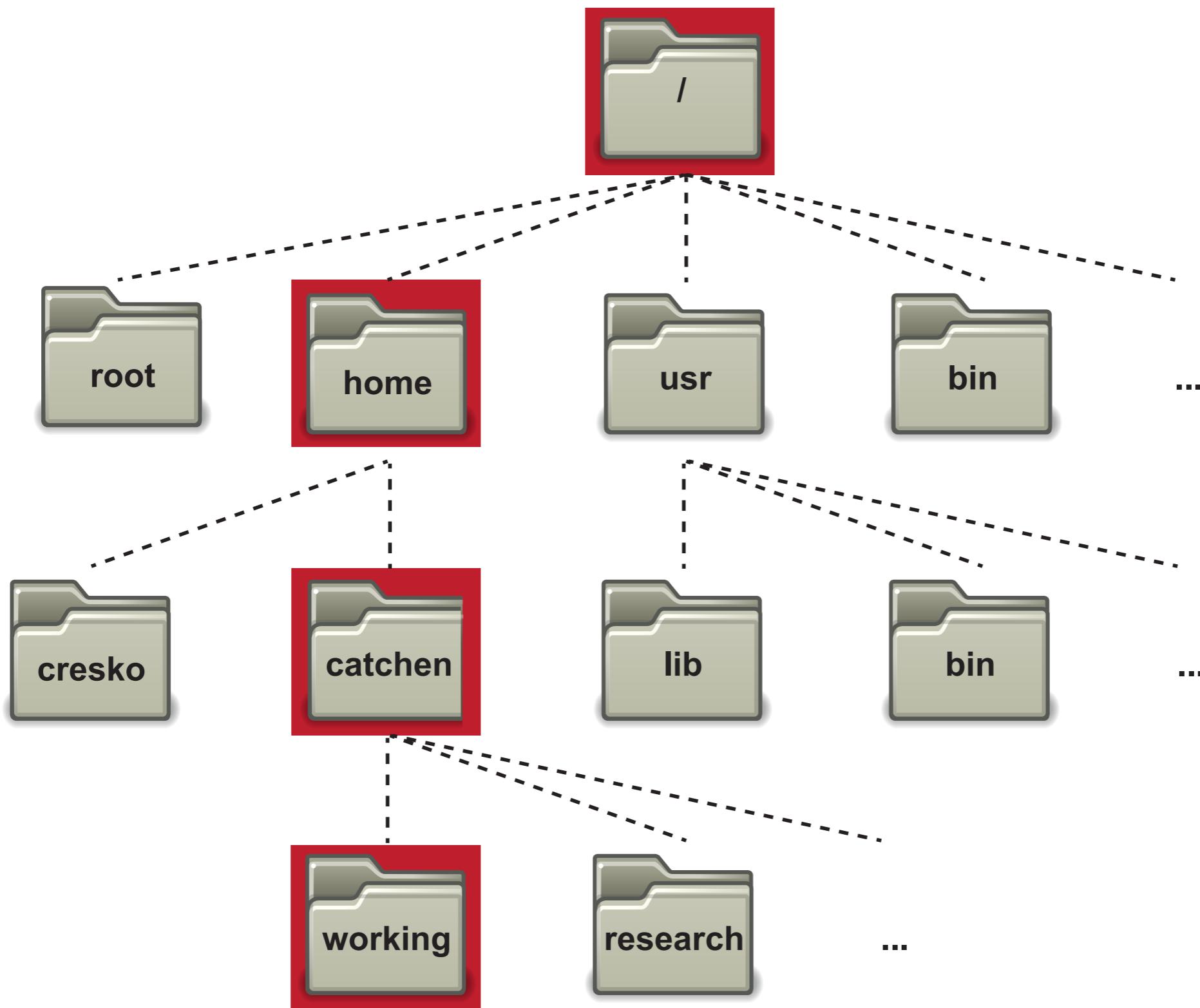
The screenshot shows a web browser window titled "Basic UNIX commands". The address bar displays the URL "mally.stanford.edu/~sr/computing/basic-unix.html". The main content area is titled "Basic UNIX commands". It contains a note about the commands being used on "turing" and differences between Solaris and SunOS. It also mentions keyboard navigation and case-sensitivity. Below this, a section titled "Files" lists various Unix commands:

- **ls** --- lists your files
  - ls -l --- lists your files in 'long format', which contains lots of useful information, e.g. the exact size of the file, who owns the file and who has the right to look at it, and when it was last modified.
  - ls -a --- lists all files, including the ones whose filenames begin in a dot, which you do not always want to see.  
There are many more options, for example to list files by size, by date, recursively etc.
- **more filename** --- shows the first part of a file, just as much as will fit on one screen. Just hit the space bar to see more or **q** to quit. You can use **/pattern** to search for a pattern.
- **emacs filename** --- is an editor that lets you create and edit a file. See the [emacs page](#).
- **mv filename1 filename2** --- moves a file (i.e. gives it a different name, or moves it into a different directory (see below)
- **cp filename1 filename2** --- copies a file
- **rm filename** --- removes a file. It is wise to use the option rm -i, which will ask you for confirmation before actually deleting anything.  
You can make this your default by making an [alias](#) in your .cshrc file.
- **diff filename1 filename2** --- compares files, and shows where they differ
- **wc filename** --- tells you how many lines, words, and characters there are in a file
- **chmod options filename** --- lets you change the read, write, and execute permissions on your files. The default is that only you can look at them and change them, but you may sometimes want to change these permissions. For example, **chmod o+r filename** will make the file readable for everyone, and **chmod o-r filename** will make it unreadable for others again. Note that for someone to be able to actually look at the file the directories it is in need to be at least executable. See [help protection](#) for more details.
- File Compression
  - **gzip filename** --- compresses files, so that they take up much less space. Usually text files compress to about half their original size, but it depends very much on the size of the file and the nature of the contents. There are other tools for this purpose, too (e.g. **compress**), but gzip usually gives the highest compression rate. Gzip produces files with the ending '.gz' appended to the original filename.
  - **gunzip filename** --- uncompresses files compressed by gzip.

# In UNIX everything is a file organized in a hierarchy



# Paths



**/home/catchen/working**

# Create a series of directories

```
ubuntu@ip-10-4-193-188:~$ mkdir shell
ubuntu@ip-10-4-193-188:~$ cd shell
ubuntu@ip-10-4-193-188:~/shell$ mkdir research
ubuntu@ip-10-4-193-188:~/shell$ ls
research
ubuntu@ip-10-4-193-188:~/shell$ cd research/
ubuntu@ip-10-4-193-188:~/shell/research$ mkdir seq
ubuntu@ip-10-4-193-188:~/shell/research$ ls
seq
ubuntu@ip-10-4-193-188:~/shell/research$ cd seq/
ubuntu@ip-10-4-193-188:~/shell/research/seq$ mkdir radtags
ubuntu@ip-10-4-193-188:~/shell/research/seq$ cd radtags/
ubuntu@ip-10-4-193-188:~/shell/research/seq/radtags$ ls
ubuntu@ip-10-4-193-188:~/shell/research/seq/radtags$ ls -la
total 8
drwxrwxr-x 2 ubuntu ubuntu 4096 2012-03-06 23:08 .
drwxrwxr-x 3 ubuntu ubuntu 4096 2012-03-06 23:08 ...
ubuntu@ip-10-4-193-188:~/shell/research/seq/radtags$ pwd
/home/ubuntu/shell/research/seq/radtags
ubuntu@ip-10-4-193-188:~/shell/research/seq/radtags$ █
```

% mkdir shell

% cd shell

% ls

# Paths, cont

This shell view of the nested directories shell, research, seq, and radtags.....

```
ubuntu@ip-10-4-193-188:~$ mkdir shell  
ubuntu@ip-10-4-193-188:~$ cd shell  
ubuntu@ip-10-4-193-188:~/shell$ mkdir research  
ubuntu@ip-10-4-193-188:~/shell$ ls  
research  
ubuntu@ip-10-4-193-188:~/shell$ cd research/  
ubuntu@ip-10-4-193-188:~/shell/research$ mkdir seq  
ubuntu@ip-10-4-193-188:~/shell/research$ ls  
seq  
ubuntu@ip-10-4-193-188:~/shell/research$ cd seq/  
ubuntu@ip-10-4-193-188:~/shell/research/seq$ mkdir radtags  
ubuntu@ip-10-4-193-188:~/shell/research/seq$ cd radtags/  
ubuntu@ip-10-4-193-188:~/shell/research/seq/radtags$ ls  
ubuntu@ip-10-4-193-188:~/shell/research/seq/radtags$ ls -la  
total 8  
drwxrwxr-x 2 ubuntu ubuntu 4096 2012-03-06 23:08 .  
drwxrwxr-x 3 ubuntu ubuntu 4096 2012-03-06 23:08 ..  
ubuntu@ip-10-4-193-188:~/shell/research/seq/radtags$ pwd  
/home/ubuntu/shell/research/seq/radtags  
ubuntu@ip-10-4-193-188:~/shell/research/seq/radtags$
```

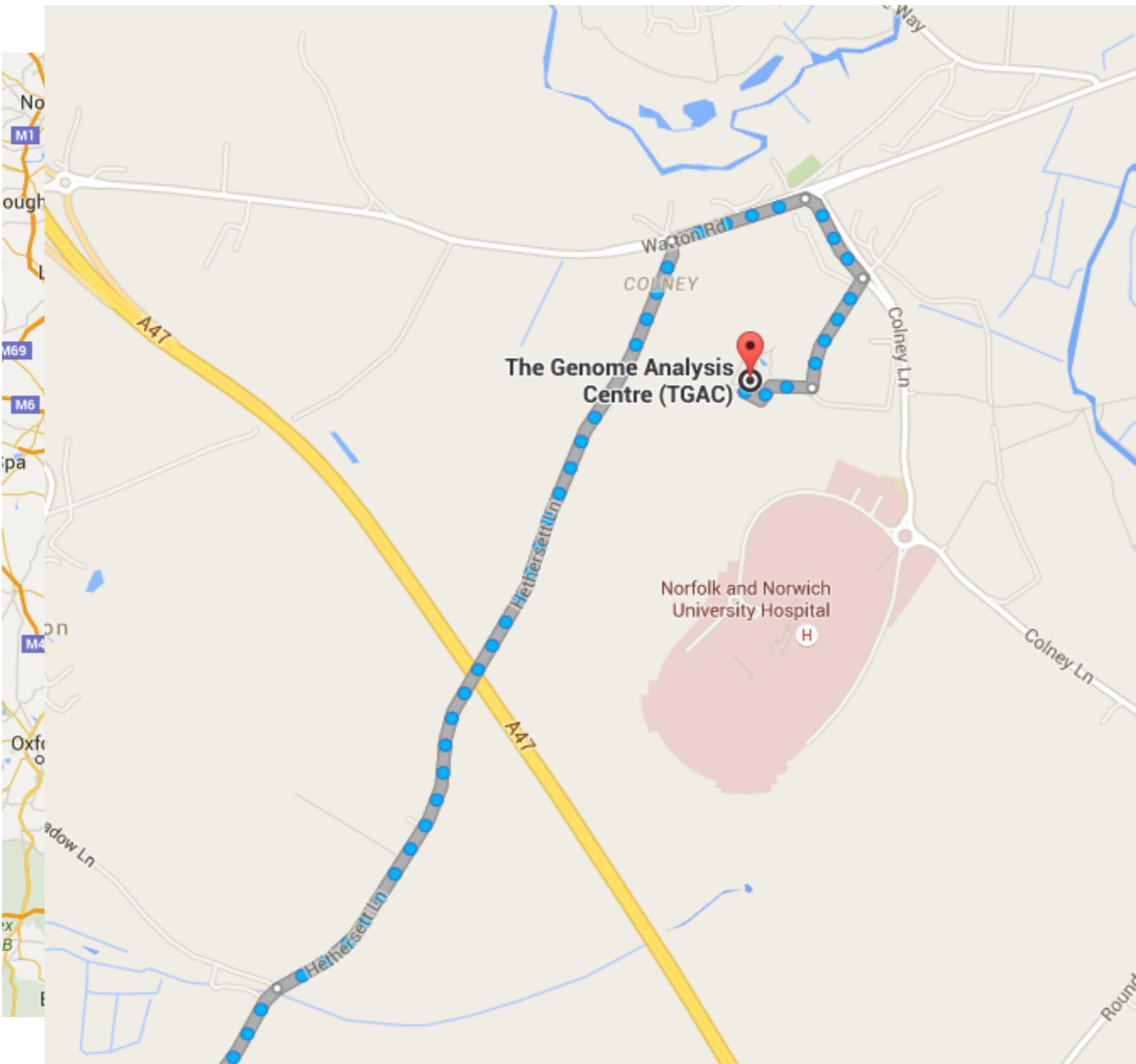
.... is equivalent to this GUI view of the same directories

Filename	Size	Modified
► bin	21 B	11/14/11 4:25 AM
► conf	20 B	11/14/11 4:23 AM
► configure_freenx.sh	992 B	11/14/11 9:12 AM
► etc	21 B	11/14/11 4:25 AM
► html	20 B	11/14/11 4:23 AM
► include	25 B	11/14/11 4:25 AM
► install	4.0 KB	11/28/11 9:49 AM
► lib	21 B	11/14/11 4:25 AM
► libexec	25 B	11/14/11 4:25 AM
► logs	20 B	11/14/11 4:23 AM
► sbin	4.0 KB	11/14/11 4:25 AM
► share	23 B	11/14/11 4:25 AM
▼ shell	4.0 KB	3/6/12 3:08 PM
▼ research	4.0 KB	3/6/12 3:08 PM
▼ seq	4.0 KB	3/6/12 3:08 PM
▼ radtags	4.0 KB	3/6/12 3:08 PM
stacks-0.998.tar.gz	330.4 KB	1/9/12 8:12 AM
tmp	4.0 KB	1/8/12 8:42 PM
var	21 B	11/14/11 4:25 AM
working	12 B	1/8/12 2:49 PM

And the radtags directory is uniquely identified by its path:  
/home/ubuntu/shell/research/seq/radtags

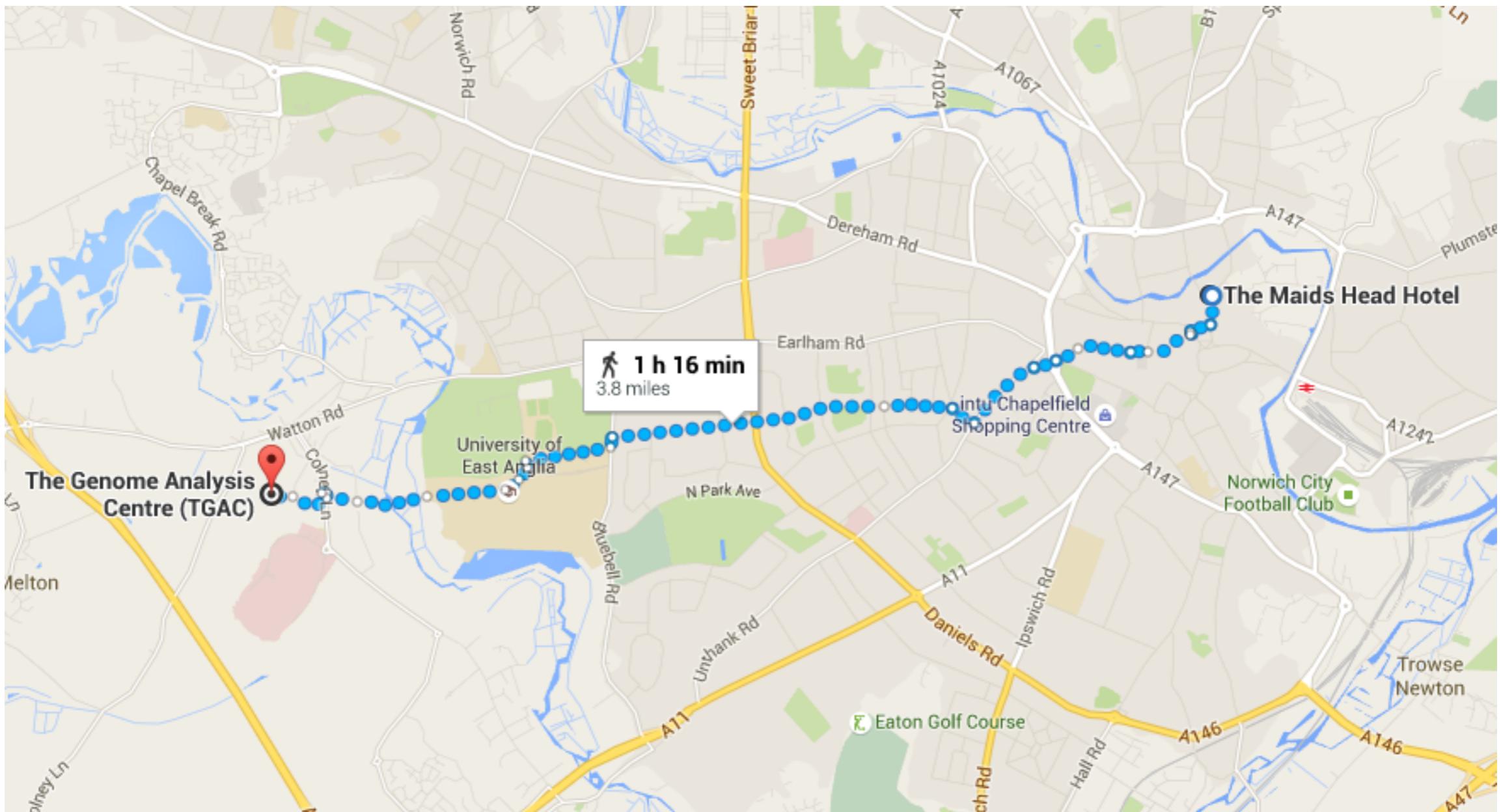
# Absolute and relative paths

*How do I get to TGAC in the UK?*

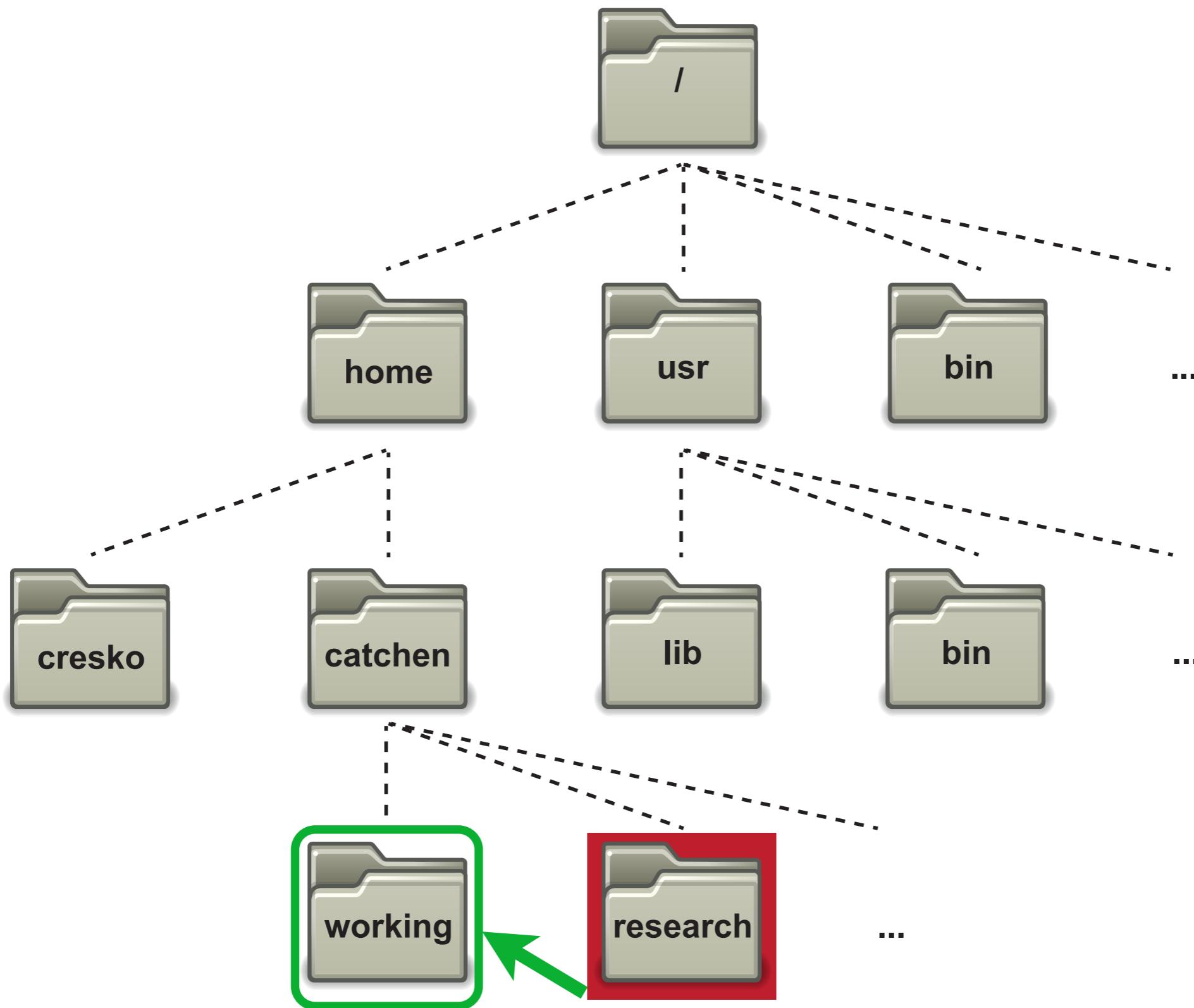


# Absolute and relative paths

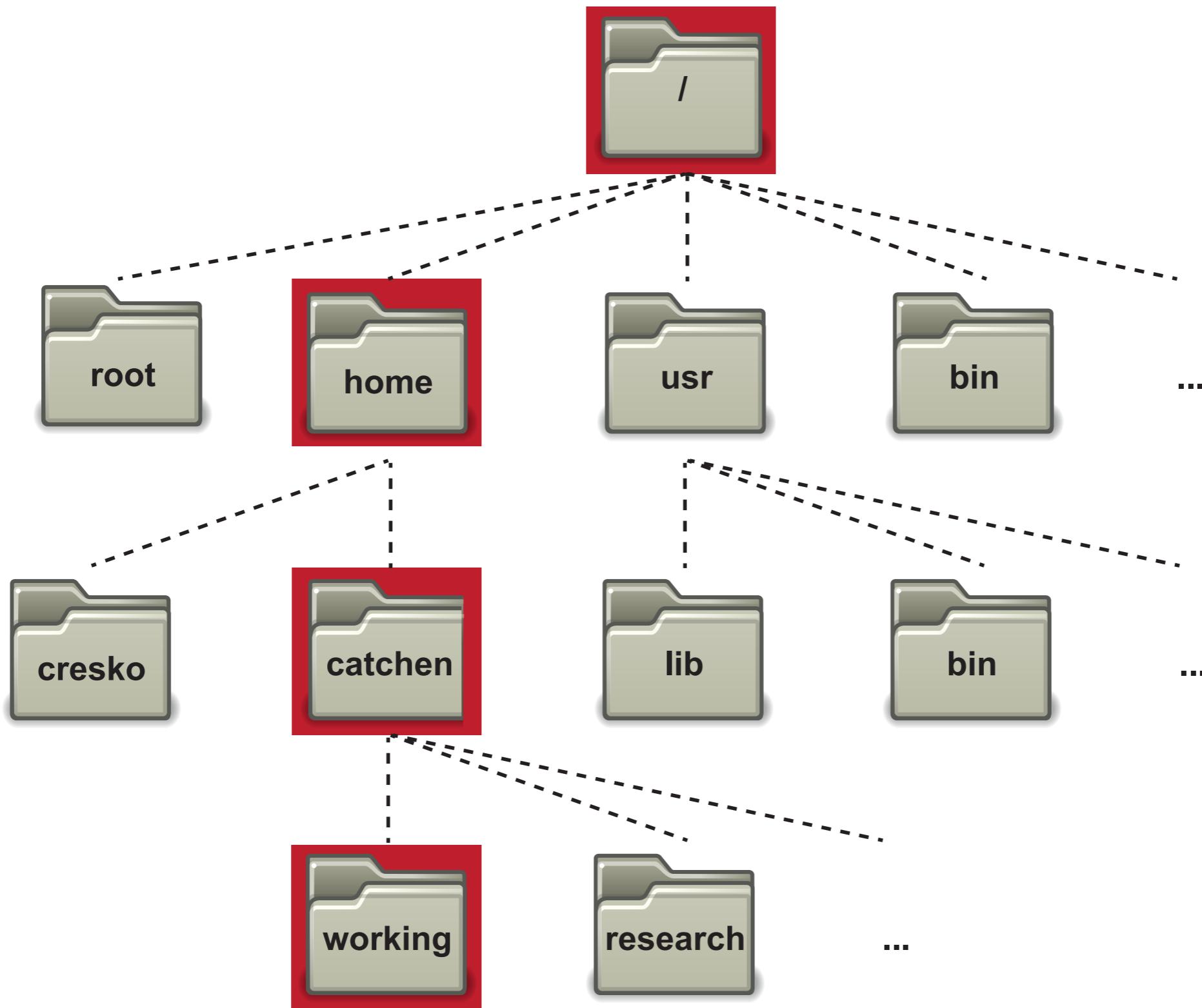
*How do I get to TGAC from The Maids Head Hotel?*



# Paths

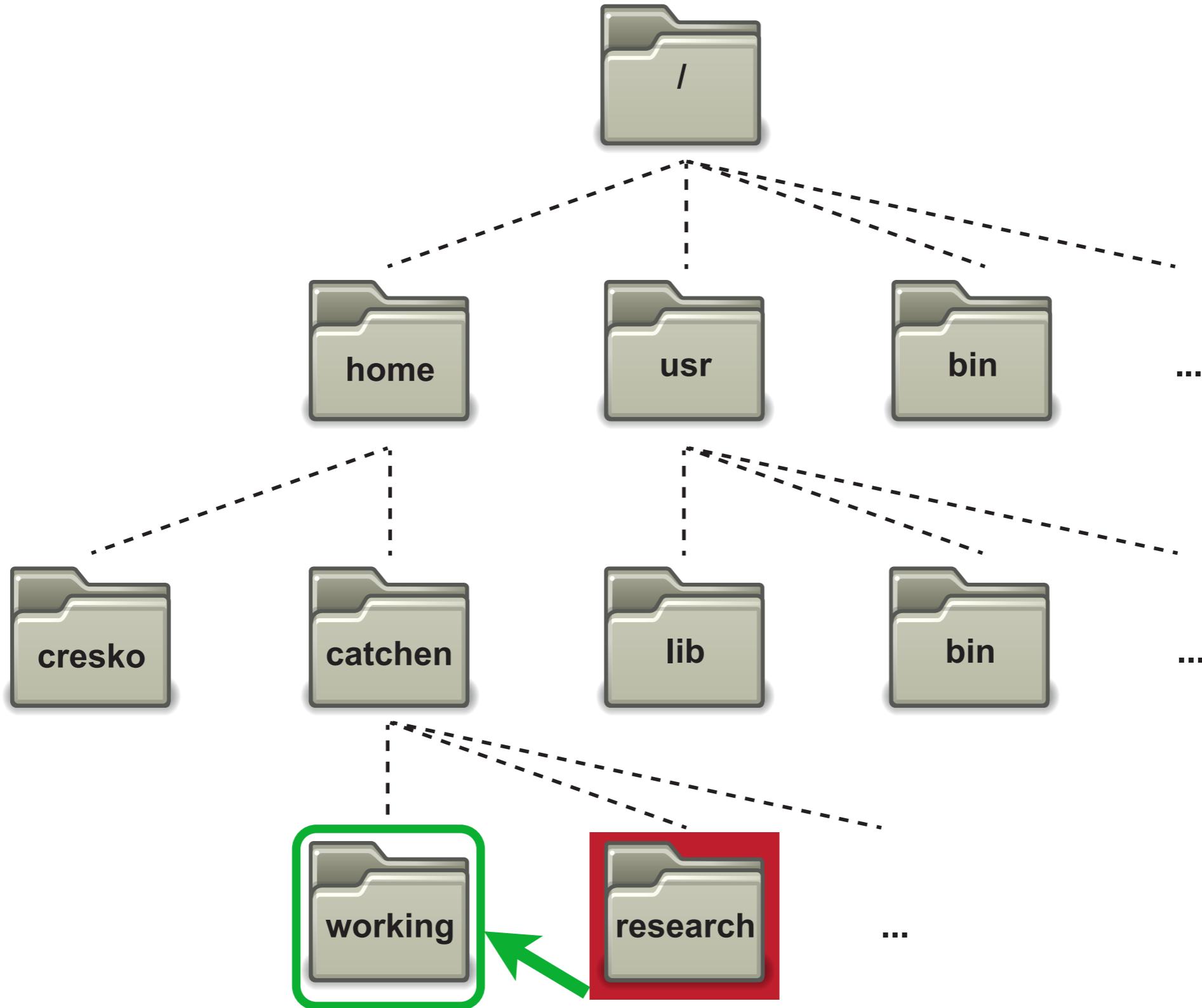


# Absolute Path

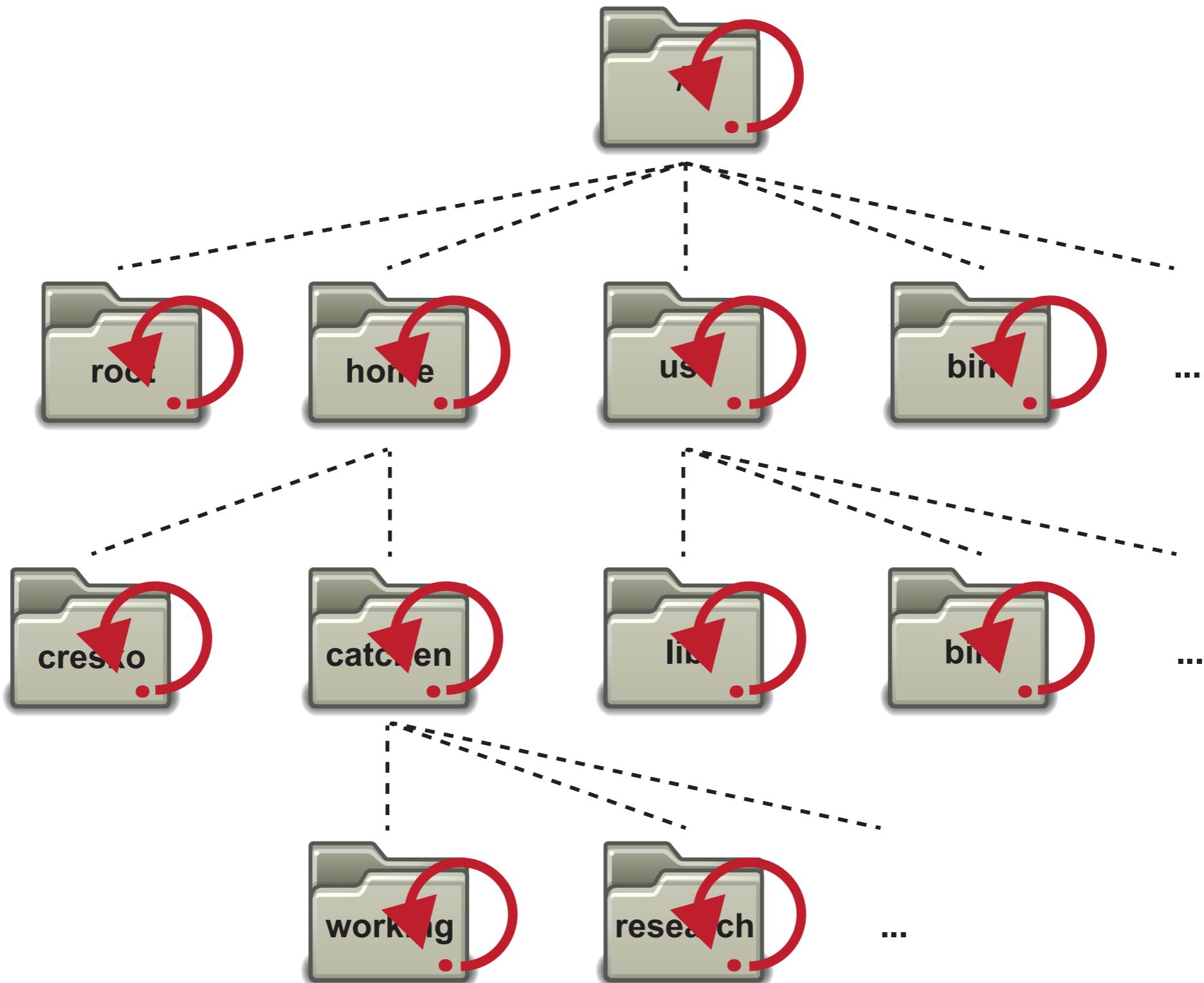


**/home/catchen/working**

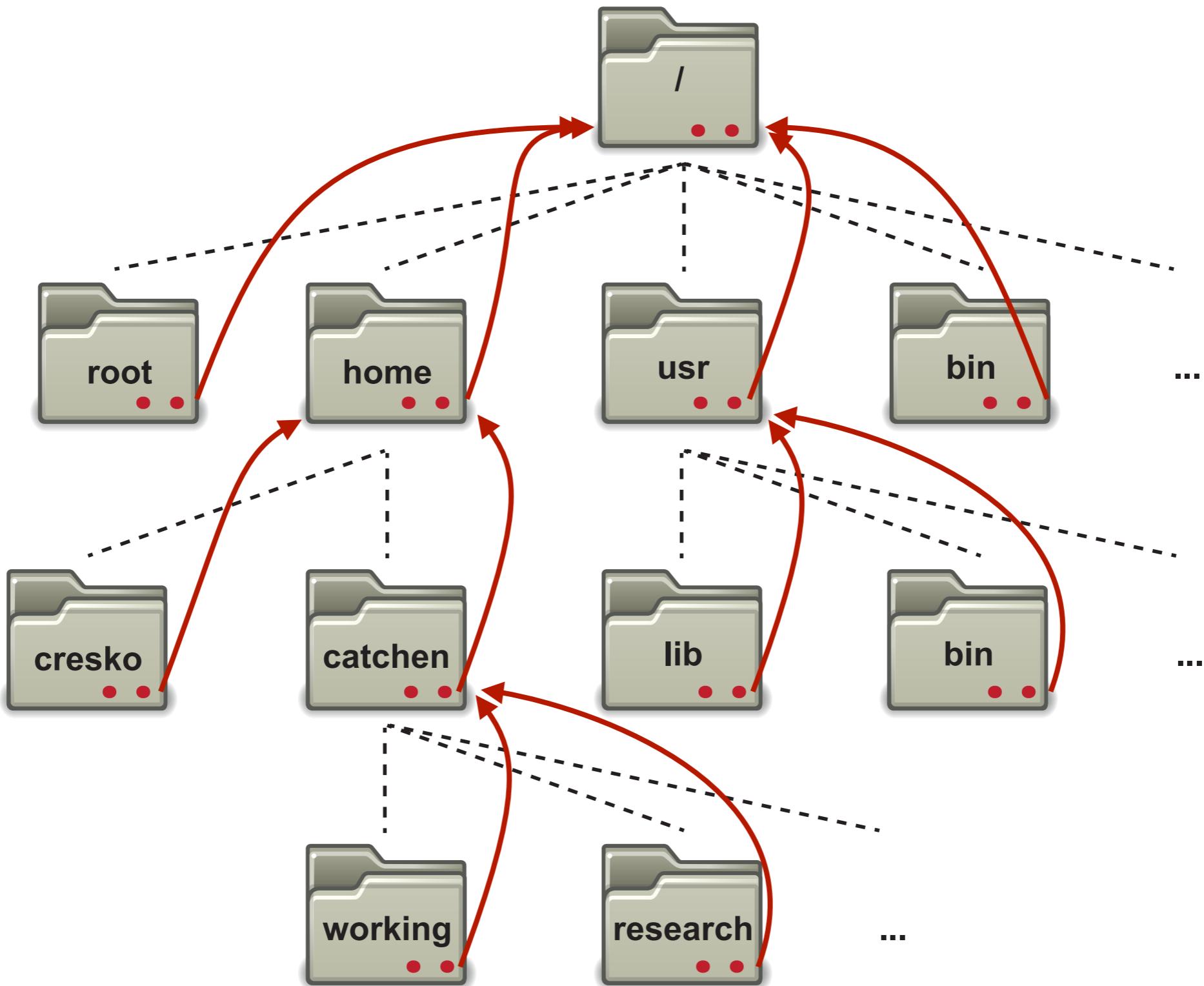
# Relative Path?



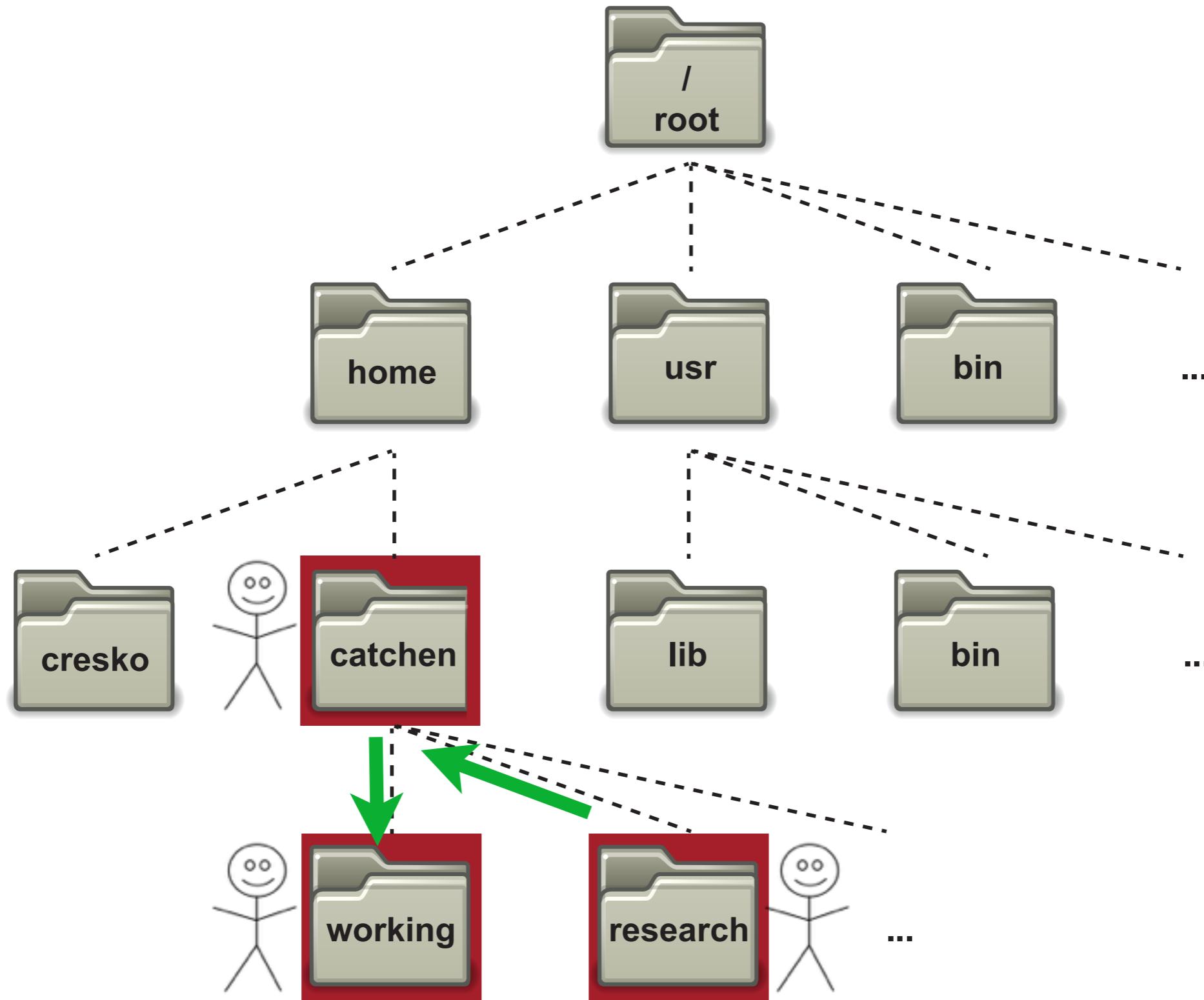
# Special files -- 'dot'



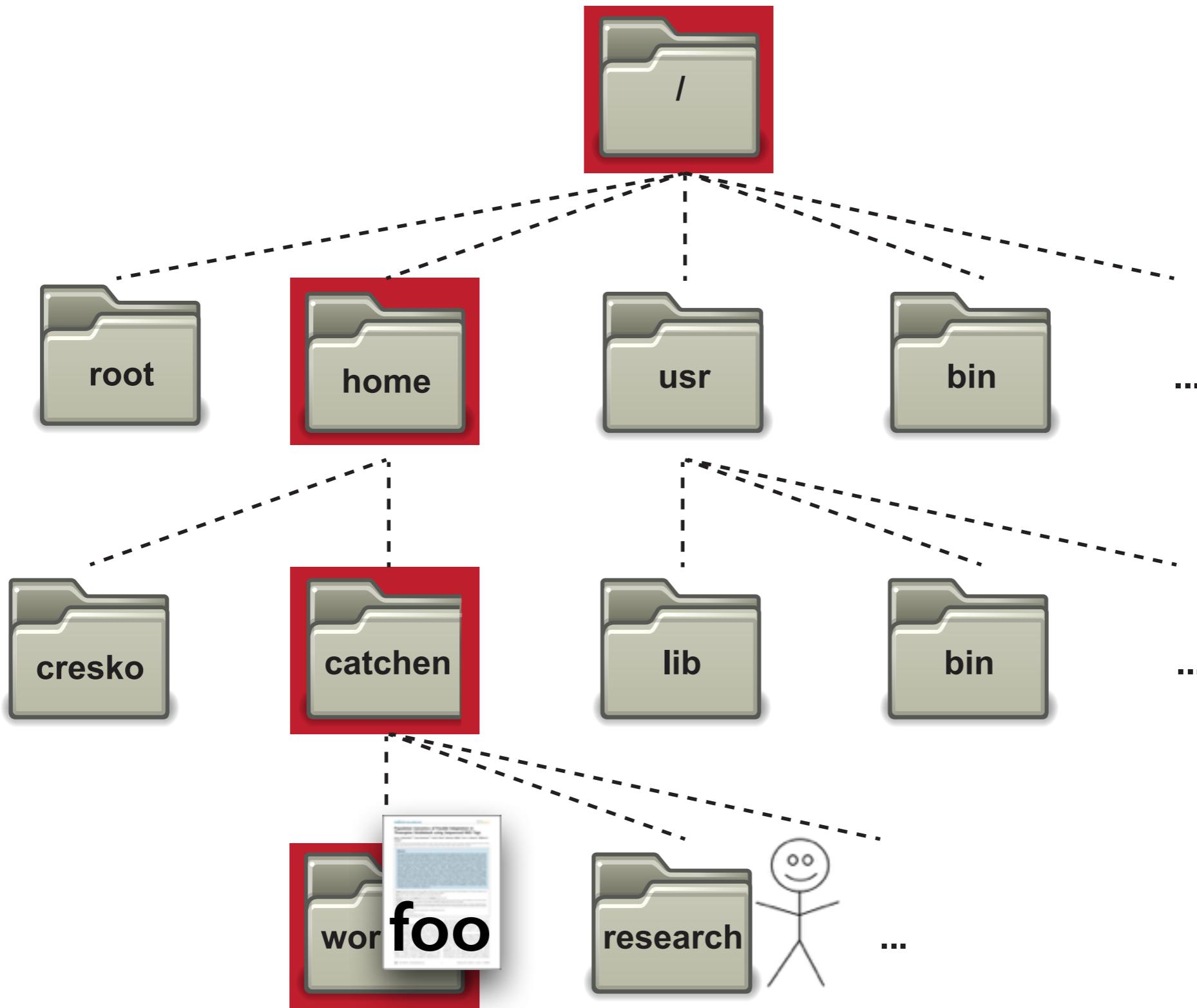
# Special files -- ‘dot dot’



# Relative Path



`../working`

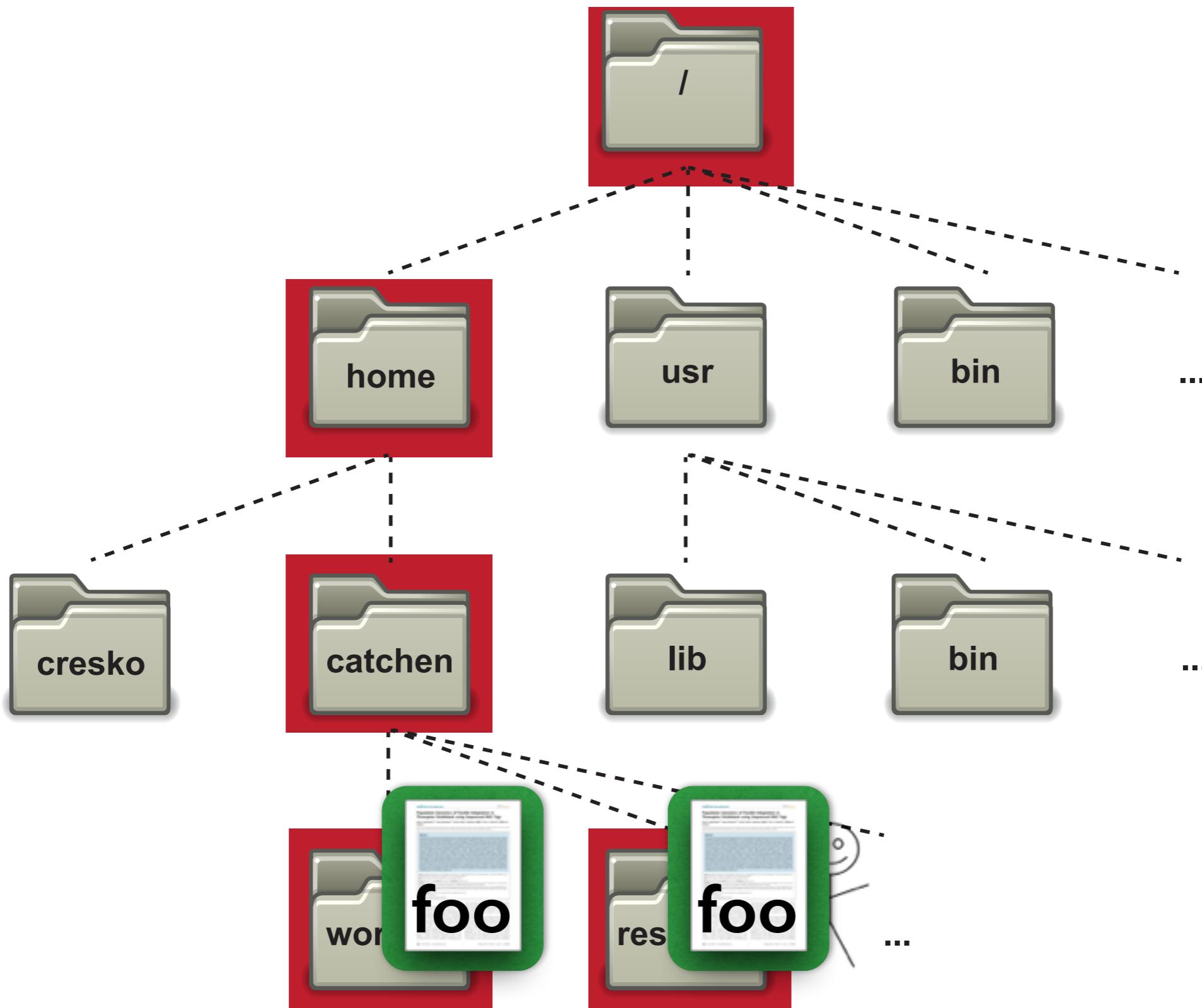


Absolute Path:

**/home/catchen/working/foo**

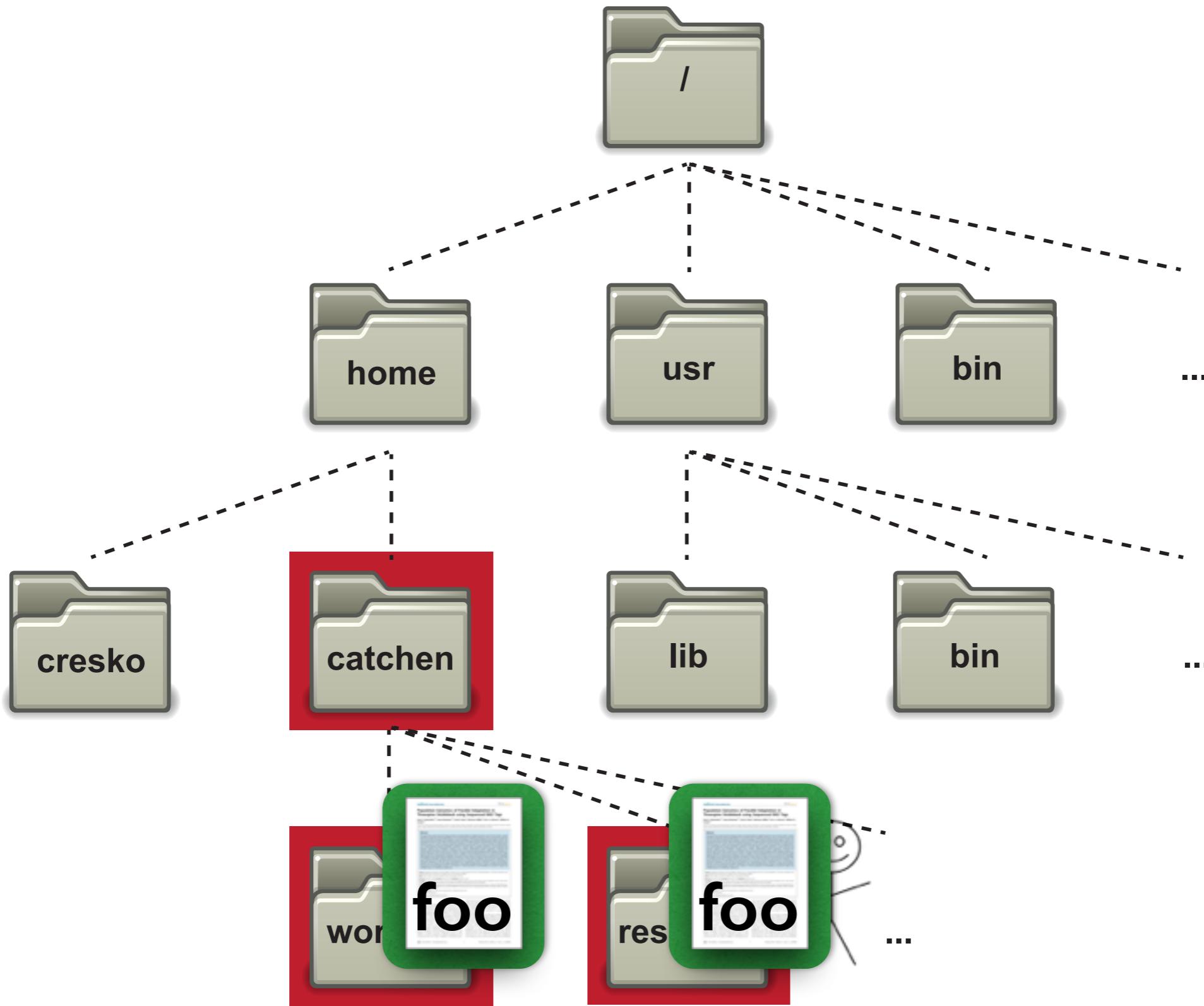
Relative Path:

**.../working/foo**



Absolute Path:  
Absolute Path:

**/home/catchen/working/foo**  
**/home/catchen/research/foo**



Relative Path: **.../working/foo**  
Relative Path: **./foo**

# Absolute and relative paths

```
ubuntu@ip-10-4-193-188:~$ mkdir shell
ubuntu@ip-10-4-193-188:~$ cd shell
ubuntu@ip-10-4-193-188:~/shell$ mkdir research
ubuntu@ip-10-4-193-188:~/shell$ ls
research
ubuntu@ip-10-4-193-188:~/shell$ cd research/
ubuntu@ip-10-4-193-188:~/shell/research$ mkdir seq
ubuntu@ip-10-4-193-188:~/shell/research$ ls
seq
ubuntu@ip-10-4-193-188:~/shell/research$ cd seq/
ubuntu@ip-10-4-193-188:~/shell/research/seq$ mkdir radtags
ubuntu@ip-10-4-193-188:~/shell/research/seq$ cd radtags/
ubuntu@ip-10-4-193-188:~/shell/research/seq/radtags$ ls
ubuntu@ip-10-4-193-188:~/shell/research/seq/radtags$ ls -la
total 8
drwxrwxr-x 2 ubuntu ubuntu 4096 2012-03-06 23:08 .
drwxrwxr-x 3 ubuntu ubuntu 4096 2012-03-06 23:08 ..
ubuntu@ip-10-4-193-188:~/shell/research/seq/radtags$ pwd
/home/ubuntu/shell/research/seq/radtags
ubuntu@ip-10-4-193-188:~/shell/research/seq/radtags$ 
```

Special Files

*dot*

*dot dot*

% ls .

% ls ..

% ls ../../..

# Binary programs - ls, cp, mkdir, etc.



```
ubuntu@ip-10-140-6-74:~$ ls /bin
bash          csh          getfacl      lsblk        ntfsc          readlink    true
bunzip2       dash         grep         lsmod       ntfsccluster   rm          unlockmgr_server
busybox       date         gunzip      lsmod       ntfscmp        rmdir       umount
bzcat         dbus-cleanup-sockets  gexe        mknod      ntfscrypt      rmano      uname
bzcmp         dbus-daemon     gzip        mktemp     ntfsdump_logfile run-parts  uncompress
bzdiff        dbus-uuidgen     hostname    more       ntfsfix        rzsh       unicode_start
bzegrep       dd           init-checkconf mount      ntfsiinfo      sed        vdir
bzexe         df           initctl2dot  mountpoint  ntfsls         setfacl     which
bzfgrep       dir          ip           mt        ntfsmftalloc  setfont    ypdomainname
bzgrep        dmesg        kbd_mode    mt-gnu     ntfsmove      setupcon   zcat
bzip2         dnsdomainname kill        mv        ntfstruncate sh        zjmp
bzip2recover  domainname    ksh         nano      ntfswipe      sh.distrib zdiff
bzless        dumpkeys     less        nc        open          sleep      zegrep
bzmore        echo         lessecho    nc.openbsd  openvt       static-sh  zfgrep
cat           ed           lessfile    netcat     pidof        stty       zforce
chacl         egrep        lesskey     netstat    ping         su        zgrep
chgrp         false        lesspipe    nisdomainname ping6       sync      zless
chmod         fgconsole    ln          ntfs-3g    Plymouth    tailf      zmore
chown        fgrep        loadkeys   ntfs-3g.probe  Plymouth-upstart-bridge tar       znew
chvt          findmnt     login      ntfs-3g.secaudit ps          tcsh      zsh
cp            fuser        lountfs-3g ntfs-3g.usermodel  pwd       tempfile  zsh4
cpio          fusermount   ls          ntfscluster  rbash      touch
```

working

research

...

% ls /bin

# Relative and absolute paths

A shortcut to your ‘home’, tilde:

~

Moving through the filesystem:

**cd**

Knowing where you are:

**pwd**

```
% ls ~/
```

```
% cd ~/
```

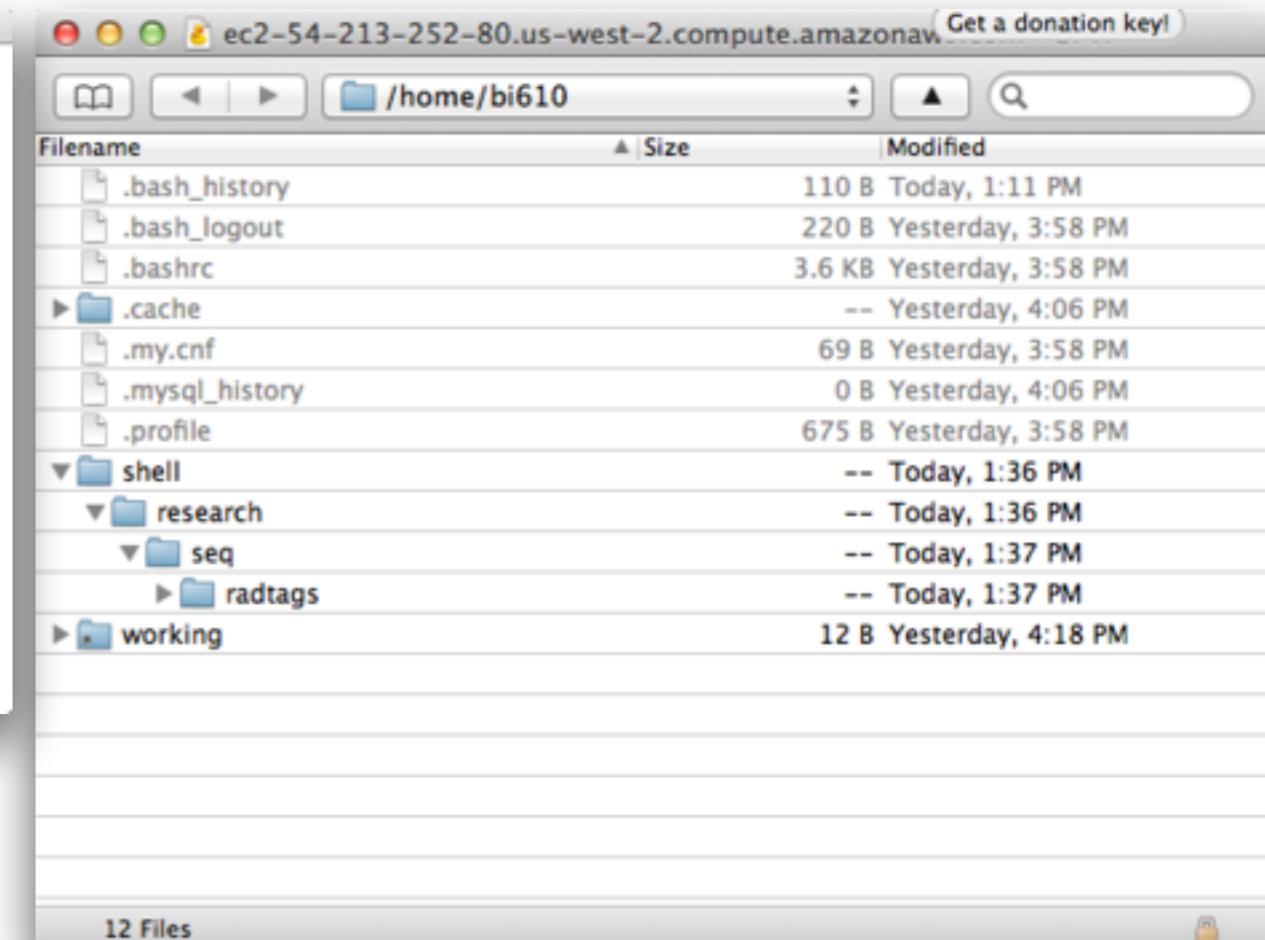
```
% cd
```

```
% pwd
```

# Relative and absolute paths

/home/tgac/shell/research/seq/radtags

```
tgac@Genotyping2:~$ mkdir shell  
tgac@Genotyping2:~$ cd shell  
tgac@Genotyping2:~/shell$ mkdir research  
tgac@Genotyping2:~/shell$ ls  
research  
tgac@Genotyping2:~/shell$ cd research/  
tgac@Genotyping2:~/research$ mkdir seq  
tgac@Genotyping2:~/shell/research$ ls  
seq  
tgac@Genotyping2:~/shell/research$ cd seq/  
tgac@Genotyping2:~/shell/research/seq$ ls  
tgac@Genotyping2:~/shell/research/seq$ mkdir radtags  
tgac@Genotyping2:~/shell/research/seq$ ls  
radtags  
tgac@Genotyping2:~/shell/research/seq$ ls -la  
total 12  
drwxrwxr-x 3 tgac tgac 4096 Nov  4 09:34 .  
drwxrwxr-x 3 tgac tgac 4096 Nov  4 09:34 ..  
drwxrwxr-x 2 tgac tgac 4096 Nov  4 09:34 radtags  
tgac@Genotyping2:~/shell/research/seq$ cd radtags/  
tgac@Genotyping2:~/shell/research/seq/radtags$ pwd  
/home/tgac/shell/research/seq/radtags  
tgac@Genotyping2:~/shell/research/seq/radtags$
```



% ls .

% ls ..

% ls ../../..

% cd ~/

% cd shell/research

% pwd

# Are you typing? You're doing it wrong.

## Tab-completion:

- Tab once to complete uniquely
- Tab twice to see all possible completions

## Up-arrow:

- Previous commands can be found by pressing “up-arrow”

## ‘history’

```
% cd /etc
```

```
% ls c <tab>
```

```
% pwd
```

```
% ls c <tab><tab>
```

```
bi610@ip-172-31-17-121:~/shell/research/seq/radtags$ cd /etc/
bi610@ip-172-31-17-121:/etc$ ls -l
total 868
-rw-r--r-- 1 root root 15070 Apr  4 04:30 a2ps.cfg
-rw-r--r-- 1 root root 2563 Apr  4 04:30 a2ps-site.cfg
drwxr-xr-x 3 root root 4096 Apr 16 18:26 acpi
-rw-r--r-- 1 root root 2981 Apr 16 18:25 adduser.conf
-rw-r--r-- 1 root root 10 Apr 16 18:27 adjtime
drwxr-xr-x 2 root root 4096 Jun 20 22:23 alternatives
drwxr-xr-x 8 root root 4096 Jun 20 21:30 apache2
drwxr-xr-x 3 root root 4096 Apr 16 18:25 apt
drwxr-xr-x 3 root root 4096 Apr 16 18:26 apparmor
drwxr-xr-x 8 root root 4096 Jun 19 22:10 apparmor.d
drwxr-xr-x 3 root root 4096 Jun 19 22:00 apport
drwxr-xr-x 6 root root 4096 Apr 16 18:27 apt
-rw-r----- 1 root daemon 144 Oct 21 2013 at.deny
-rw-r--r-- 1 root root 2177 Apr  9 01:03 bash.bashrc
-rw-r--r-- 1 root root 45 Mar 22 20:57 bash_completion
drwxr-xr-x 2 root root 4096 Jun 20 21:44 bash_completion.d
-rw-r--r-- 1 root root 356 Jan  1 2012 bindresvport.blacklist
-rw-r--r-- 1 root root 321 Apr 16 16:07 blkid.conf
lrwxrwxrwx 1 root root 15 Jun  3 20:54 blkid.tab -> /dev/.blkid.tab
drwxr-xr-x 2 root root 4096 Apr 16 18:26 byobu
drwxr-xr-x 3 root root 4096 Apr 16 18:25 ca-certificates
-rw-r--r-- 1 root root 7464 Apr 16 18:26 ca-certificates.conf
drwxr-xr-x 2 root root 4096 Apr 16 18:26 calendar
drwxr-s--- 2 root dip 4096 Apr 16 18:26 chatscripts
drwxr-xr-x 4 root root 4096 Apr 16 18:27 cloud
drwxr-xr-x 2 root root 4096 Apr 16 18:25 console-setup
drwxr-xr-x 2 root root 4096 Jun 19 22:10 cron.d
drwxr-xr-x 2 root root 4096 Jun 19 22:10 cron.daily
drwxr-xr-x 2 root root 4096 Apr 16 18:25 cron.hourly
drwxr-xr-x 2 root root 4096 Apr 16 18:25 cron.monthly
-rw-r--r-- 1 root root 722 Feb  9 2013 crontab
drwxr-xr-x 2 root root 4096 Jun 20 22:23 cron.weekly
drwxr-xr-x 1 root root 54 Apr 16 18:26 crypttab
drwxr-xr-x 4 root root 4096 Apr 16 18:26 dbus-1
-rw-r--r-- 1 root root 2969 Feb 23 14:30 debconf.conf
-rw-r--r-- 1 root root 11 Feb 20 02:43 debian_version
drwxr-xr-x 3 root root 4096 Jun 20 22:23 default
```

# Three variants to `ls`

<code>ls -l</code>	<code>ls -la</code>	<code>ls -lh</code>
provides a <i>long</i> listing	includes <i>all</i> files, even hidden files	displays file sizes in <i>human</i> readable numbers

# Four ways to view a text file

more	head	tail	cat
<b>view a text file one screen full at a time</b>	<b>view the top 15 lines of a file</b>	<b>view the last 15 lines of a file</b>	<b>spit the whole file at once</b>
<b>space-bar:</b> <b>scroll</b> <b>q: quit</b>	<b>-n num</b> <b>controls the number of lines</b>	<b>-n num</b> <b>controls the number of lines</b>	

# Explore the file hierarchy

## I. Move to the directory /etc

- What is the first line of the file ‘hosts’ in the directory /etc?
- What is the relative file path to get to /var/log from here?
- What is the absolute path?

## 2. Move to the directory /var/log/

- What is the contents on line 73 of the dmesg file?
- Without changing directories, what is the second line of the cpuinfo file in the proc directory? (proc directory is under root)
  - What is the command to read this file with a relative path?
  - An absolute path?

## 3. Move back to the root, what directories do you see?

## 4. Move back home, what are three ways to move home from the root?

# Copy example files

Return to the directory in your home  
called ‘shell’.

TSV file:

/data/unix/batch\_1.genotypes\_1.loc.gz

FASTQ file:

/data/unix/s\_1\_sequence.txt.gz

Tar Archive:

/data/unix/samples.tar.gz

# What is a tar archive?



**tar = tape archive**

# Compress / Decompress

**gzip / gunzip**

batch\_I.genotypes\_I.loc.gz

s\_I\_sequence.txt.gz

Gzipped Tar archive

**tar xvzf**

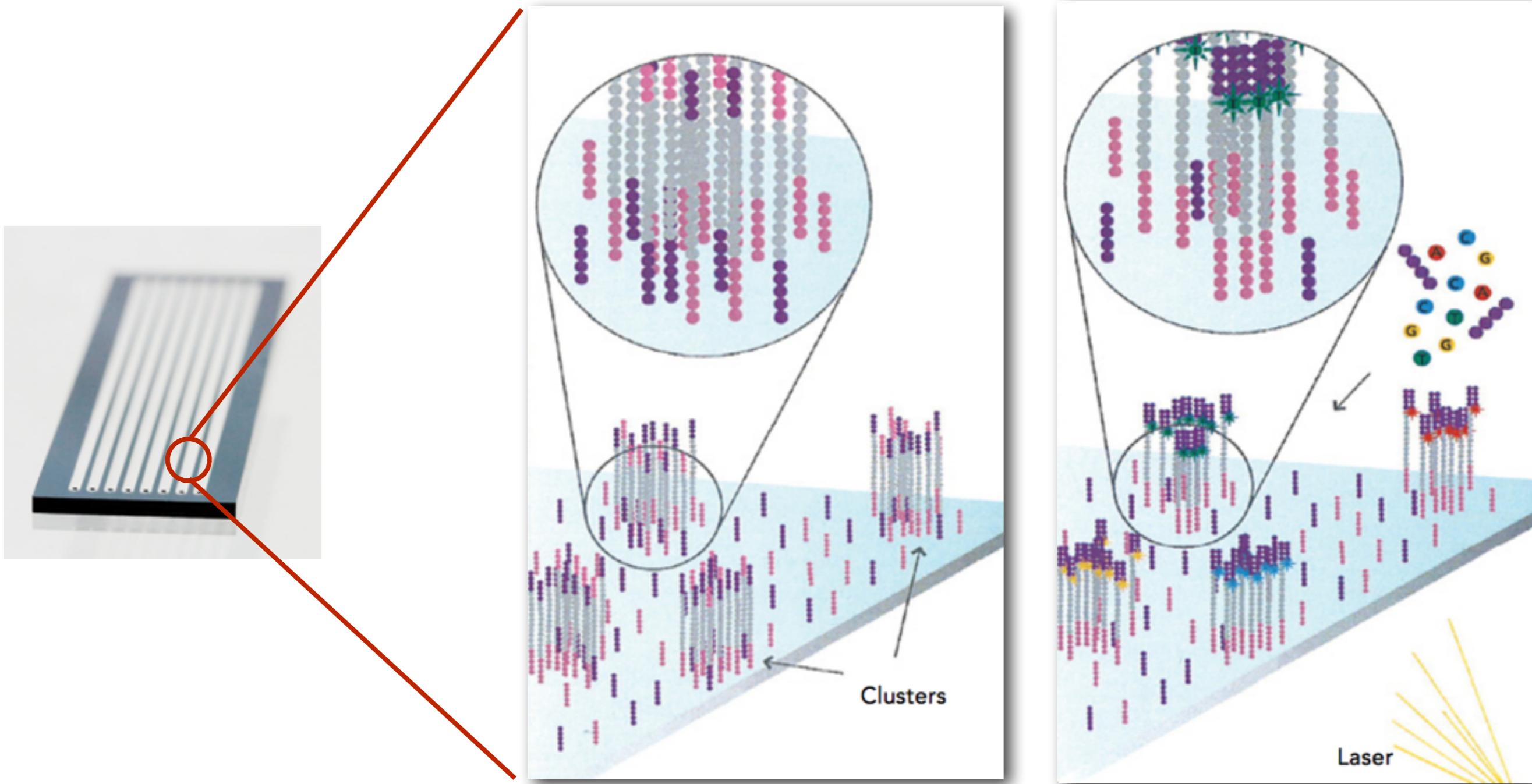
samples.tar.gz

Tar archive

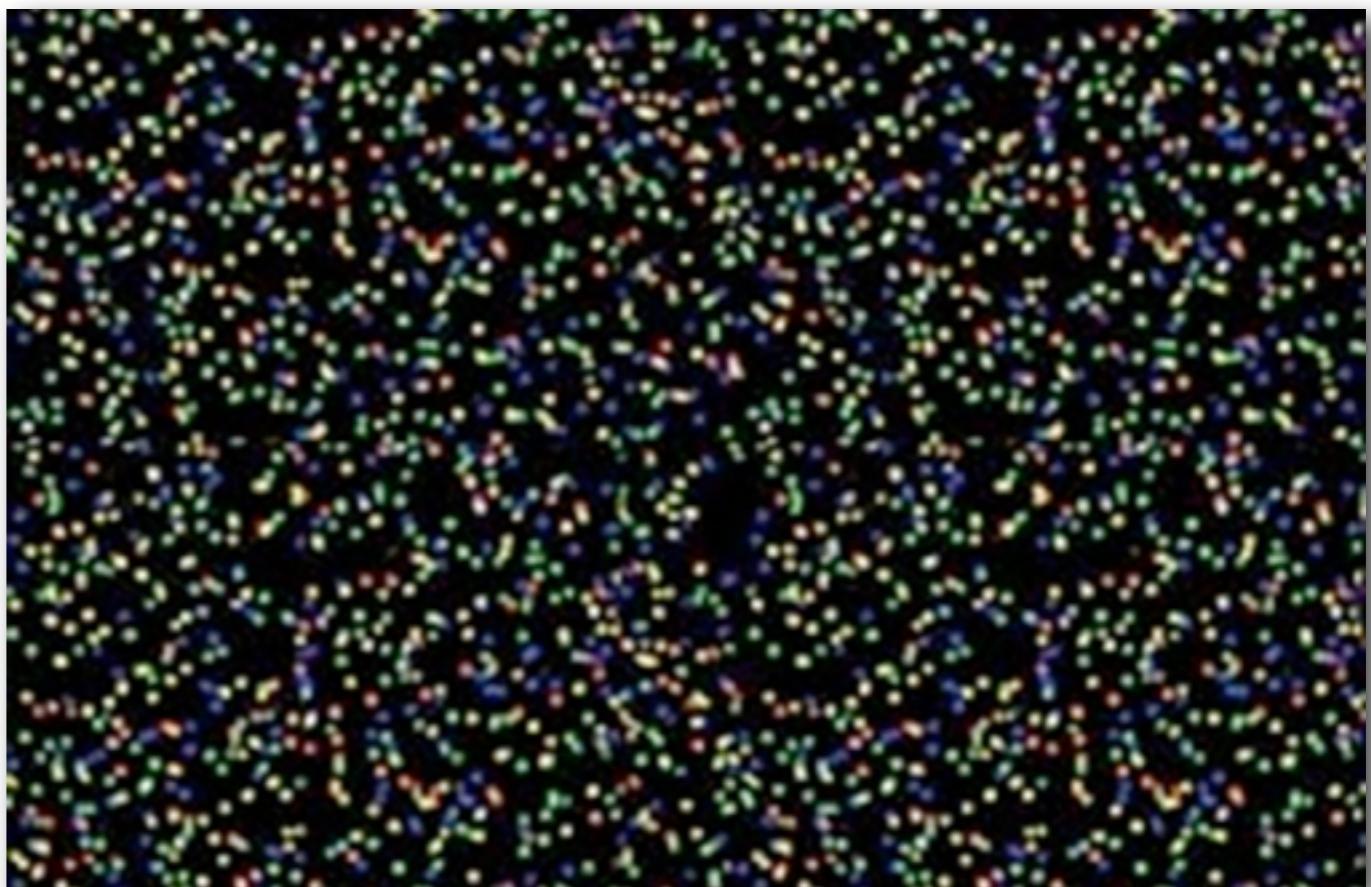
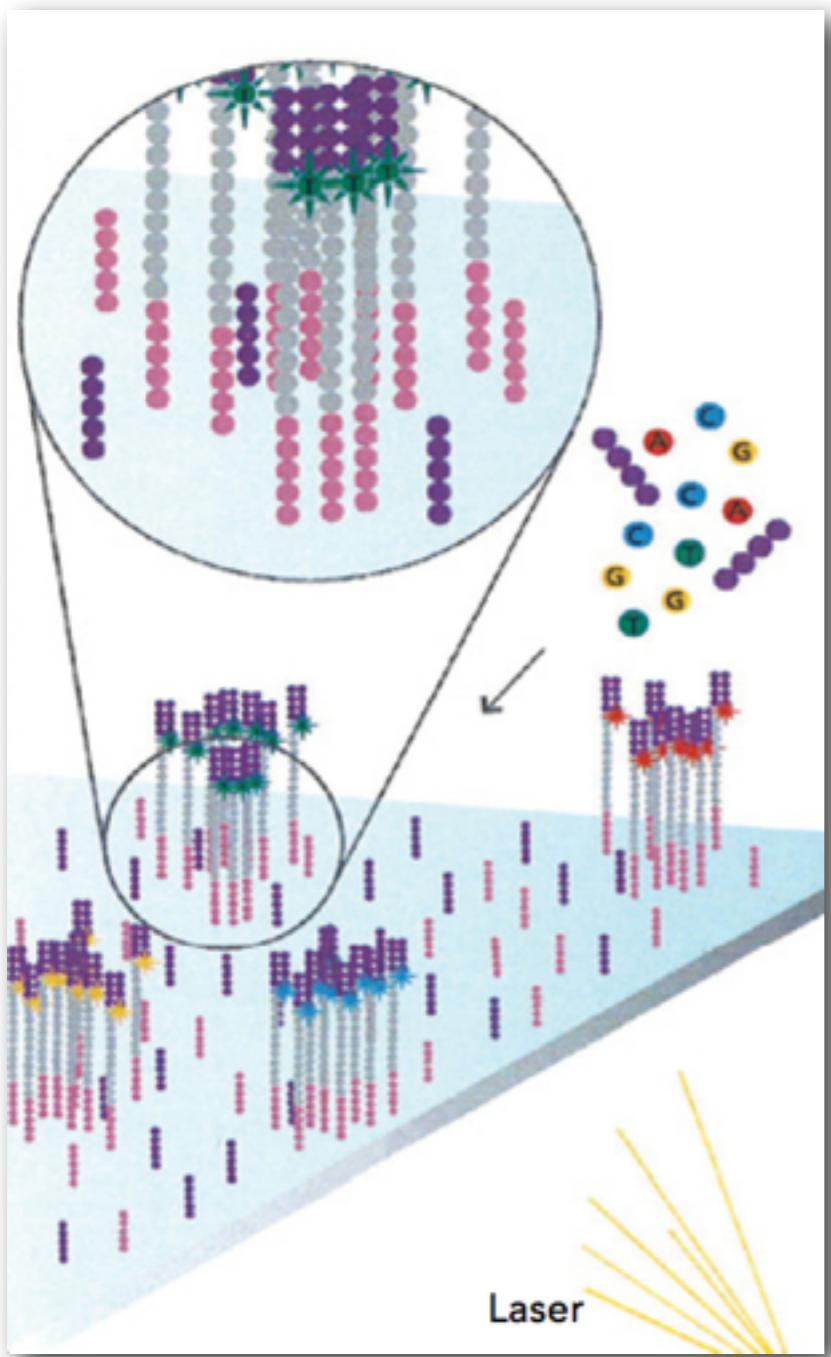
**tar xvf**

samples.tar

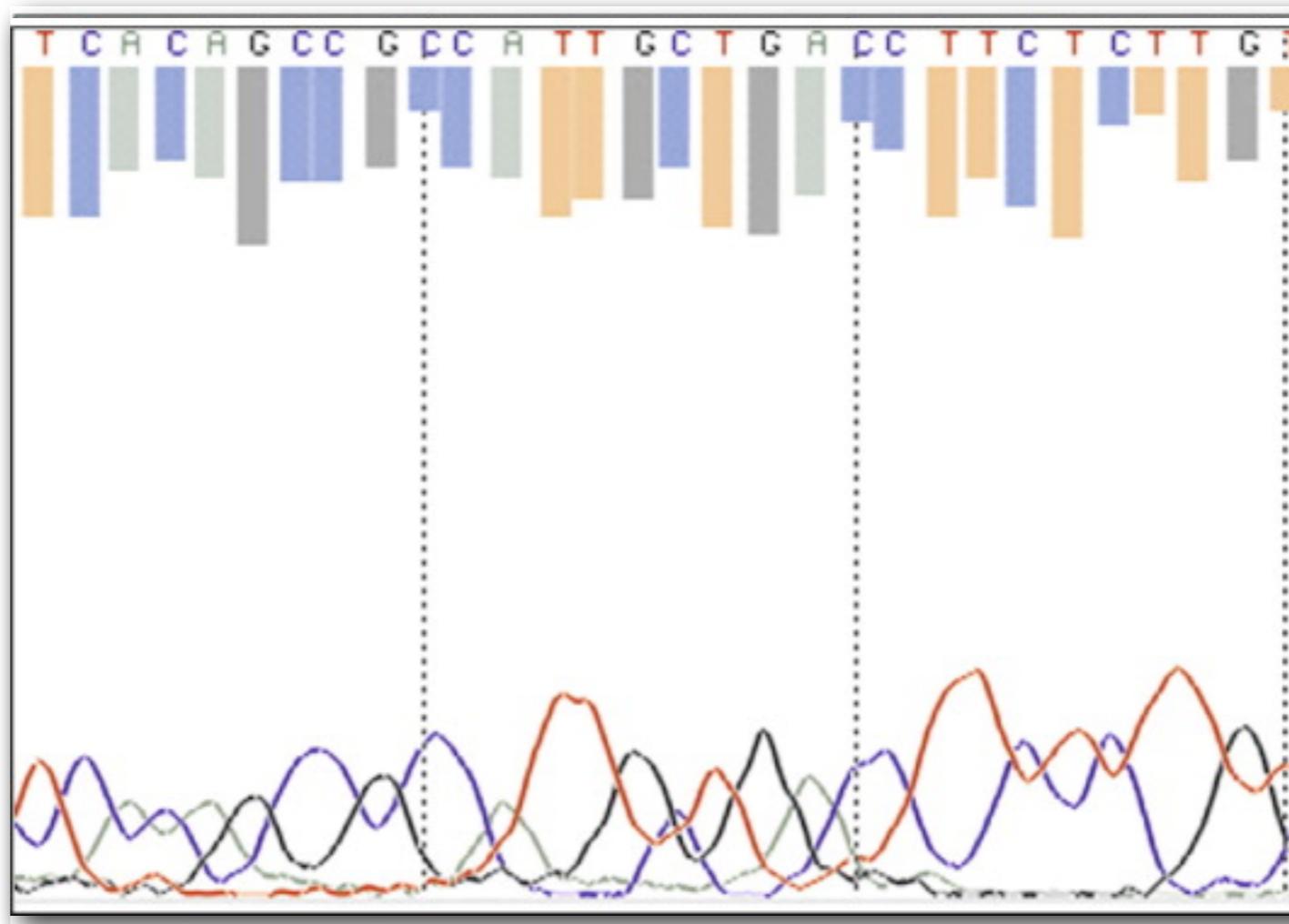
# Sequencing on Illumina's Flow cell



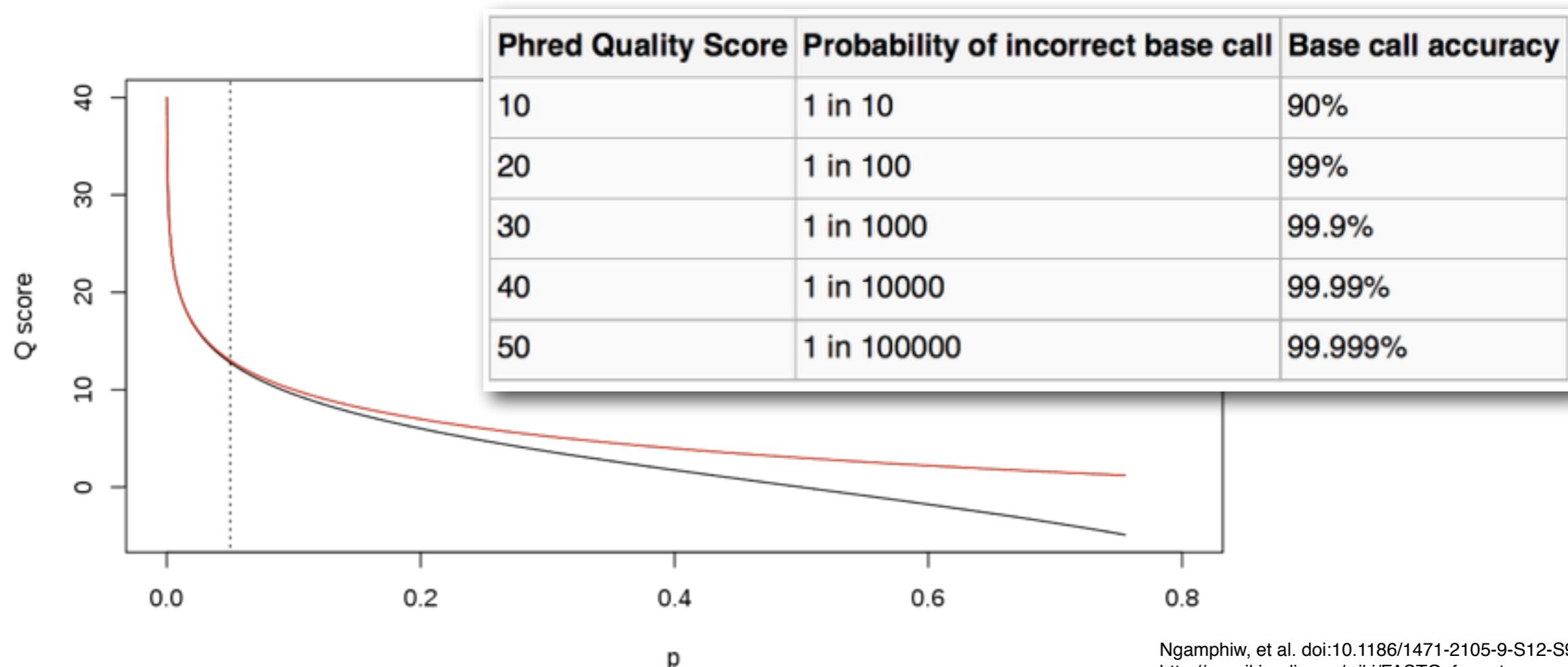
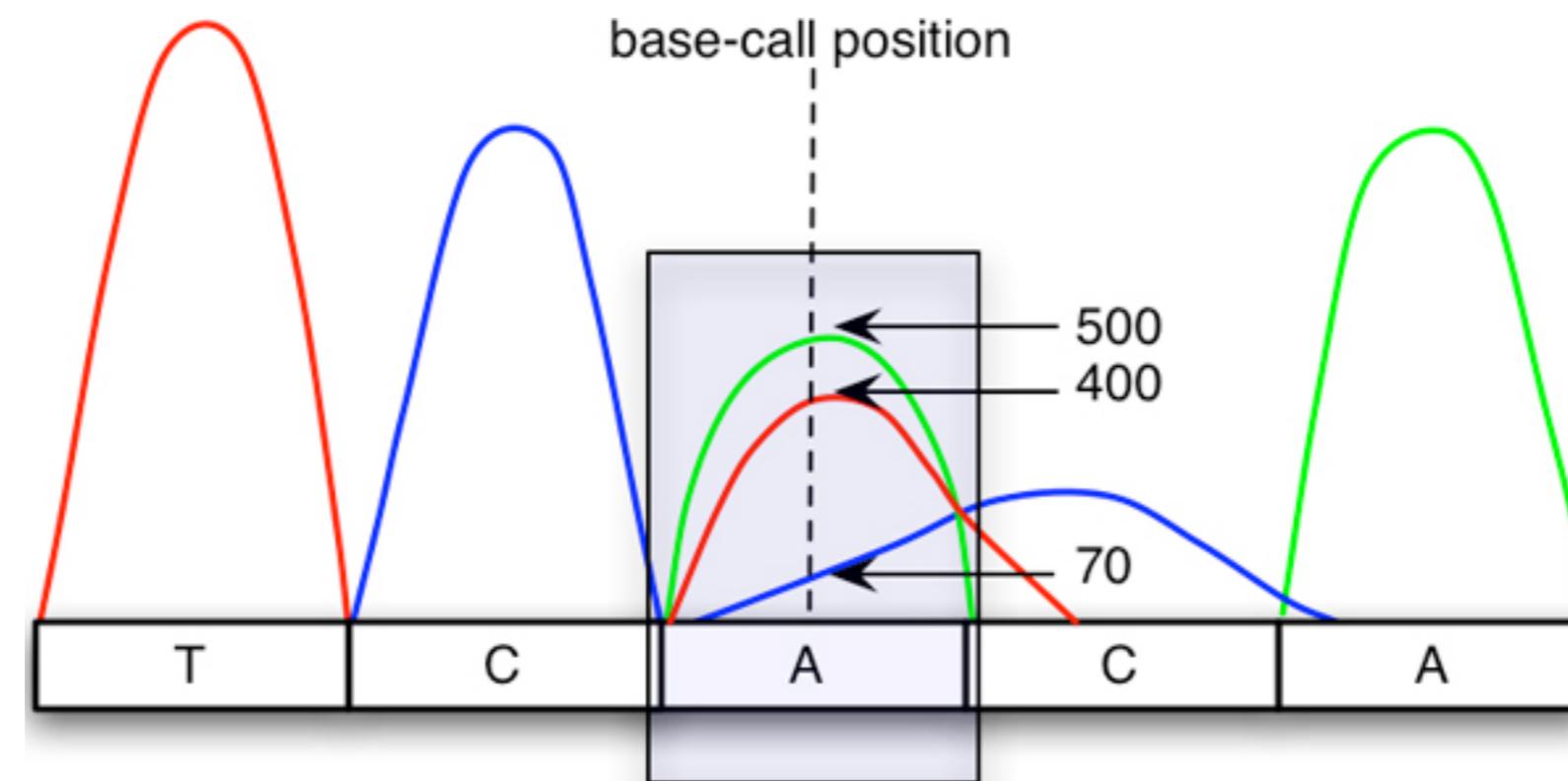
# Sequencing on Illumina's Flow cell, ctd.



# Sequencing on Illumina's Flow cell, ctd.



# Phred Quality Score



# The FASTQ File Format

## FASTA

```
>chromosome7
TTTGTCTGCAGGGGGACACGTCAAAGTCAAACGCAGGCAAGTTGTGTTATGTCCAGTGGATCTTTGATT
ACATACTGCAGGGTCAGGAGGATTATCTCCTCTGCAAGGTAACGCCTGCTGTAACC GTTCTTCATCCTTT
CCTAACTGCAGGGCTGTCTGTCAGGTCTGACAAGACATATGCAGGGCTCAATTGAGATAATTGCTCAATATA
```

## FASTQ

```
@Sequence_137
TTTGTCTGCAGGGGGACACGTCAAAGTCAAACGCAGGCAAGTTGTGTTATGTCCAGTGGATCTTTGATT
+Sequence_137
<?@DDDDDFHHFBB@GGIACFHGGHBGHGCDHBEAHACHI=@CH.=7ACAHHADECDBCC66(6>@C>5@CACCA
```

```
@HWI-ST0747:162:C03AJACXX:3:1108:19763:106771 1:N:0:
TTTGTCTGCAGGGGGACACGTCAAAGTCAAACGCAGGCAAGTTGTGTTATGTCCAGTGGATCTTTGATT
+
<?@DDDDDFHHFBB@GGIACFHGGHBGHGCDHBEAHACHI=@CH.=7ACAHHADECDBCC66(6>@C>5@CACCA
```

# ASCII Code

0	<NUL>	32	<SPC>	64	@	96	'	128	Ä	160	†	192	�	224	‡
1	<SOH>	33	!	65	A	97	a	129	Å	161	°	193	i	225	.
2	<STX>	34	"	66	B	98	b	130	Ç	162	¢	194	¬	226	,
3	<ETX>	35	#	67	C	99	c	131	�	163	£	195	�	227	"
4	<EOT>	36	\$	68	D	100	d	132	�	164	§	196	f	228	%o
5	<ENQ>	37	%	69	E	101	e	133	�	165	•	197	�	229	�
6	<ACK>	38	&	70	F	102	f	134	�	166	¶	198	�	230	�
7	<BEL>	39	'	71	G	103	g	135	�	167	�	199	�	231	�
8	<BS>	40	(	72	H	104	h	136	�	168	�	200	»	232	�
9	<TAB>	41	)	73	I	105	i	137	�	169	�	201	...	233	�
10	<LF>	42	*	74	J	106	j	138	�	170	�	202	...	234	�
11	<VT>	43	+	75	K	107	k	139	�	171	�	203	�	235	�
12	<FF>	44	,	76	L	108	l	140	�	172	�	204	�	236	�
13	<CR>	45	-	77	M	109	m	141	�	173	�	205	�	237	�
14	<SO>	46	.	78	N	110	n	142	�	174	�	206	�	238	�
15	<SI>	47	/	79	O	111	o	143	�	175	�	207	�	239	�
16	<DLE>	48	0	80	P	112	p	144	�	176	�	208	-	240	�
17	<DC1>	49	1	81	Q	113	q	145	�	177	�	209	-	241	�
18	<DC2>	50	2	82	R	114	r	146	�	178	�	210	�	242	�
19	<DC3>	51	3	83	S	115	s	147	�	179	�	211	�	243	�
20	<DC4>	52	4	84	T	116	t	148	�	180	�	212	�	244	�
21	<NAK>	53	5	85	U	117	u	149	�	181	�	213	�	245	�
22	<SYN>	54	6	86	V	118	v	150	�	182	�	214	�	246	�
23	<ETB>	55	7	87	W	119	w	151	�	183	�	215	�	247	�
24	<CAN>	56	8	88	X	120	x	152	�	184	�	216	�	248	�
25	<EM>	57	9	89	Y	121	y	153	�	185	�	217	�	249	�
26	<SUB>	58	:	90	Z	122	z	154	�	186	�	218	/	250	�
27	<ESC>	59	;	91	[	123	{	155	�	187	�	219	�	251	�
28	<FS>	60	<	92	\	124		156	�	188	�	220	<	252	�
29	<GS>	61	=	93	]	125	}	157	�	189	�	221	>	253	�
30	<RS>	62	>	94	^	126	~	158	�	190	�	222	fi	254	�
31	<US>	63	?	95	_	127	<DEL>	159	�	191	�	223	fl	255	�

# The FASTQ File Format, ctd

```
@HWI-ST0747:162:C03AJACXX:3:1108:19763:106771 1:N:0:  
TTTGTCTGCAGGGGGACACGTCAAAGTCAAACGCAGGCAAGTTGTGTTATGTCCAGTGGATCTTGATTT  
+  
<?@DDDDDHFHHFBB@GGIACFHGGHBGHGCDHBEAHACHI=@CH.=7ACAHHADECDBCC66(6>@C>5@CACCA
```

# Quality Scores

S - Sanger Phred+33, raw reads typically (0, 40)  
 X - Solexa Solexa+64, raw reads typically (-5, 40)  
 I - Illumina 1.3+ Phred+64, raw reads typically (0, 40)  
 J - Illumina 1.5+ Phred+64, raw reads typically (3, 40)  
     with 0=unused, 1=unused, 2=Read Segment Quality Control Indicator (bold)  
     (Note: See discussion above).  
 L - Illumina 1.8+ Phred+33, raw reads typically (0, 41)

**ASCII values 33 - 73 = 0 - 40**

**‘F’ = 70**

$$70 - 33 = 37$$

# The FASTQ File Format, ctd

```
@HWI-ST0747:162:C03AJACXX:3:1108:19763:106771 1:N:0:  
TTTGTCTGCAGGGGGACACGTCAAAGTCAAACGCAGGCAAGTTGTGTTATGTCCAGTGGATCTTTGATT  
+  
<?@DDDDDFHHFBB@GGIACFHGGHBGHGCDHBEAHACHI=@CH.=7ACAHHADECDBCC66(6>@C>5@CACCA
```

$$70 - 33 = 37$$

Phred Quality Score	Probability of incorrect base call	Base call accuracy
10	1 in 10	90%
20	1 in 100	99%
30	1 in 1000	99.9%
40	1 in 10000	99.99%
50	1 in 100000	99.999%

## Count raw reads:

```
wc -l s_1_sequence.txt
```

```
grep "@" s_1_sequence.txt  
grep -c "@" s_1_sequence.txt
```

```
grep -v "@" s_1_sequence.txt  
grep -v -c "@" s_1_sequence.txt
```

## Count reads with barcode:

```
grep -c "^CGATA" s_1_sequence.txt
```

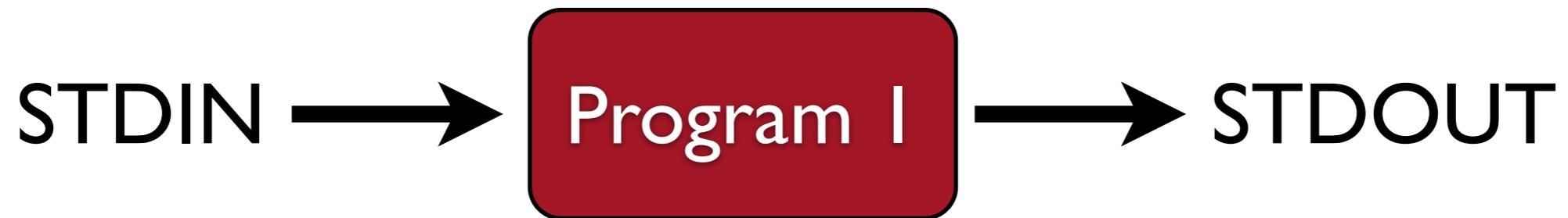
# Special Files

**STDIN, STDOUT, STDERR**

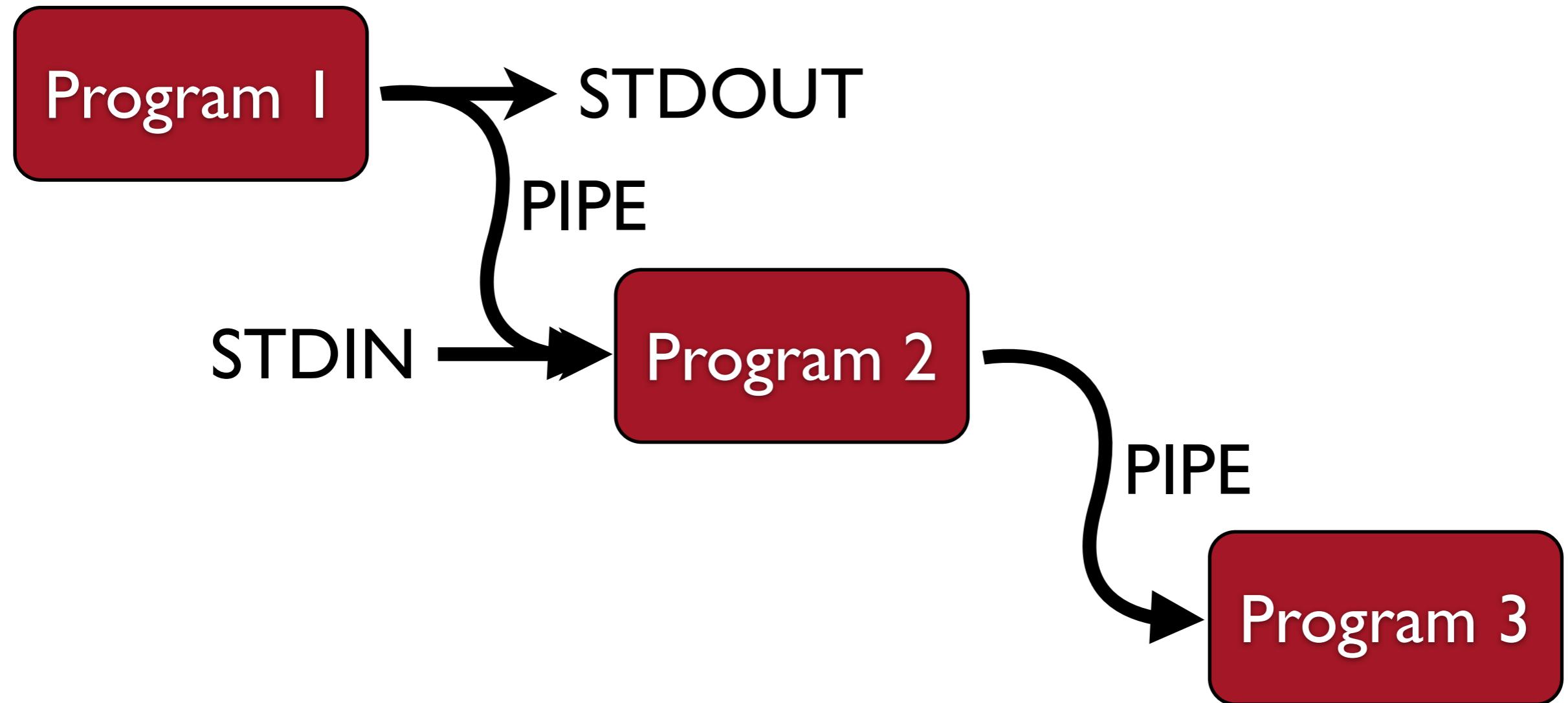
**The Shell's Killer App: Pipes**



# The Shell's Killer App: Pipes, ctd.



# The Shell's Killer App: Pipes, ctd.



So what is the purpose of the program **cat**?

**cut**

```
cut -f 10 batch_1.genotypes_1.loc
```

**cut, capture the output**

```
cut -f 1-10 batch_1.genotypes_1.loc > genos
```

**cut, pipe the output to grep**

```
cut -f 2 batch_1.genotypes_1.loc | grep -c "nnxnp"
```

```
cut -f 1-10,15,17 batch_1.genotypes_1.loc | grep "nnxnp" > genos2
```

**Examine a marker, translating the output**

```
cat batch_1.genotypes_1.loc | tr " " "," | grep "^96053"
```

## s\_I\_sequence.txt.gz

Decompress the file

1. Count the number of raw reads (250,000)
2. Count the number of reads with barcode CGATA (19,501)
3. Capture all FASTQ records for ACCAT into a file called sample\_01.fq (you should get 18352 records, 73408 lines)
4. Determine the count of all barcodes in the file

```
286 CTAGT  
7900 TCAGA  
10659 ACTGC  
10931 TGACC  
11536 GAGAT  
11871 CTGAA  
14409 CGGCG  
14508 TGGTT  
18226 GAAGC  
18352 ACCAT  
18375 TCGAG  
19501 CGATA  
23012 AATTT  
26336 GCATT  
31136 CTAGG
```

1. Use **head** when building a command, **cat** once the command is working
2. Look at the **-n** option for the **head** command, the **-l** option for **wc**
3. The “**^**” character means “must occur at beginning of line” in a grep search
4. Look at the **grep** options: **-c**, **-v**, **-A**, **-B**
5. Read the man pages for **sort** and **uniq** to learn how to combine them

```
ls  
gunzip  
man  
more  
cat  
wc  
head  
cut  
grep  
sort  
uniq  
>  
|
```