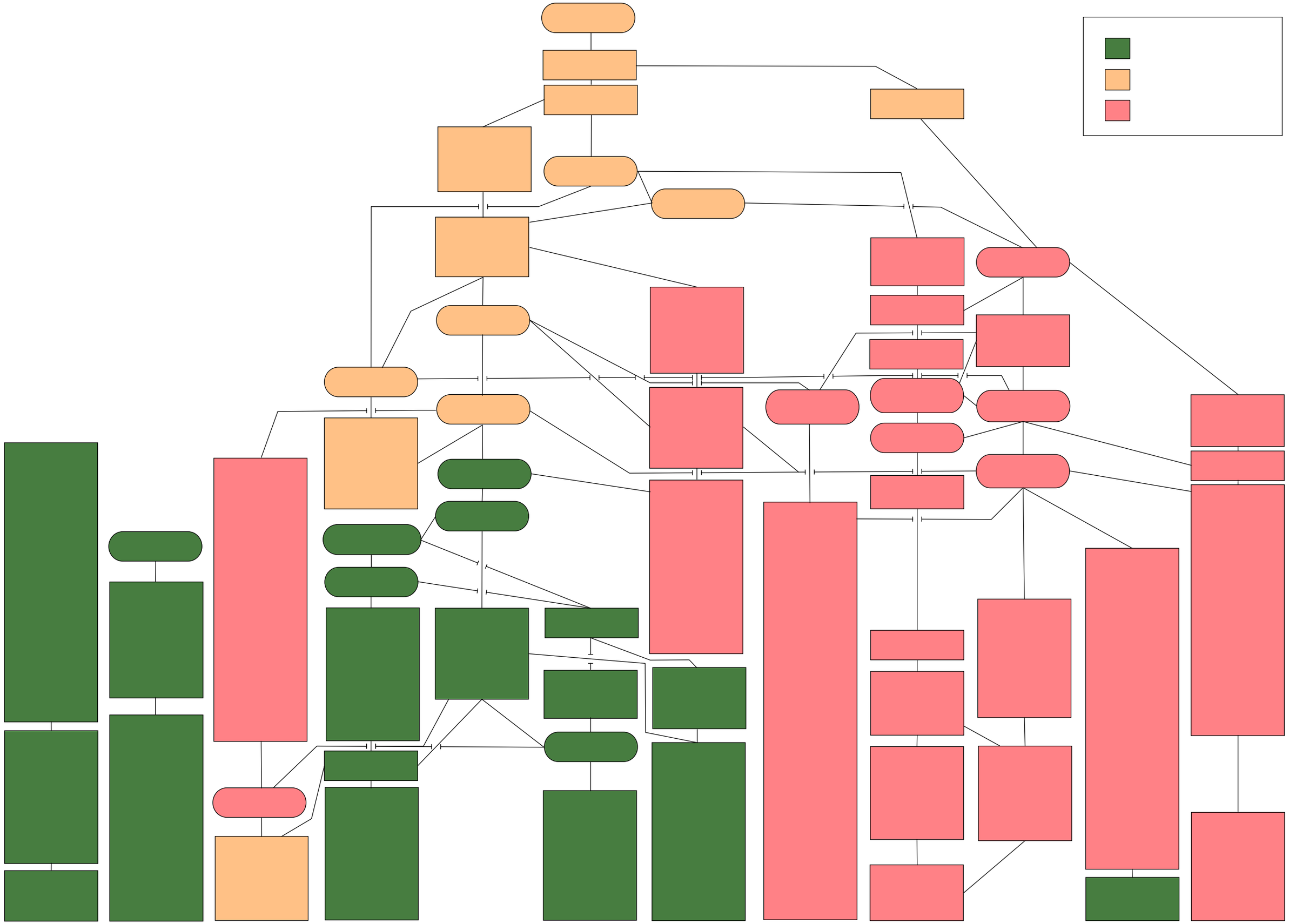
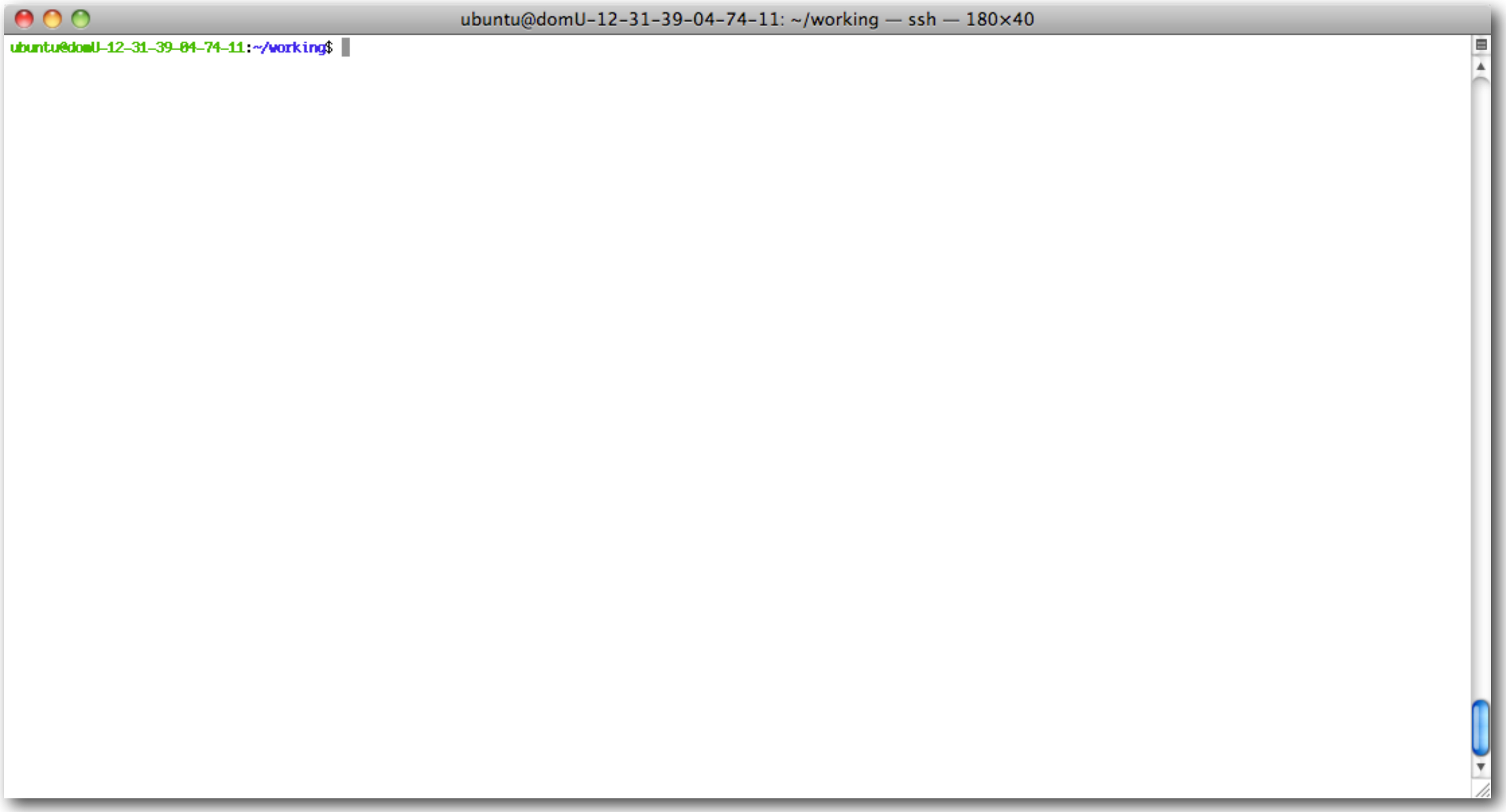
Unix History

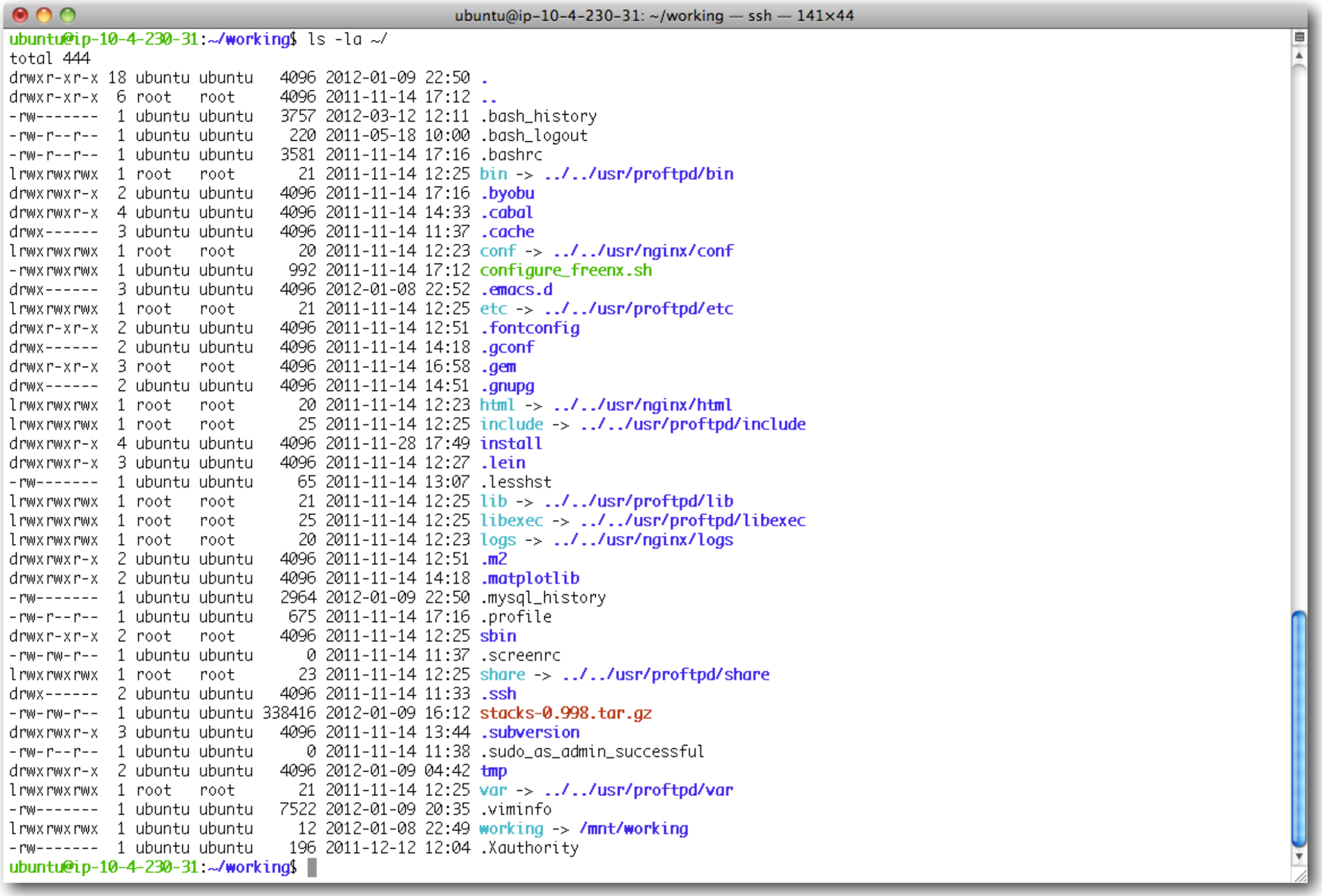


What computers can run Unix?



- Apple OS X Macs
- Google's Android phones
- Most airplane entertainment systems
- Wireless internet routers

The Terminal Window

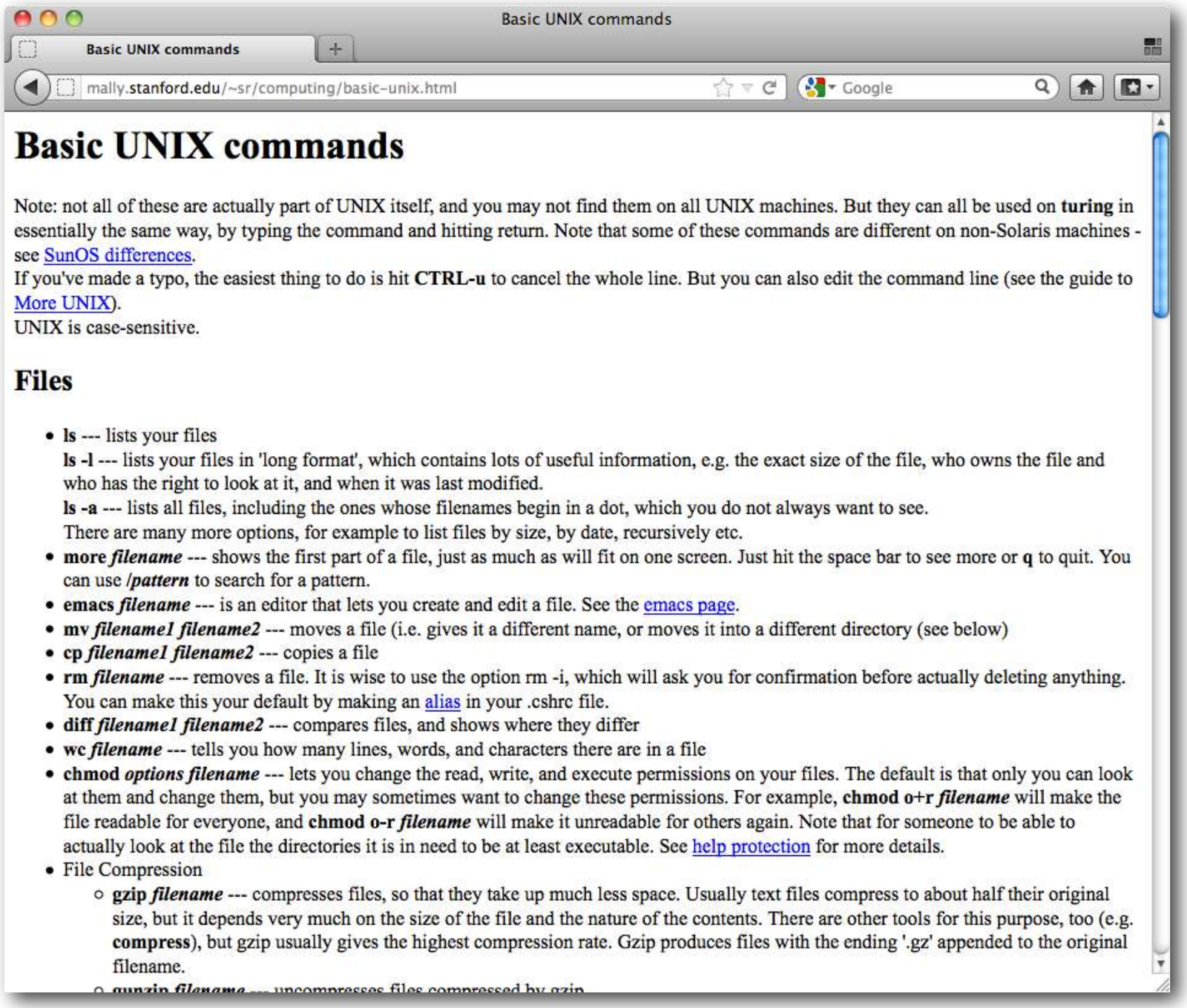


Make it comfortable to work in:

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

Obtain a cheat sheet

google “unix commands”



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$
```

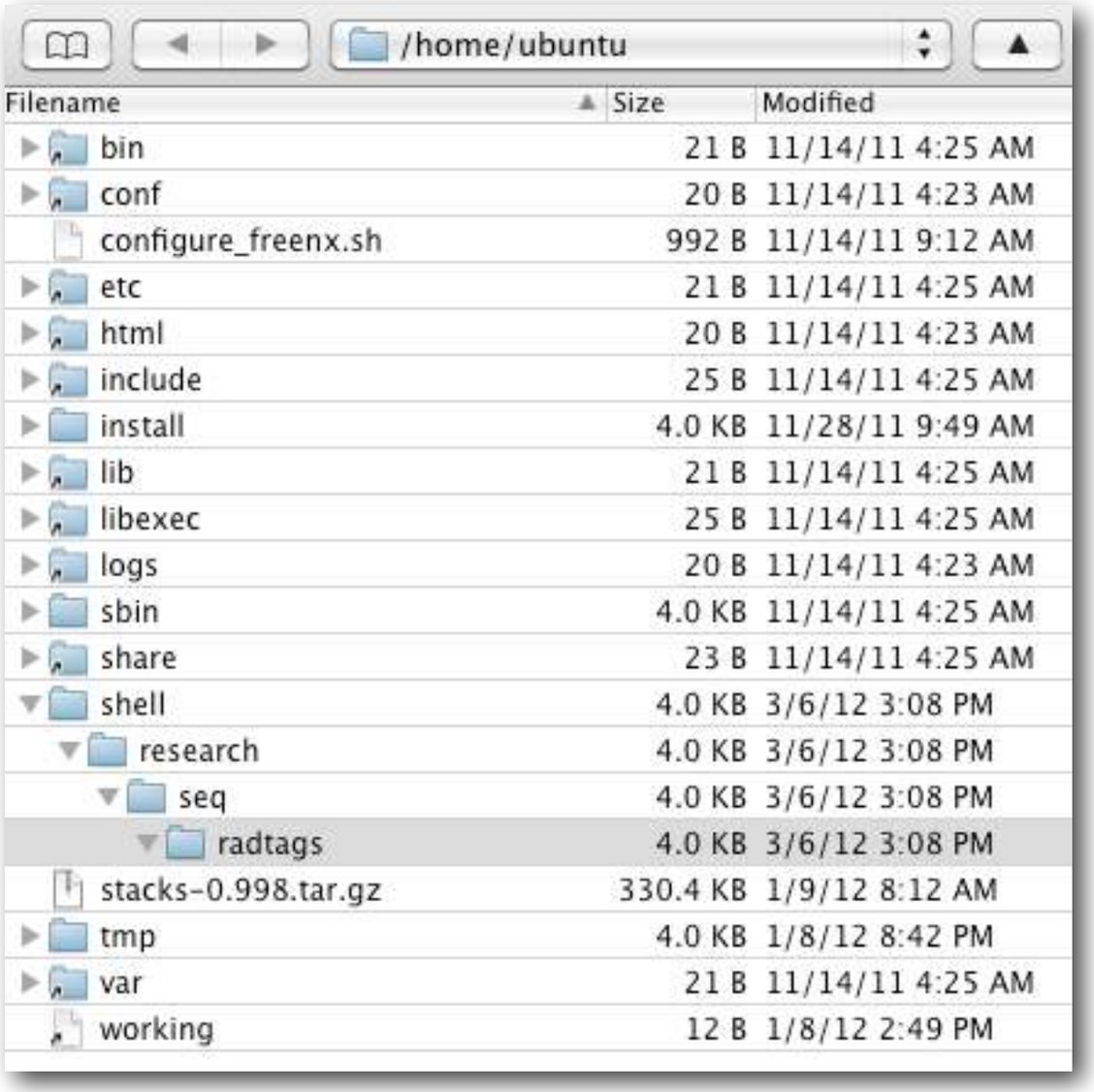
```
% mkdir shell        % cd shell        % ls
```

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

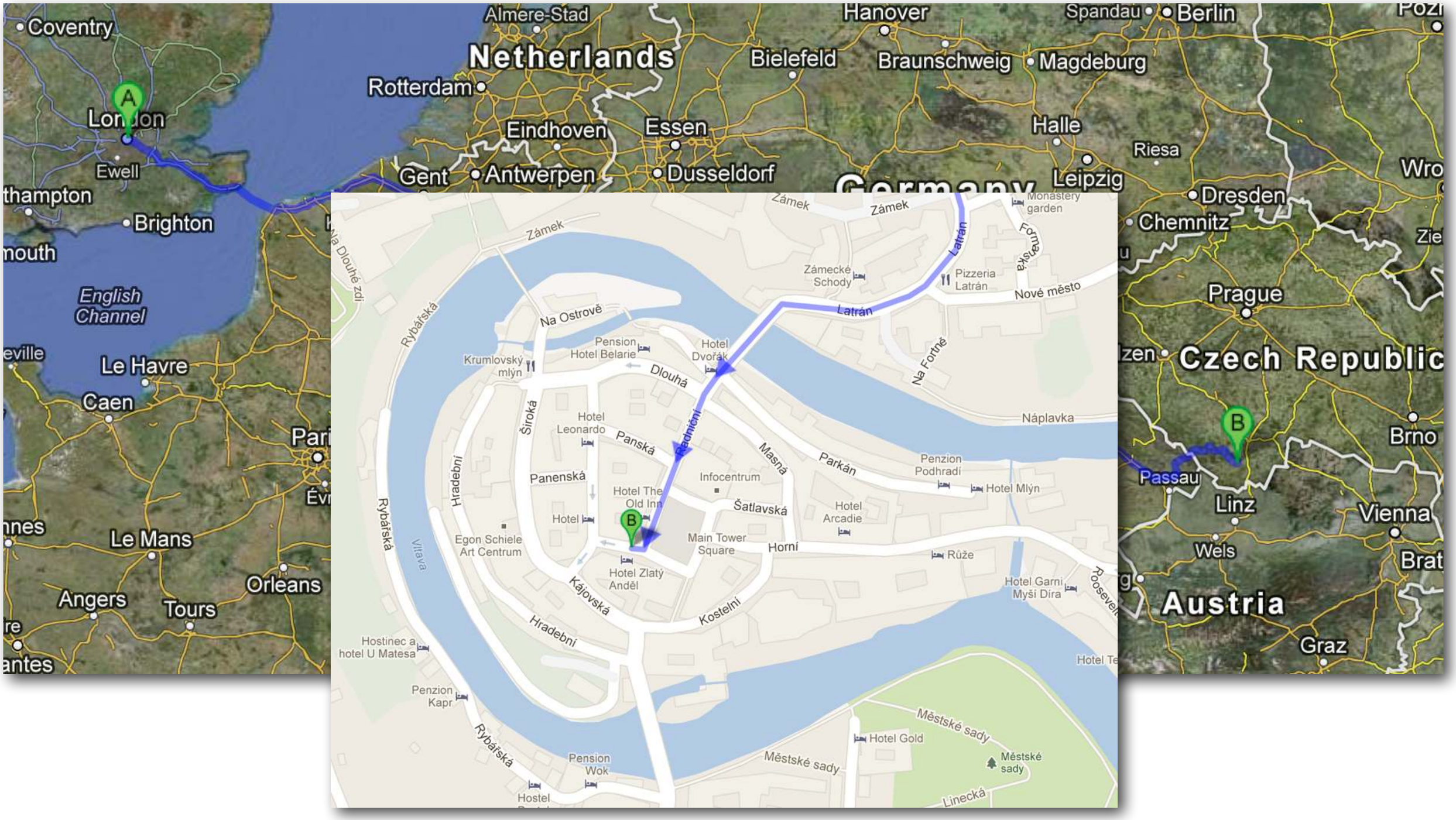


And the `radtags` directory is uniquely identified by its path:

`/home/ubuntu/shell/research/seq/radtags`

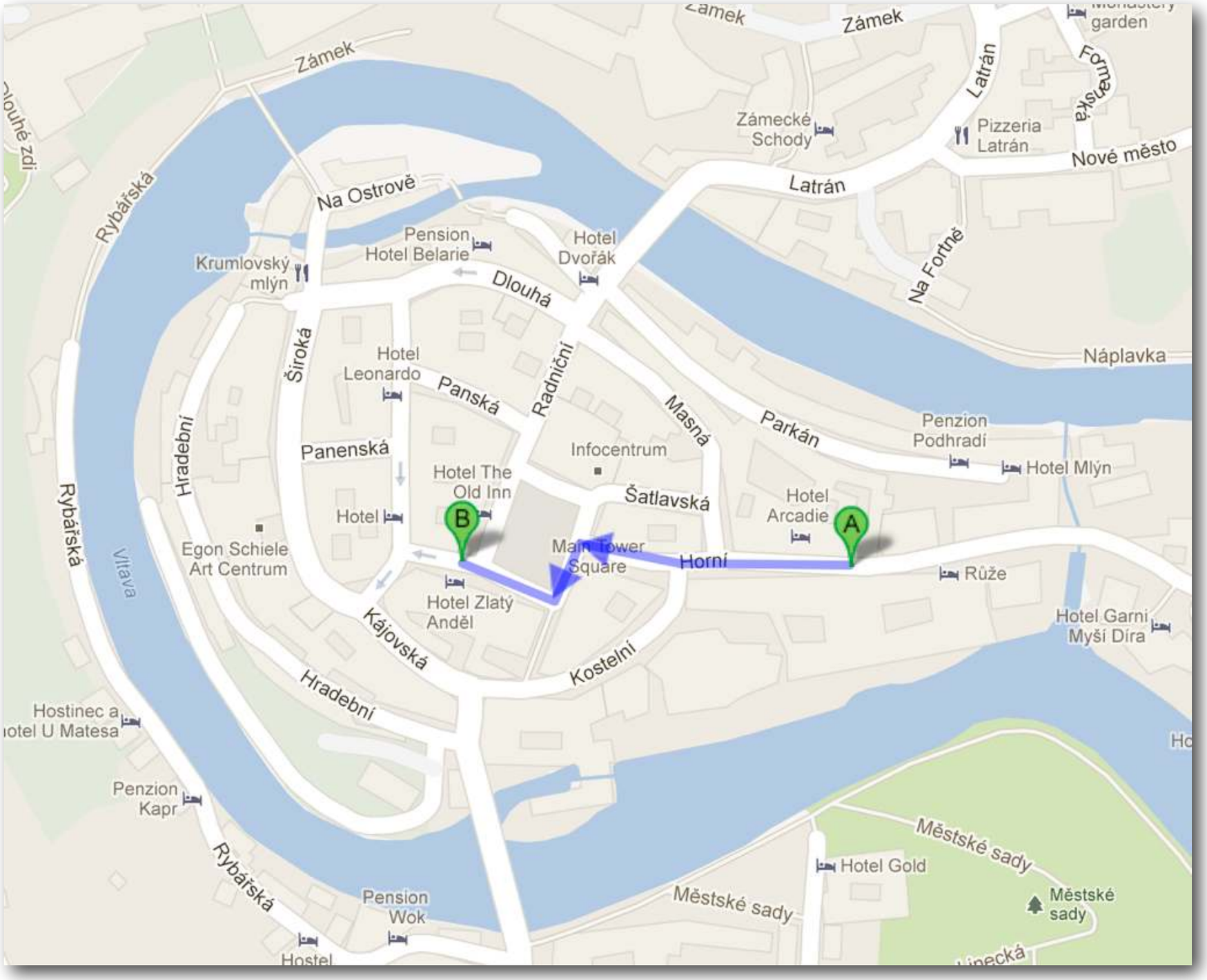
Absolute and relative paths

How do I get to the Hotel Zlaty Andel?

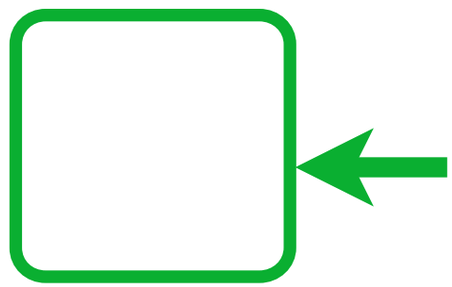


Absolute and relative paths

How do I get to the Hotel Zlaty Andel?



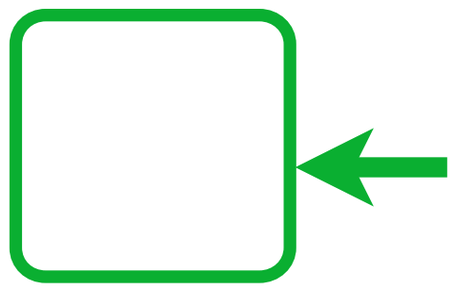
Paths



Absolute Path

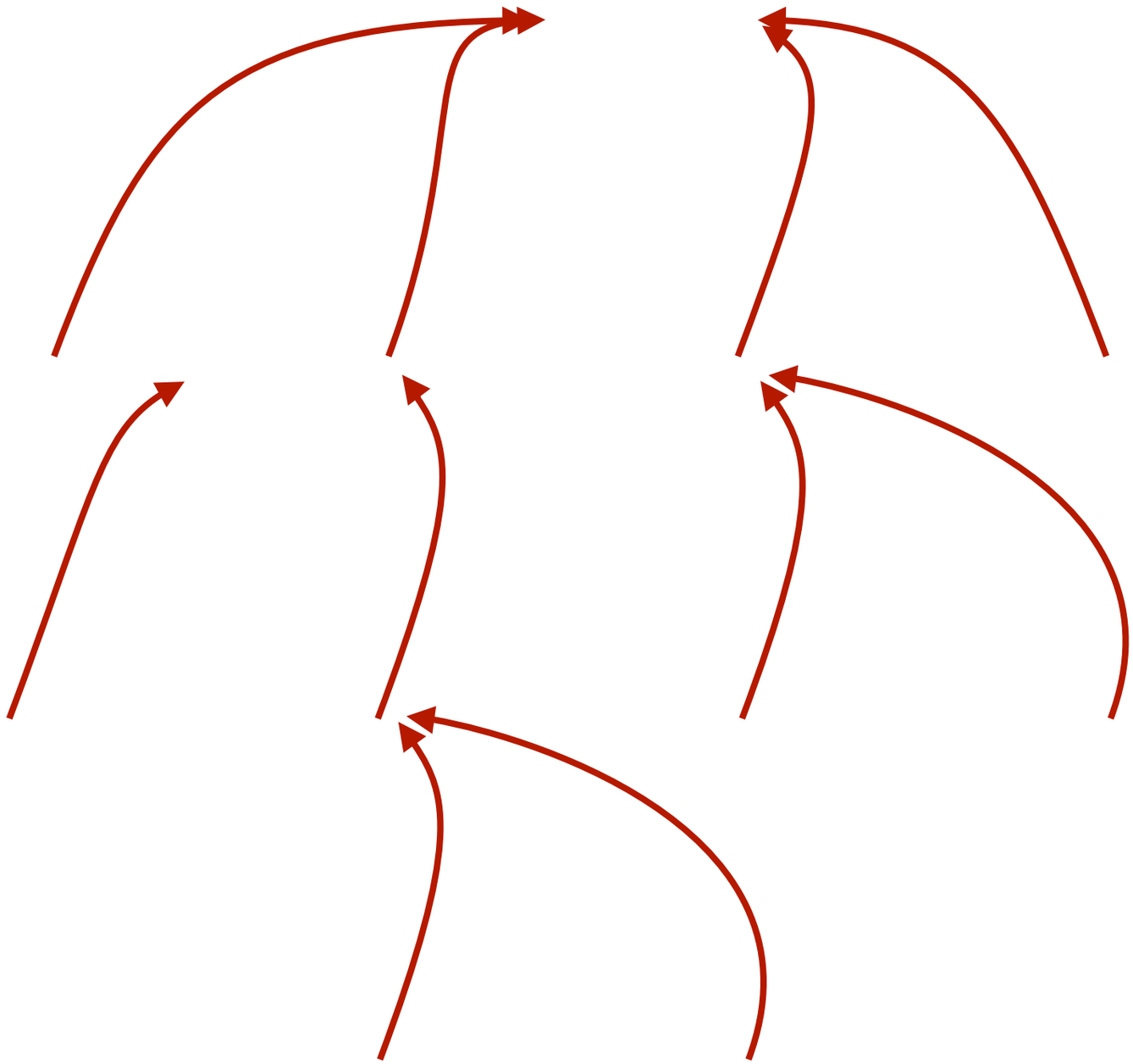
/home/catchen/working

Relative Path?

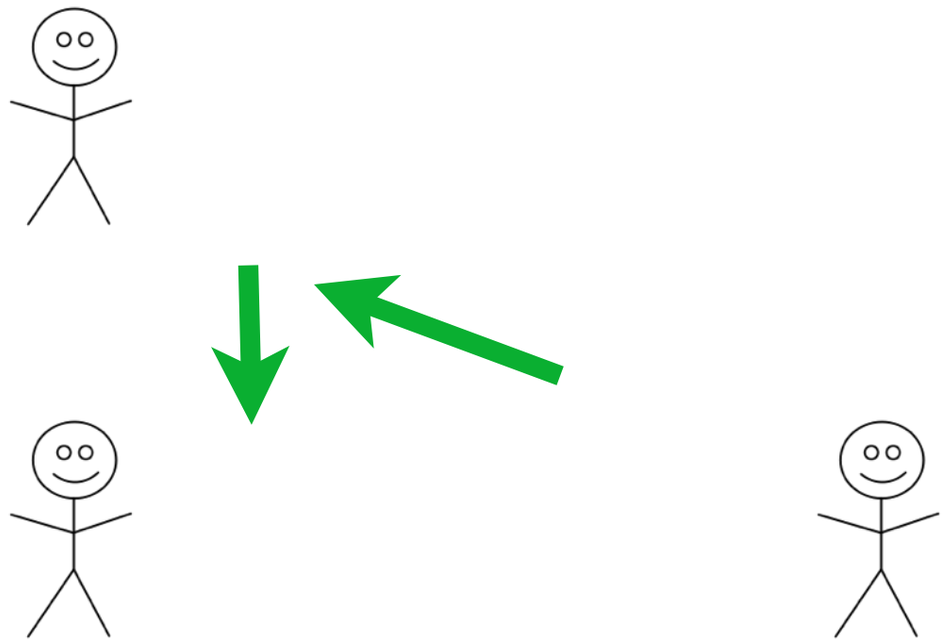


Special files -- '*dot*'

Special files -- '*dot dot*'



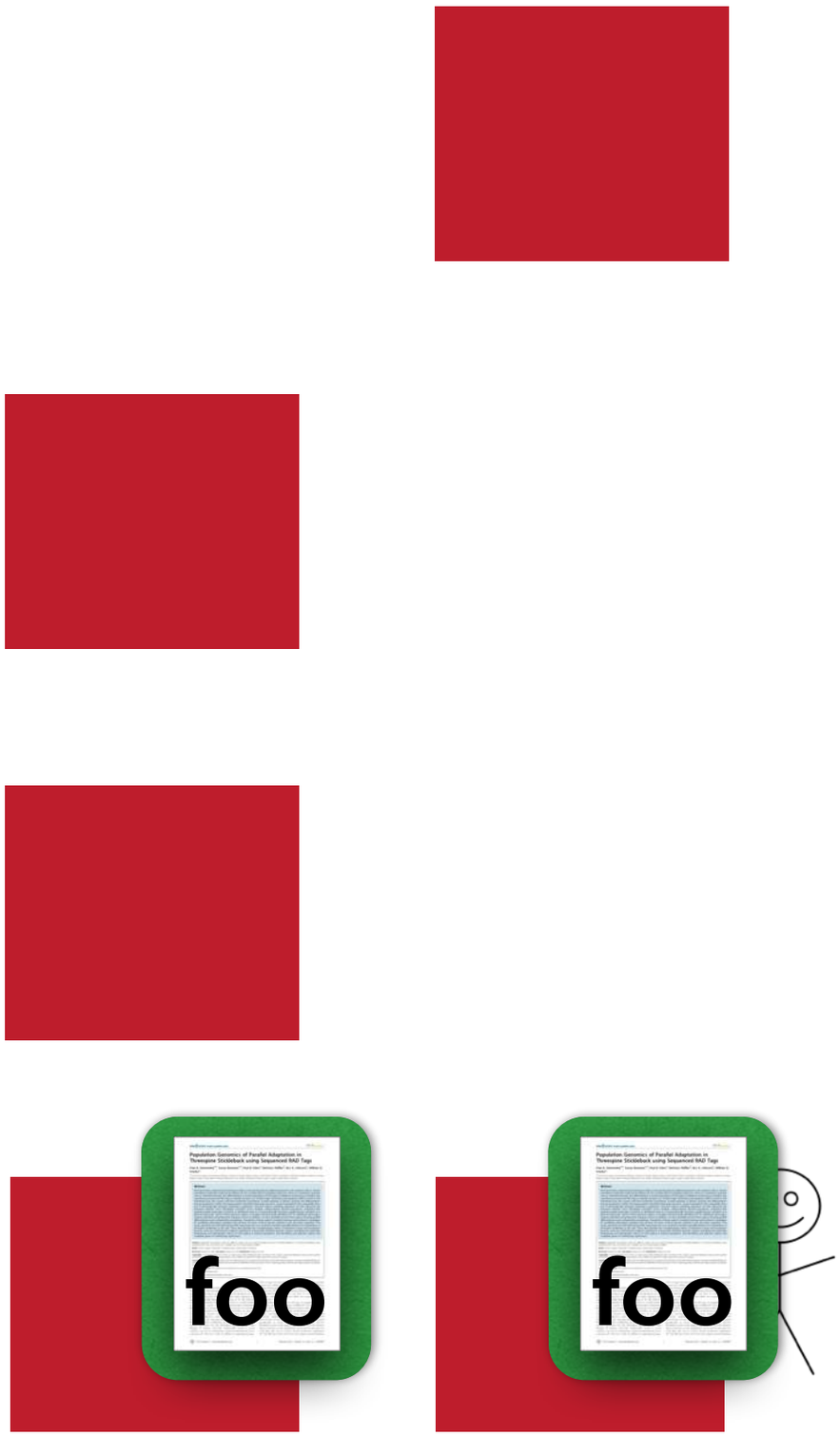
Relative Path



`../working`

Absolute Path: **`/home/catchen/working/foo`**

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



Absolute Path: **`/home/catchen/working/foo`**

Absolute Path: **`/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: ~ — ssh — 133×24

```
ubuntu@ip-10-140-6-74:~$ ls /bin
bash            csh                 getfacl     lsblk       ntfsck          readlink    true
bunzip2         dash                grep        lsmod       ntfscluster     rm          ulockmgr_server
busybox         date                gunzip      mkdir       ntfscmp         rmdir       umount
bzcat           dbus-cleanup-sockets gzexe      mknod       ntfsdecrypt     rnano       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    ntfsinfo        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          zcmp
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               lowntfs-3g  ntfs-3g.usermap pwd         tempfile    zsh4
cpio            fusermount          ls          ntfscat     rbash           touch
ubuntu@ip-10-140-6-74:~$
```

% ls /bin

Relative and absolute paths

A shortcut to your ‘home’, tilde:

~

Moving through the filesystem:

cd

Knowing where you are:

pwd

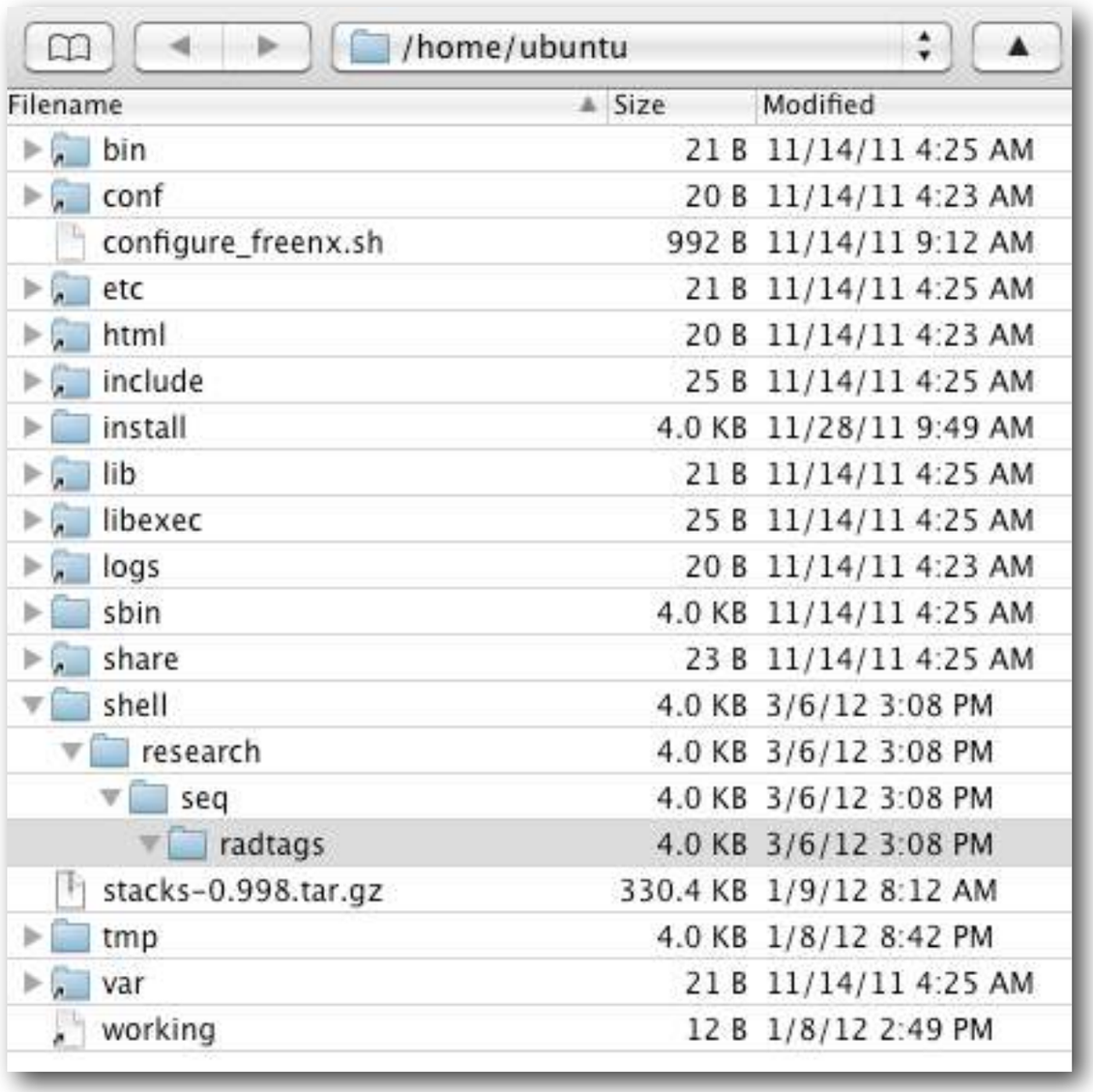
```
% ls ~/          % cd ~/          % cd          % pwd
```

Relative and absolute paths

`/home/ubuntu/shell/research/seq/radtags`

Create a series of directories under shell:

```
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$
```




```
% 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'

```
root@ubuntu:~# cd /etc/
root@ubuntu:/etc# ls -la
total 788
drwxr-xr-x 82 root root    4096 2012-06-25 01:56 .
drwxr-xr-x 23 root root    4096 2012-04-18 09:17 ..
-rw-r--r--  1 root root    2981 2012-04-16 10:09 adduser.conf
drwxr-xr-x  2 root root    4096 2012-04-16 10:23 alternatives
drwxr-xr-x  3 root root    4096 2012-04-16 10:19 apm
drwxr-xr-x  3 root root    4096 2012-04-16 10:21 apparmor
drwxr-xr-x  8 root root    4096 2012-04-16 12:09 apparmor.d
drwxr-xr-x  3 root root    4096 2012-04-16 12:09 apport
drwxr-xr-x  6 root root    4096 2012-04-16 12:08 apt
-rw-r-----  1 root daemon   144 2011-05-16 03:32 at.deny
-rw-r--r--  1 root root    1975 2011-05-18 03:00 bash.bashrc
-rw-r--r--  1 root root   58753 2011-10-04 13:53 bash_completion
drwxr-xr-x  3 root root   12288 2012-04-16 12:09 bash_completion.d
-rw-r--r--  1 root root     344 2011-10-04 12:56 bindresvport.blacklist
-rw-r--r--  1 root root     321 2011-08-09 09:16 blkid.conf
lrwxrwxrwx  1 root root      15 2011-08-09 09:16 blkid.tab -> /dev/.blkid.tab
drwxr-xr-x  2 root root    4096 2012-04-16 10:23 byobu
drwxr-xr-x  3 root root    4096 2012-04-16 10:18 ca-certificates
-rw-r--r--  1 root root    7014 2012-04-16 10:21 ca-certificates.conf
drwxr-xr-x  2 root root    4096 2012-04-16 10:22 calendar
drwxr-s---  2 root dip     4096 2012-04-16 10:22 chatscripts
drwxr-xr-x  4 root root    4096 2012-04-16 12:10 cloud
drwxr-xr-x  2 root root    4096 2012-04-16 10:10 console-setup
drwxr-xr-x  2 root root    4096 2012-04-16 10:10 cron.d
drwxr-xr-x  2 root root    4096 2012-04-16 12:09 cron.daily
drwxr-xr-x  2 root root    4096 2012-04-16 10:10 cron.hourly
drwxr-xr-x  2 root root    4096 2012-04-16 10:10 cron.monthly
-rw-r--r--  1 root root     724 2011-09-19 17:04 crontab
drwxr-xr-x  2 root root    4096 2012-04-16 10:22 cron.weekly
drwxr-xr-x  4 root root    4096 2012-04-16 10:21 dbus-1
-rw-r--r--  1 root root    2969 2011-06-23 00:01 debconf.conf
-rw-r--r--  1 root root      11 2011-07-08 10:13 debian_version
drwxr-xr-x  2 root root    4096 2012-04-16 12:09 default
```

```
% cd /etc
% ls c <tab>
```

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

Three variants to `ls`

`ls -l`	`ls -la`	`ls -lh`
provides a *long* listing	includes *all* files, even hidden files	displays file sizes in *human* readable numbers

Four ways to view a text file

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

Explore the file hierarchy

1. Move to the directory /etc
    - What is the first line of the files '`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?
        - 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:
/unix_data/batch_1.genotypes_1.loc.gz

FASTQ file:
/unix_data/s_1_sequence.txt.gz

Tar Archive:
/unix_data/samples.tar.gz

What is a tar archive?

tar = tape archive

Compress / Decompress

`gzip` / `gunzip`

batch_1.genotypes_1.loc.gz

s_1_sequence.txt.gz

Gzipped Tar archive

`tar xvfz`

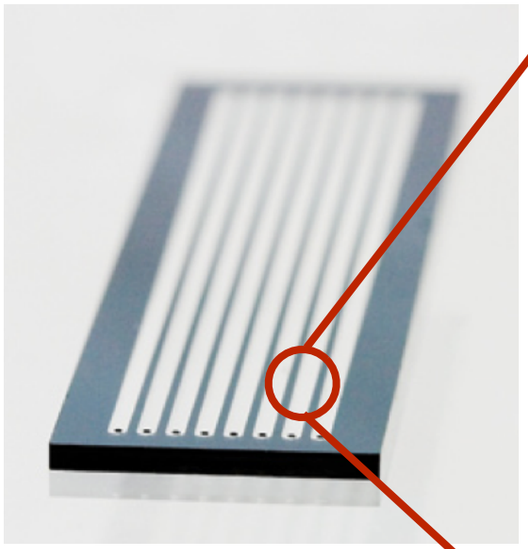
samples.tar.gz

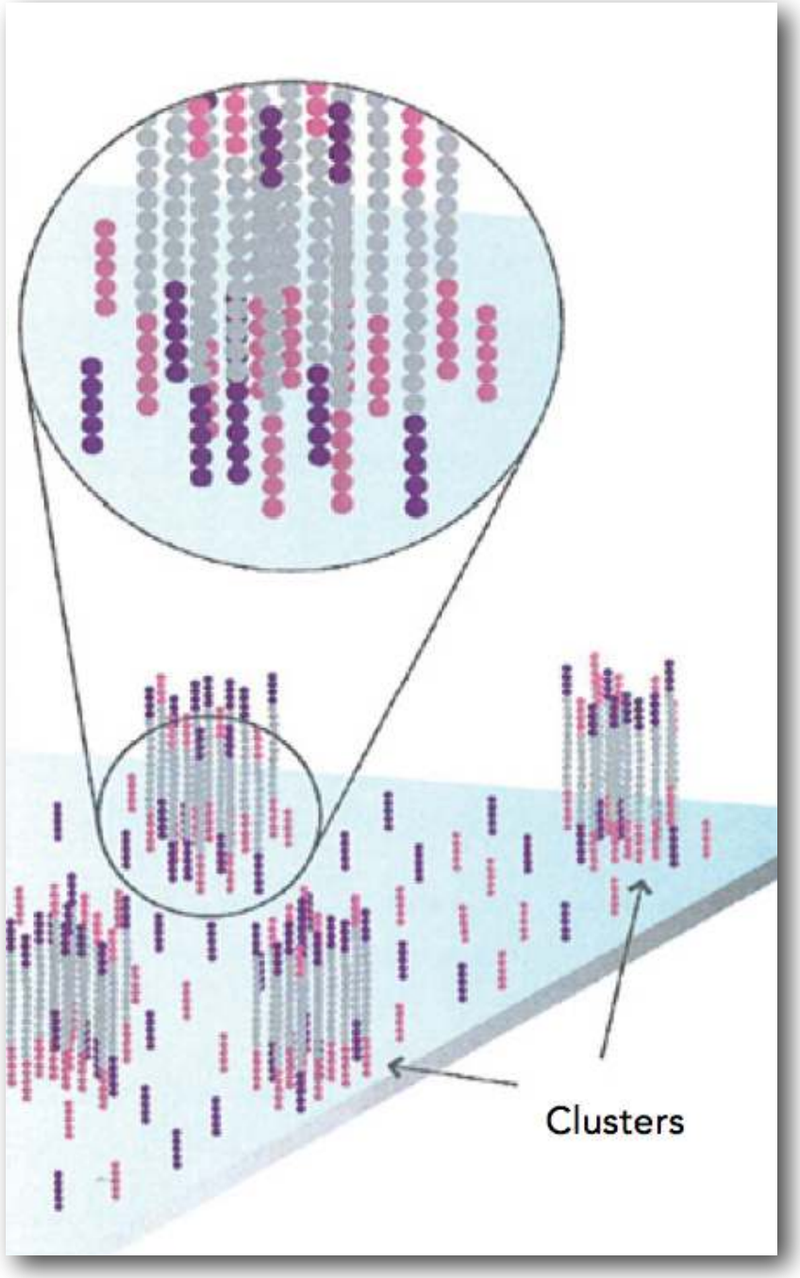
Tar archive

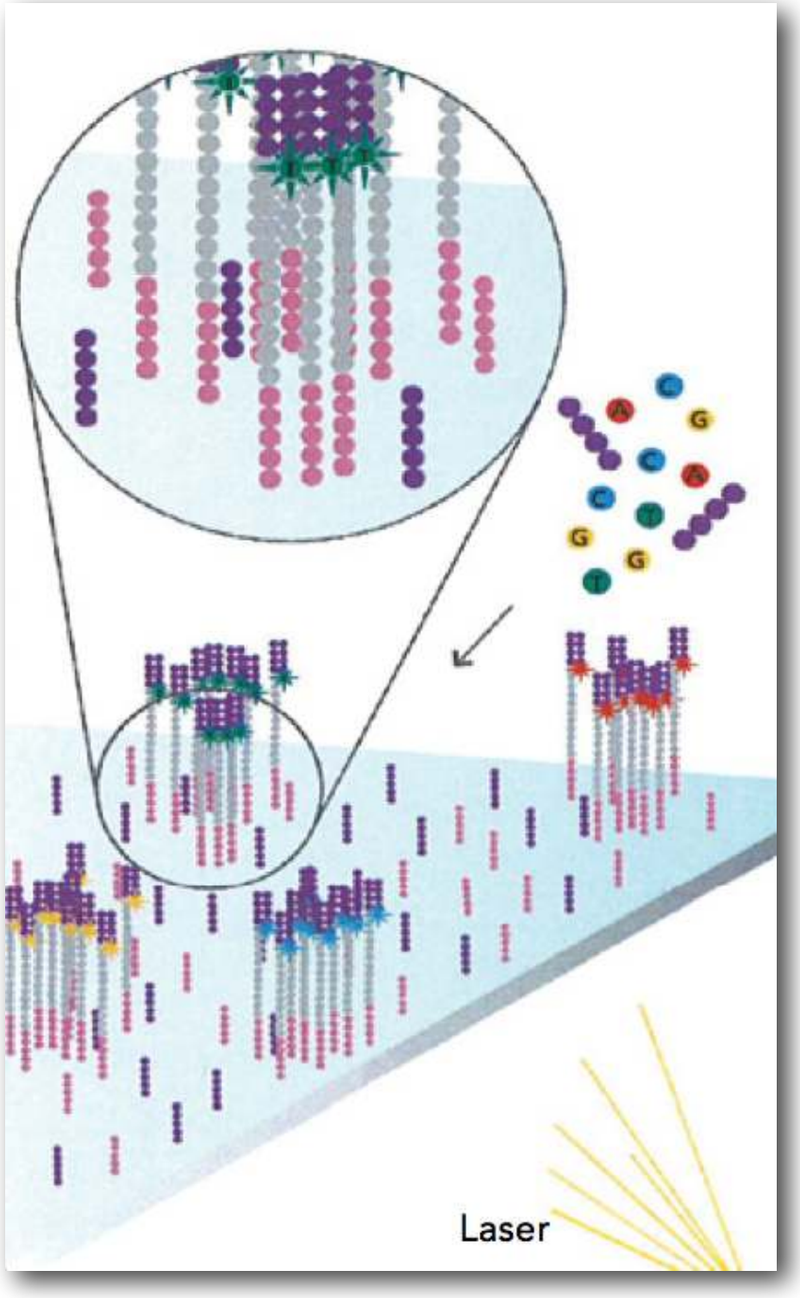
`tar xvf`

samples.tar

Sequencing on Illumina's Flow cell

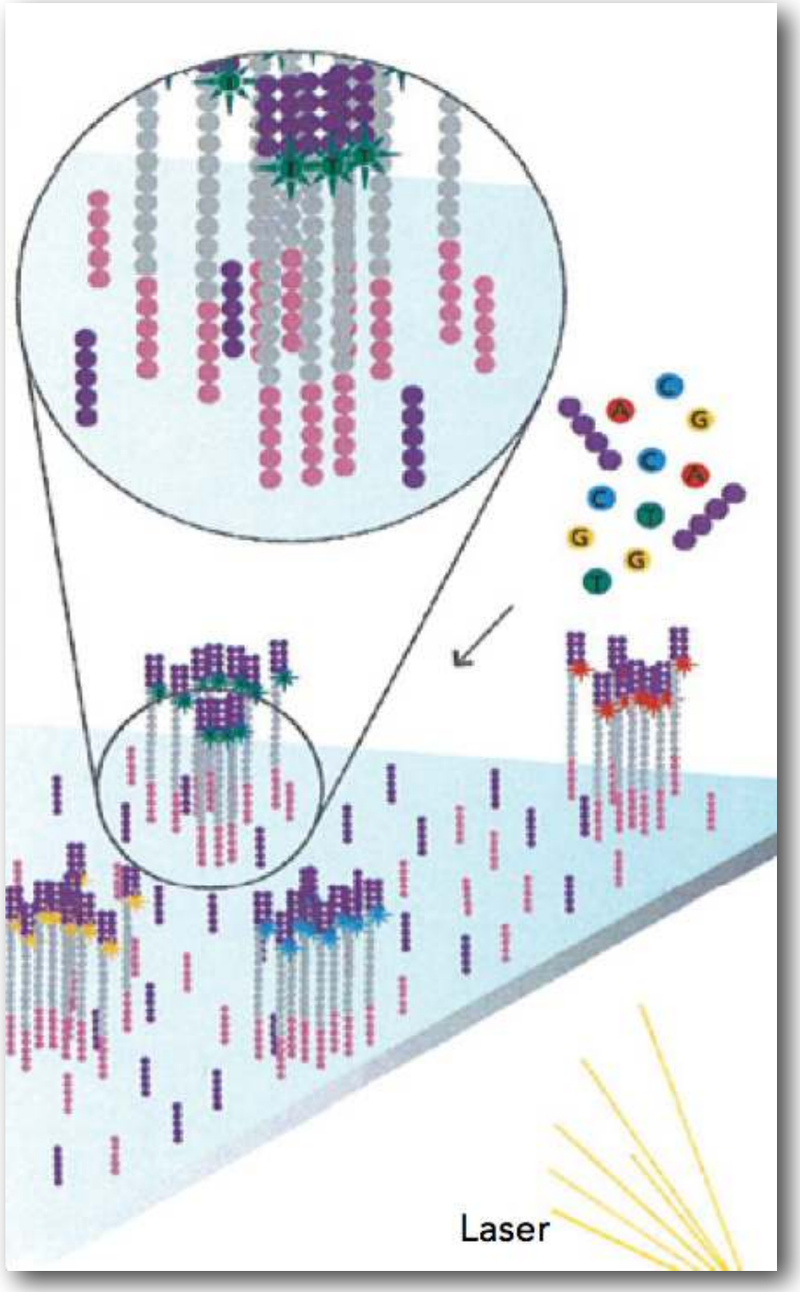


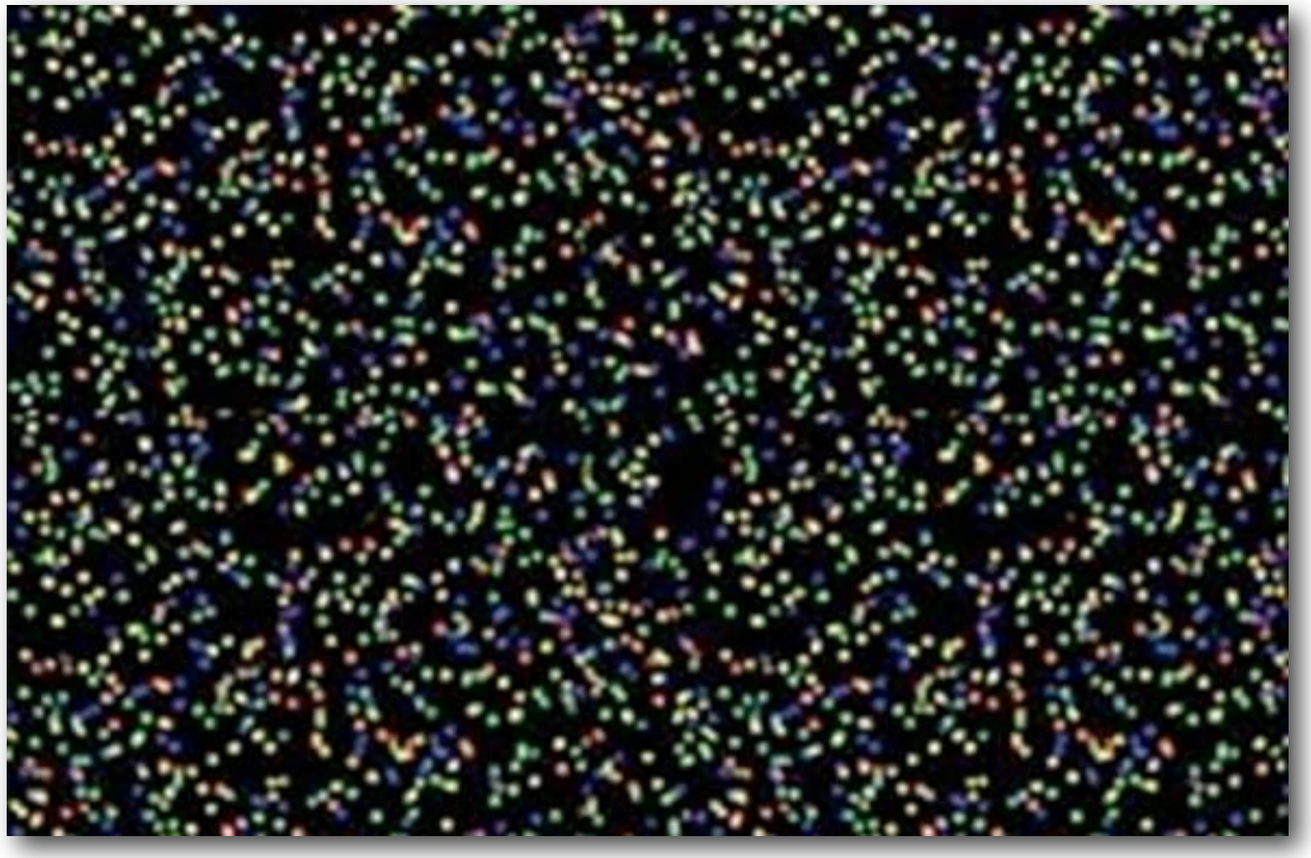




Source: Illumina Publication 770-2007-002 01May07

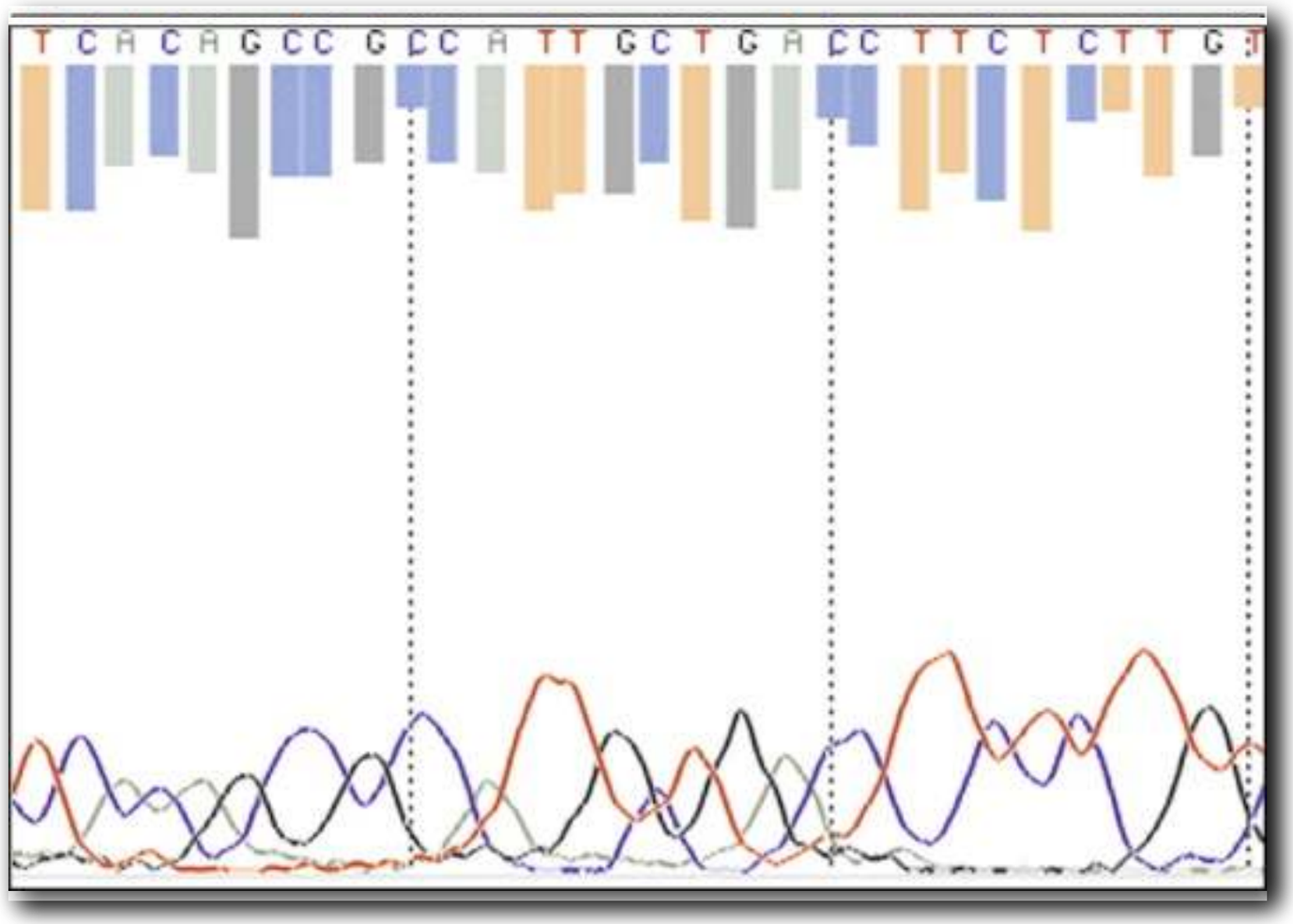
Sequencing on Illumina's Flow cell, ctd.





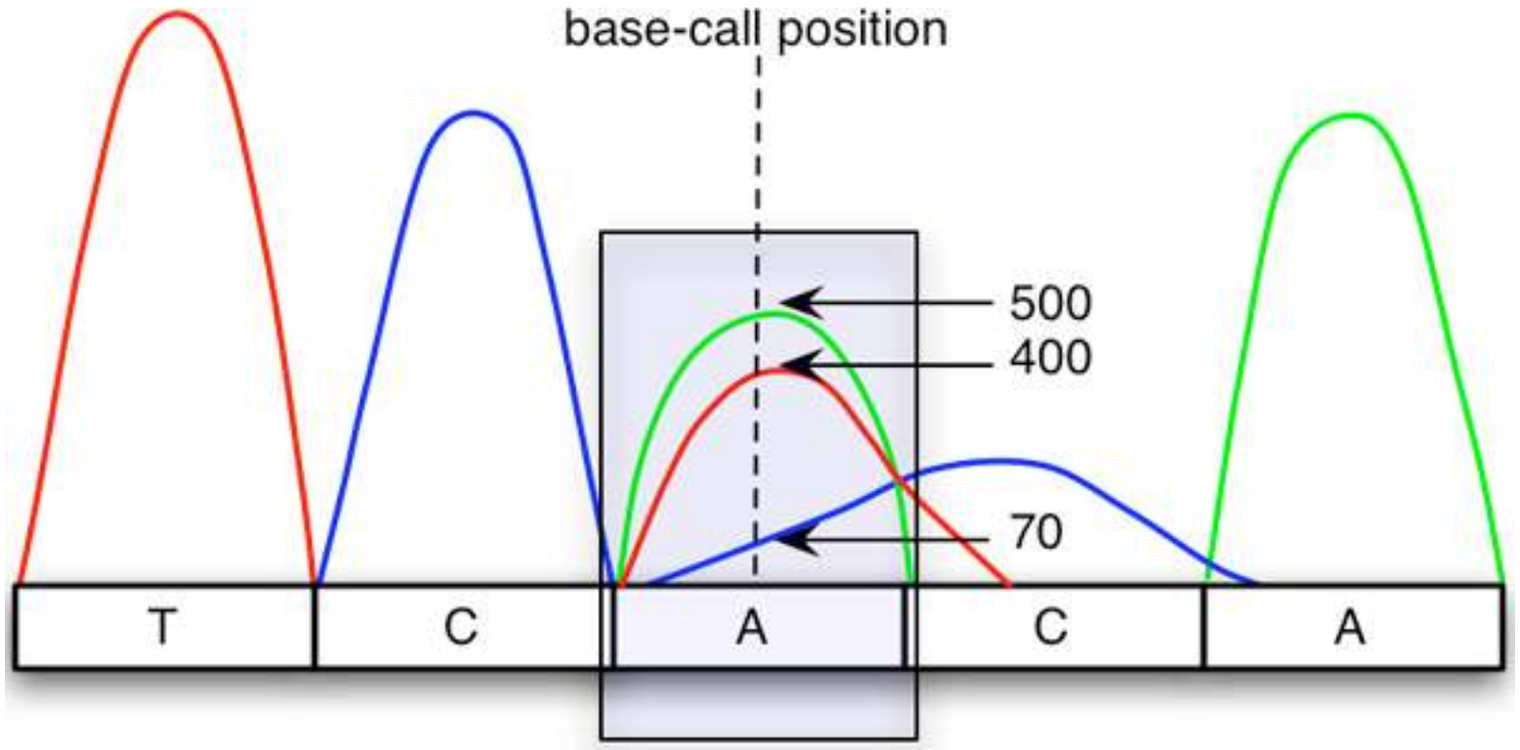
Source: Illumina Publication 770-2007-002 01May07

Sequencing on Illumina's Flow cell, ctd.

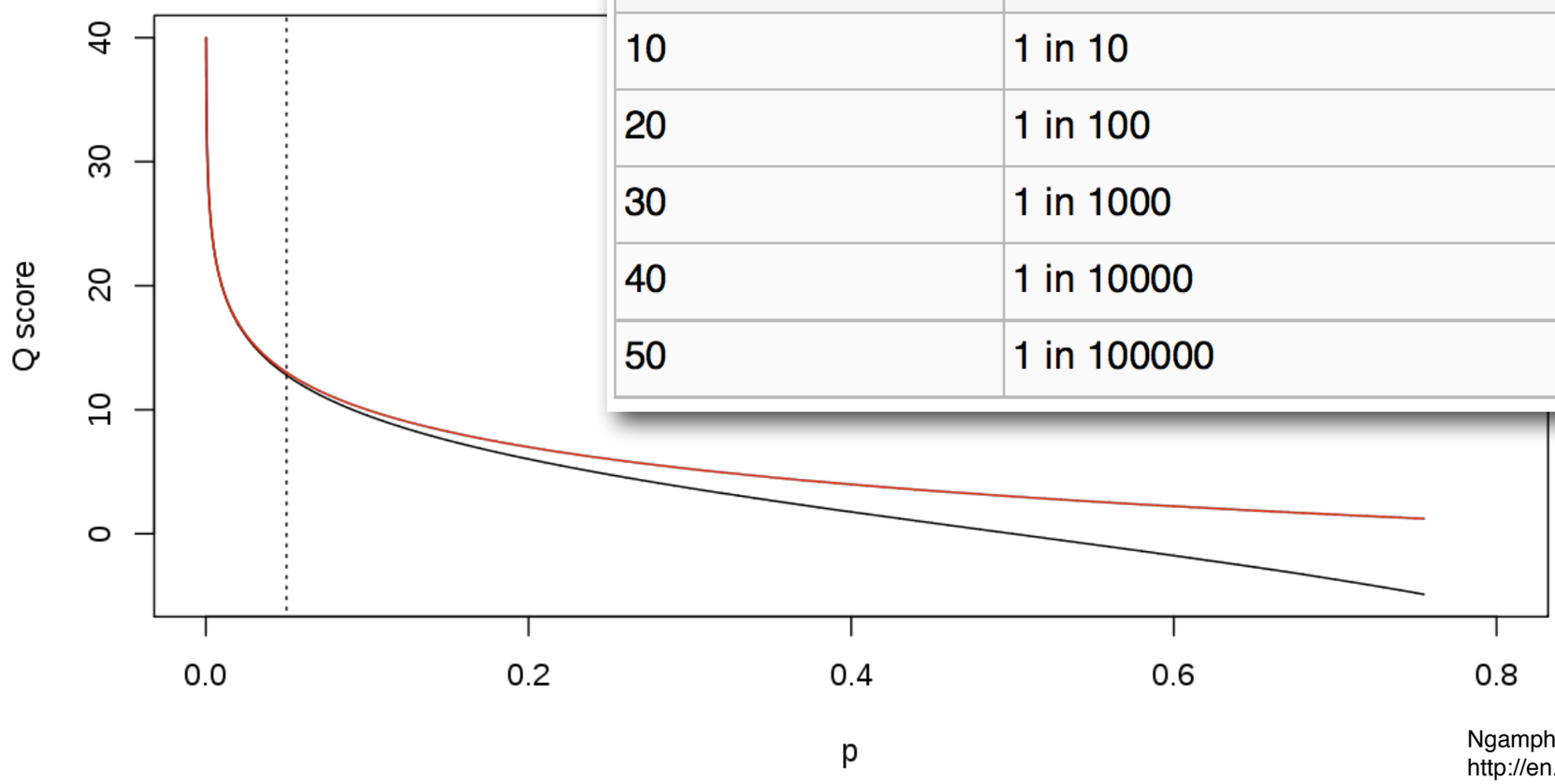


Source: Kim, et al. doi:10.1016/j.pnsc.2007.12.011

Phred Quality Score



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%



Ngamphiw, et al. doi:10.1186/1471-2105-9-S12-S9
http://en.wikipedia.org/wiki/FASTQ_format

The FASTQ File Format

FASTA

```
>chromosome7
TTTGTCTGCAGGGGGACACGTCAAAGTCAAACGCAGGCAAGTTTGTGTTTATGTCCAGTGGATCTTTTGATTTT
ACATACTGCAGGGTCAGGAGGATTATCTCCTCTGCAAGGTAACGCCTGCTGTAACCGTTGTTCTTCATCCTTTT
CCTAACTGCAGGGCTGTCTTGTCAGGTCTGACAAGACATATGCAGGGCTCAATTTGAGATAATTGCTCAATATA
```

FASTQ

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

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

http://4toc.com/fb/FBHelp/gfx/AppF_ASCIITable.png

ASCII Code

0	<NUL>	32	<SPC>	64	@	96	`	128	Ä	160	†	192	¿	224	‡
1	<SOH>	33	!	65	A	97	a	129	Å	161	°	193	¡	225	·
2	<STX>	34	"	66	B	98	b	130	Ç	162	¢	194	¬	226	‚
3	<ETX>	35	#	67	C	99	c	131	É	163	£	195	√	227	„
4	<EOT>	36	$	68	D	100	d	132	Ñ	164	§	196	ƒ	228	‰
5	<ENQ>	37	%	69	E	101	e	133	Ö	165	•	197	≈	229	Â
6	<ACK>	38	&	70	F	102	f	134	Ü	166	¶	198	Δ	230	Ê
7	<BEL>	39	'	71	G	103	g	135	á	167	ß	199	«	231	Á
8	<BS>	40	(	72	H	104	h	136	à	168	®	200	»	232	Ë
9	<TAB>	41	)	73	I	105	i	137	â	169	©	201	…	233	È
10	<LF>	42	*	74	J	106	j	138	ä	170	™	202		234	Í
11	<VT>	43	+	75	K	107	k	139	ã	171	´	203	À	235	Î
12	<FF>	44	,	76	L	108	l	140	å	172	¨	204	Ã	236	Ï
13	<CR>	45	-	77	M	109	m	141	ç	173	≠	205	Õ	237	Ì
14	<SO>	46	.	78	N	110	n	142	é	174	Æ	206	Œ	238	Ó
15	<SI>	47	/	79	O	111	o	143	è	175	Ø	207	œ	239	Ô
16	<DLE>	48	0	80	P	112	p	144	ê	176	∞	208	–	240	
17	<DC1>	49	1	81	Q	113	q	145	ë	177	±	209	—	241	Ò
18	<DC2>	50	2	82	R	114	r	146	í	178	≤	210	“	242	Ú
19	<DC3>	51	3	83	S	115	s	147	ì	179	≥	211	”	243	Û
20	<DC4>	52	4	84	T	116	t	148	î	180	¥	212	‘	244	Ù
21	<NAK>	53	5	85	U	117	u	149	ï	181	µ	213	’	245	ı
22	<SYN	54	6	86	V	118	v	150	ñ	182	∂	214	÷	246	ˆ
23	<ETB>	55	7	87	W	119	w	151	ó	183	∑	215	◊	247	˜
24	<CAN>	56	8	88	X	120	x	152	ò	184	∏	216	ÿ	248	¯
25	<EM>	57	9	89	Y	121	y	153	ô	185	π	217	Ÿ	249	˘
26	<SUB>	58	:	90	Z	122	z	154	ö	186	∫	218	/	250	˙
27	<ESC>	59	;	91	[	123	{	155	õ	187	ª	219	€	251	°
28	<FS>	60	<	92	\	124	|	156	ú	188	º	220	‹	252	¸
29	<GS>	61	=	93	]	125	}	157	ù	189	Ω	221	›	253	˝
30	<RS>	62	>	94	^	126	~	158	û	190	æ	222	fi	254	˛
31	<US>	63	?	95	_	127	<DEL>	159	ü	191	ø	223	fl	255	ˇ

8 bits = 2^8 combinations = 256

0	1	1	1	1	0	0	1

$0x2^7+1x2^6+1x2^5+1x2^4+1x2^3+0x2^2+0x2^1+1x2^0 = 121 =$ **y**

The FASTQ File Format, ctd

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

Quality Scores

```
  SSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSS.....................................
  .........................XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX.....................
  ............................IIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIII.....................
  ..............................JJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJJ.....................
  LLLLLLLLLLLLLLLLLLLLLLLLLLLLLLLLLLLLLLLLL.....................................
  !"#$%&'()*+,-./0123456789:;<=>?@ABCDEFGHIJKLMNOPQRSTUVWXYZ[\]^_`abcdefghijklmnopqrstuvwxyz{|}~
  |                          |    |        |                              |                     |
 33                         59   64       73                            104                   126

 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**

http://en.wikipedia.org/wiki/FASTQ_format

The FASTQ File Format, ctd

```
@HWI-ST0747:162:C03AJACXX:3:1108:19763:106771 1:N:0:
TTTGTCTGCAGGGGGACACGTCAAAGTCAAACGCAGGCAAGTTTGTGTTTATGTCCAGTGGATCTTTTGATTTT
+
<?@DDDDDHFHHFBB@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%

http://en.wikipedia.org/wiki/Phred_quality_score

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
```

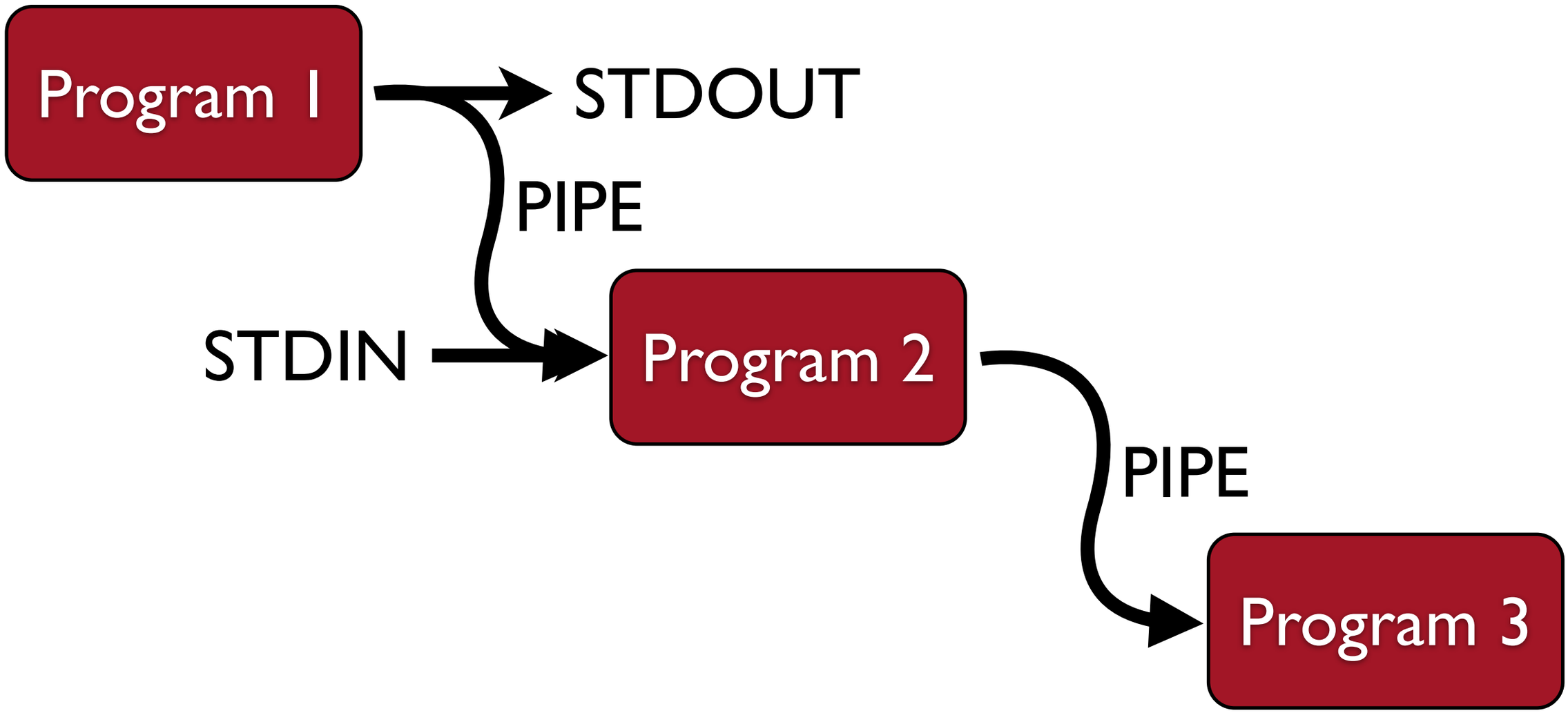
Special Files

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_1_sequence.txt.gz

1. Decompress the file
2. Count the number of raw reads (250,000)
3. Count the number of reads with barcode `CGATA` (19,501)
4. Capture all FASTQ records for `ACCAT` into a file called sample_01.fq (you should get 18352 records, 73408 lines)
5. Determine the count of all barcodes in the file

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

```
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

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"
```

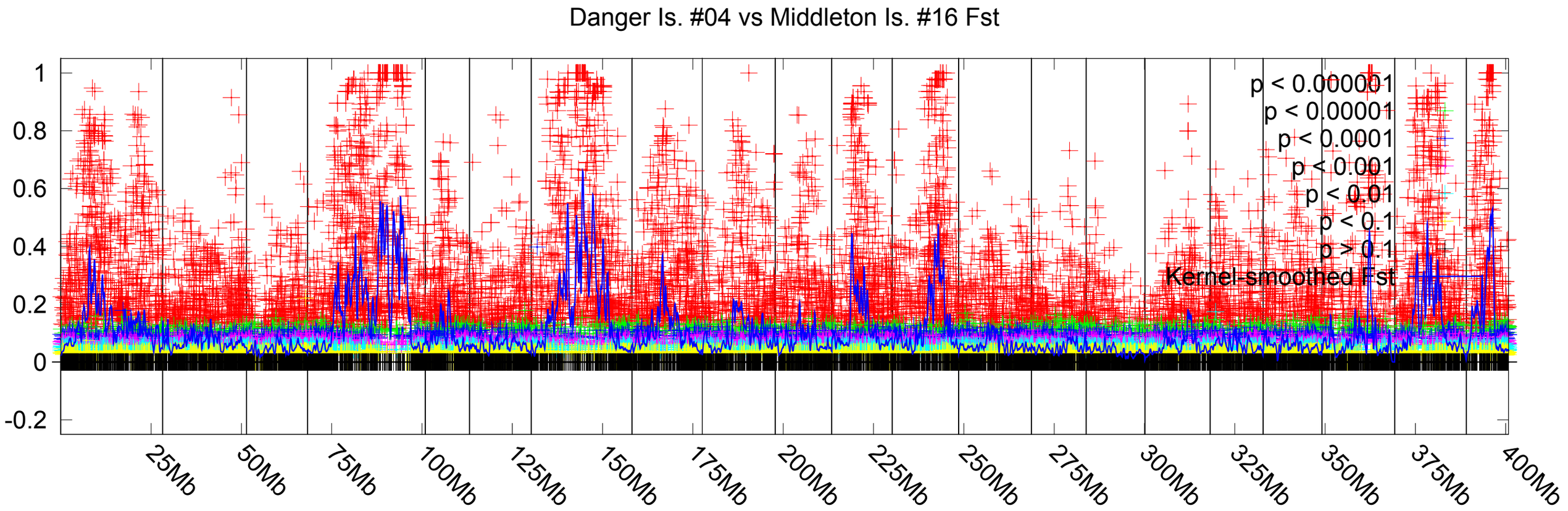
Count raw reads:

```
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
```

Problem Set #1



# Batch ID	Locus ID		Pop 1 ID	Pop 2 ID	Chr	BP	Column	Overall Pi	Fst	Fisher's P	Odds Ratio
2	7894	2	3	groupI	11832	19	0.428182	-0.0076252913	0.191294		0.687192
2	7896	2	3	groupI	11900	83	0.328622	0.1775694587	4.35747e-08		5.44667
2	9611	2	3	groupI	49756	48	0.090426	-0.1127451906	0.00109115		0.0898072
2	9611	2	3	groupI	49765	57	0.0132887	-0.1522407447	0.518395		0.493113
2	9611	2	3	groupI	49766	58	0.0133776	-0.1563725438	0.520033		0.501401
...											
2	7083	2	3	groupIX	20172984	57	0.0480227	-0.1249752728	0.0230205		0.152263
2	7083	2	3	groupIX	20173004	77	0.0181808	-0.1405892833	0.27288		0.355556
2	7083	2	3	groupIX	20173016	89	0.0424072	-0.1243625038	0.0442229		0.17037
2	7083	2	3	groupIX	20173018	91	0.148072	0.2305903127	2.24832e-07		13.2936
2	7084	2	3	groupIX	20172960	29	0.388239	-0.0102170169	0.307185		0.758001
2	7084	2	3	groupIX	20172995	64	0.368259	-0.0130602425	0.233889		0.70084
2	7084	2	3	groupIX	20173013	82	0.0121578	-0.1359057575	0.51309		0.46472