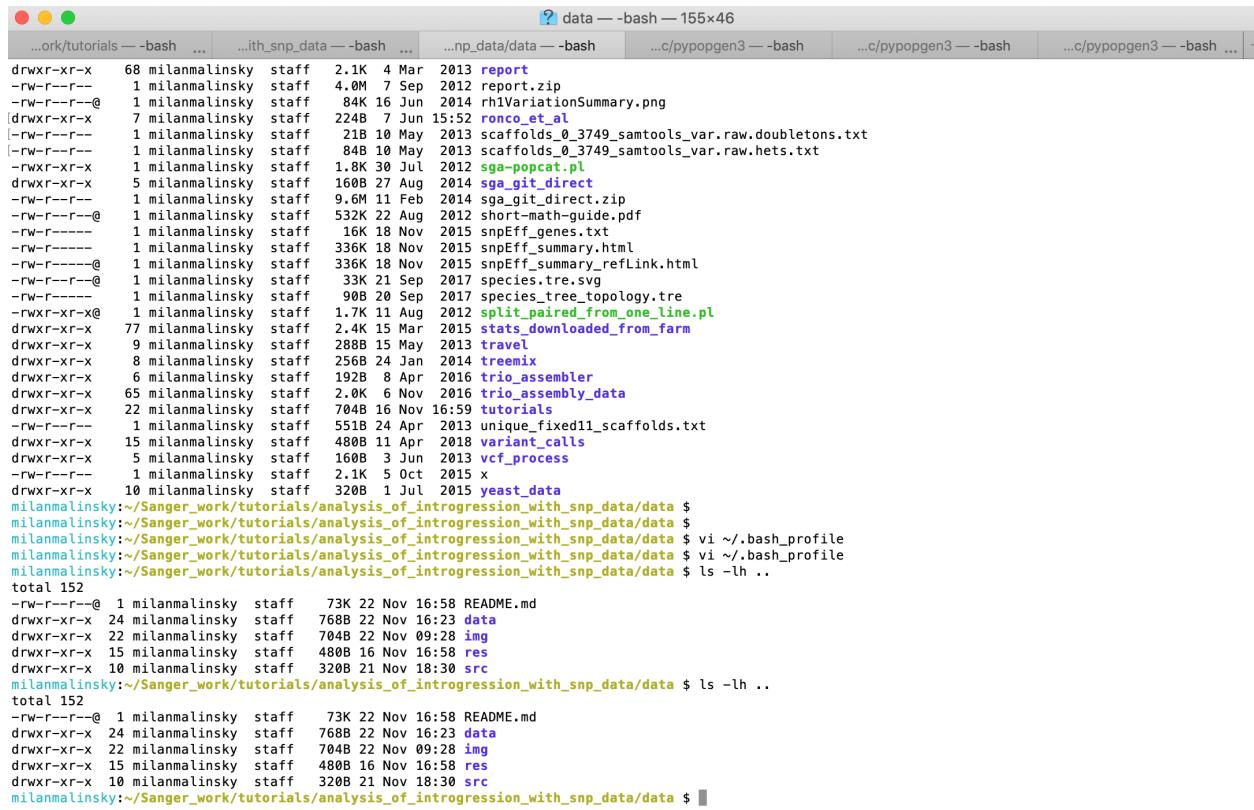


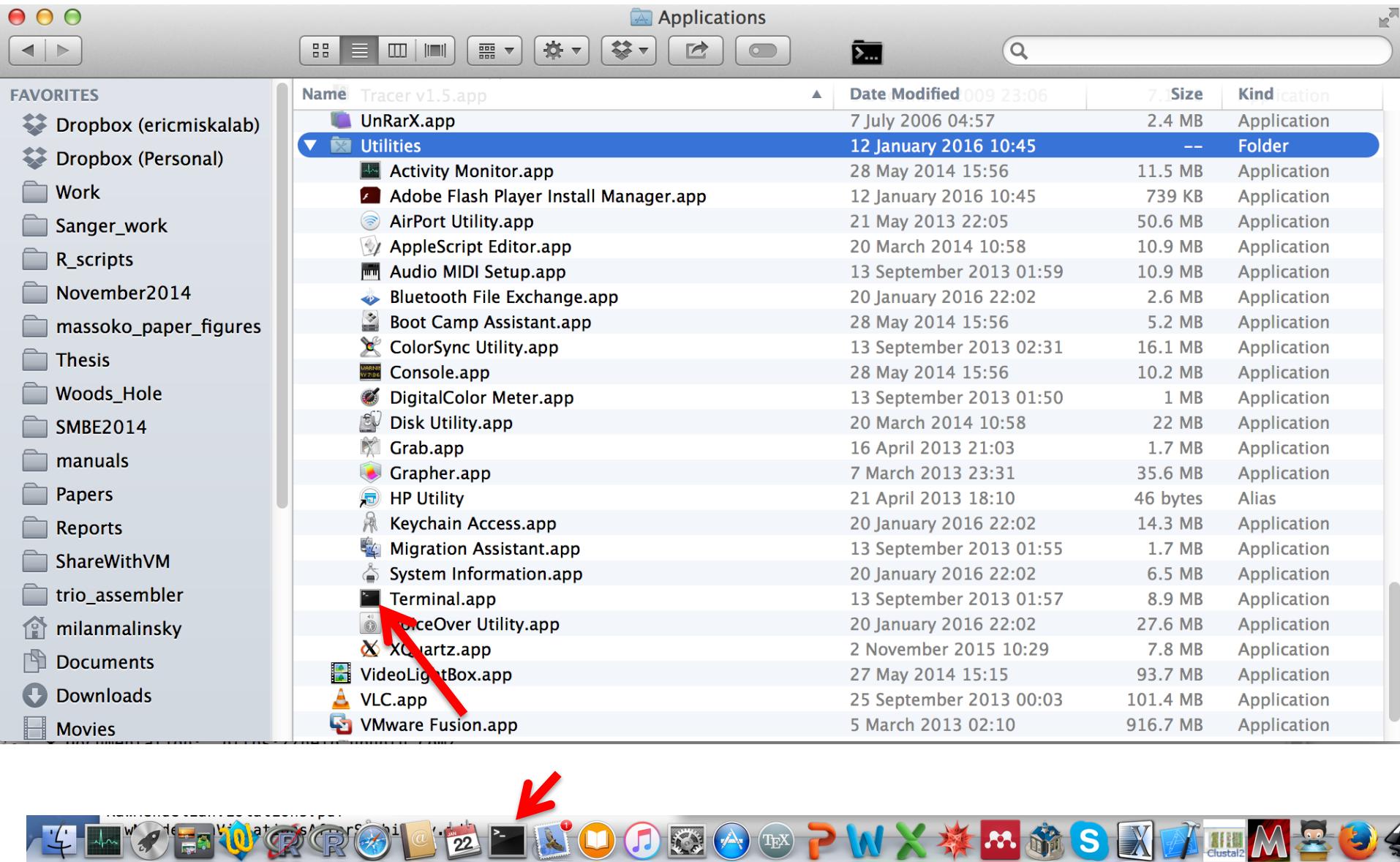
# The UNIX terminal



A screenshot of a terminal window titled "data -- bash -- 155x46". The window shows a file listing and command history. The file listing includes columns for permissions, file size, owner, group, modification date, and name. The command history at the bottom shows several commands related to Sanger work and tutorials.

```
...ork/tutorials -- bash ... ...ith.snp_data -- bash ... ...np_data/data -- bash ...c/pypopgen3 -- bash ...c/pypopgen3 -- bash ... +  
drwxr-xr-x 68 milanmalinsky staff 2.1K 4 Mar 2013 report  
-rw-r--r-- 1 milanmalinsky staff 4.0M 7 Sep 2012 report.zip  
drwxr-xr-x 7 milanmalinsky staff 224B 7 Jun 15:52 ronco_et_al  
-rw-r--r-- 1 milanmalinsky staff 21B 10 May 2013 scaffolds_0_3749_samtools_var.raw.doubletons.txt  
-rw-r--r-- 1 milanmalinsky staff 84B 10 May 2013 scaffolds_0_3749_samtools_var.raw.hets.txt  
-rwxr-xr-x 5 milanmalinsky staff 1.8K 30 Jul 2012 sga-popcat.pl  
drwxr-xr-x 5 milanmalinsky staff 160B 27 Aug 2014 sga_git_direct  
-rw-r--r-- 1 milanmalinsky staff 9.6M 11 Feb 2014 sga_git_direct.zip  
-rw-r--r--@ 1 milanmalinsky staff 532K 22 Aug 2013 short-math-guide.pdf  
-rw-r----- 1 milanmalinsky staff 16K 18 Nov 2015 snpEff_genes.txt  
-rw-r----- 1 milanmalinsky staff 336K 18 Nov 2015 snpEff_summary.html  
-rw-r--r--@ 1 milanmalinsky staff 336K 18 Nov 2015 snpEff_summary_RefLink.html  
-rw-r--r--@ 1 milanmalinsky staff 33K 21 Sep 2017 species.tre.svg  
-rw-r--r-- 1 milanmalinsky staff 90B 20 Sep 2017 species_tree_topology.tre  
-rwxr-xr-x@ 1 milanmalinsky staff 1.7K 11 Aug 2012 split_paired_from_one_line.pl  
drwxr-xr-x 77 milanmalinsky staff 2.4K 15 Mar 2013 stats_downloaded_from_farm  
drwxr-xr-x 9 milanmalinsky staff 288B 15 May 2013 travel  
drwxr-xr-x 8 milanmalinsky staff 256B 24 Jan 2014 treemix  
drwxr-xr-x 6 milanmalinsky staff 192B 8 Apr 2016 trio_assembler  
drwxr-xr-x 65 milanmalinsky staff 2.0K 6 Nov 2016 trio_assembly_data  
drwxr-xr-x 22 milanmalinsky staff 704B 16 Nov 16:59 tutorials  
-rw-r--r-- 1 milanmalinsky staff 551B 24 Apr 2013 unique_fixed11_scaffolds.txt  
drwxr-xr-x 15 milanmalinsky staff 480B 11 Apr 2018 variant_calls  
drwxr-xr-x 5 milanmalinsky staff 160B 3 Jun 2013 vcf_process  
-rw-r--r-- 1 milanmalinsky staff 2.1K 5 Oct 2015 x  
drwxr-xr-x 10 milanmalinsky staff 320B 1 Jul 2015 yeast_data  
milanmalinsky:~/Sanger_work/tutorials/analysis_of_introgression_with.snp_data/data $  
milanmalinsky:~/Sanger_work/tutorials/analysis_of_introgression_with.snp_data/data $  
milanmalinsky:~/Sanger_work/tutorials/analysis_of_introgression_with.snp_data/data $ vi ~/.bash_profile  
milanmalinsky:~/Sanger_work/tutorials/analysis_of_introgression_with.snp_data/data $ vi ~/.bash_profile  
milanmalinsky:~/Sanger_work/tutorials/analysis_of_introgression_with.snp_data/data $ ls -lh ..  
total 152  
-rw-r--r--@ 1 milanmalinsky staff 73K 22 Nov 16:58 README.md  
drwxr-xr-x 24 milanmalinsky staff 768B 22 Nov 16:23 data  
drwxr-xr-x 22 milanmalinsky staff 704B 22 Nov 09:28 img  
drwxr-xr-x 15 milanmalinsky staff 480B 16 Nov 16:58 res  
drwxr-xr-x 10 milanmalinsky staff 320B 21 Nov 18:30 src  
milanmalinsky:~/Sanger_work/tutorials/analysis_of_introgression_with.snp_data/data $ ls -lh ..  
total 152  
-rw-r--r--@ 1 milanmalinsky staff 73K 22 Nov 16:58 README.md  
drwxr-xr-x 24 milanmalinsky staff 768B 22 Nov 16:23 data  
drwxr-xr-x 22 milanmalinsky staff 704B 22 Nov 09:28 img  
drwxr-xr-x 15 milanmalinsky staff 480B 16 Nov 16:58 res  
drwxr-xr-x 10 milanmalinsky staff 320B 21 Nov 18:30 src  
milanmalinsky:~/Sanger_work/tutorials/analysis_of_introgression_with.snp_data/data $
```

# On a Mac computer



# Why use something from the 1960s?

```
data — bash — 155x46  
..np.data/data — bash ..c/pypopgen3 — bash ...c/pypopgen3 — bash ..c/pypopgen3 — bash ..  
drwxr-xr-x 68 milamalinsky staff 2.1K 4 Mar 2013 report  
-rv-r--r--@ 1 milamalinsky staff 4.0M 7 Sep 2012 report.zip  
drwxr-xr-x 7 milamalinsky staff 84K 16 Jun 2014 rh1VariationSummary.png  
-rv-r--r--@ 1 milamalinsky staff 224B 7 Jun 15:52 ronchet.al  
drwxr-xr-x 1 milamalinsky staff 208 14 May 2014 scaffolds_0_3749_samtools_var.raw.doubletons.txt  
-rv-r--r-- 1 milamalinsky staff 849 10 Mar 2013 scaffolds_0_3749_samtools_var.raw.hets.txt  
-rv-r--r-- 1 milamalinsky staff 1.8K 30 Jul 2012 sga-popcat.pl  
drwxr-xr-x 5 milamalinsky staff 160B 27 Aug 2014 sga_git_direct  
-rv-r--r-- 1 milamalinsky staff 9.6M 11 Feb 2014 sga_git_direct.zip  
-rv-r--r--@ 1 milamalinsky staff 536B 13 Aug 2015 snpff_genomes.pdf  
-rv-r--r-- 1 milamalinsky staff 16K 18 Nov 2015.snpff_genes.txt  
-rv-r--r--@ 1 milamalinsky staff 336K 18 Nov 2015.snpff.summary.html  
-rv-r--r--@ 1 milamalinsky staff 336K 18 Nov 2017.species.tre.svg  
-rv-r--r-- 1 milamalinsky staff 471 11 Mar 2017.tre.rerooted.phylogeny.tre  
-rvx-r--r--@ 1 milamalinsky staff 1.7K 11 Aug 2012.split_paired_from_one_line.pl  
drwxr-xr-x 77 milamalinsky staff 2.4K 15 Mar 2015.stats_downloaded_from_farm  
drwxr-xr-x 9 milamalinsky staff 288B 15 May 2013.travel  
drwxr-xr-x 8 milamalinsky staff 256B 24 Jan 2014.treml1  
drwxr-xr-x 6 milamalinsky staff 1.6K 12 Mar 2015.trio_assembler  
drwxr-xr-x 65 milamalinsky staff 2.8K 6 Nov 2016.trio_assembly.data  
drwxr-xr-x 22 milamalinsky staff 704B 16 Nov 16:59.tutorials  
-rv-r--r-- 1 milamalinsky staff 551B 24 Apr 2013.unique_fixed11_scaffolds.txt  
drwxr-xr-x 18 milamalinsky staff 488B 11 Apr 2018.variant_calls  
drwxr-xr-x 5 milamalinsky staff 168B 12 Dec 2018.vcf_process  
-rv-r--r-- 1 milamalinsky staff 2.1K 5 Oct 2015.x  
drwxr-xr-x 10 milamalinsky staff 320B 1 Jul 2015.yeast_data  
milamalinsky:~/Sanger_work/tutorials/analysis_of_introgression_with_snp_data$ ls -l ..  
total 152  
-rv-r--r--@ 1 milamalinsky staff 73K 22 Nov 16:58 README.md  
drwxr-xr-x 24 milamalinsky staff 768B 22 Nov 16:23 data  
drwxr-xr-x 22 milamalinsky staff 704B 22 Nov 09:28 img  
drwxr-xr-x 15 milamalinsky staff 488B 16 Nov 16:58 res  
drwxr-xr-x 10 milamalinsky staff 320B 21 Nov 18:30 src  
milamalinsky:~/Sanger_work/tutorials/analysis_of_introgression_with_snp_data$ ls -lh ..  
total 152  
-rv-r--r--@ 1 milamalinsky staff 73K 22 Nov 16 Nov 16:58 README.md  
drwxr-xr-x 24 milamalinsky staff 768B 22 Nov 16:23 data  
drwxr-xr-x 22 milamalinsky staff 704B 22 Nov 09:28 img  
drwxr-xr-x 15 milamalinsky staff 488B 16 Nov 16:58 res  
drwxr-xr-x 10 milamalinsky staff 320B 21 Nov 18:30 src  
milamalinsky:~/Sanger_work/tutorials/analysis_of_introgression_with_snp_data$
```



# Why use something from the 1960s?

```
data — bash — 155x46  
.../np.data/data — bash .../c/pypopgen3 — bash .../c/pypopgen3 — bash .../  
/ork/tutorials — bash .../ith.snp_data — bash .../np.data/data — bash .../c/pypopgen3 — bash .../c/pypopgen3 — bash .../  
  
drwxr-xr-x  68 milamalinsky staff   2.1K  4 Mar  2013 report  
-rw-r--r--@  1 milamalinsky staff    4.0M  7 Sep  2012 report.zip  
drwxr-xr-x  7 milamalinsky staff  224B  7 Jun 15:52 ronchet.al  
-rw-r--r--@  1 milamalinsky staff    4.0M  7 Jun 2012 scaffolds_0_3749_samtools_var.raw.doubletons.txt  
-rw-r--r--@  1 milamalinsky staff    4.0M 10 May 2012 scaffolds_0_3749_samtools_var.raw.hets.txt  
-rwxr-xr-x  1 milamalinsky staff 1.8K 30 Jul 2012 sga-popcat.pl  
drwxr-xr-x  5 milamalinsky staff 160B 27 Aug 2014 sga_git_direct  
-rwxr--r--@  1 milamalinsky staff 532B 15 Sep 2014 sga_git_direct.pdf  
-rwxr--r--@  1 milamalinsky staff 16K 18 Nov 2015 smpff_genes.txt  
-rwxr--r--@  1 milamalinsky staff 336K 18 Nov 2015 smpff_summary.html  
-rwxr--r--@  1 milamalinsky staff 336K 18 Nov 2017 species.tre.svg  
-rwxr--r--@  1 milamalinsky staff 4.0M 19 Dec 2014 split_paired_reads_treology.tre  
-rwxr--r-x@  3 milamalinsky staff 1.7K 11 Aug 2012 split_paired_from_one_line.pl  
drwxr-xr-x  77 milamalinsky staff 2.4K 15 Mar 2015 stats_downloaded_from_farm  
drwxr-xr-x  9 milamalinsky staff 288B 15 May 2013 travel  
drwxr-xr-x  8 milamalinsky staff 256B 24 Jan 2014 treeall  
drwxr-xr-x  6 milamalinsky staff 1.6K 15 Feb 2015 trio_assembler  
drwxr-xr-x  65 milamalinsky staff 2.8K  6 Nov 2016 trio_assembly.data  
drwxr-xr-x  22 milamalinsky staff 704B 16 Nov 16:59 tutorials  
-rwxr--r--  1 milamalinsky staff 551B 24 Apr 2013 unique_fixed11_scaffolds.txt  
drwxr-xr-x 16 milamalinsky staff 488B 11 Apr 2018 variant_calls  
drwxr-xr-x  5 milamalinsky staff 168B 21 Nov 15:29 vcf_process  
-rwxr--r--  1 milamalinsky staff 2.1K  5 Oct 2015 x  
drwxr-xr-x 10 milamalinsky staff 320B  1 Jul 2015 yeast_data  
milamalinsky:/Sanger_work/tutorials/analysis_of_introgression_with.snp_data/data $  
milamalinsky:/Sanger_work/tutorials/analysis_of_introgression_with.snp_data/data $ ls -l  
milamalinsky:/Sanger_work/tutorials/analysis_of_introgression_with.snp_data/data $ vi ~/.bash_profile  
milamalinsky:/Sanger_work/tutorials/analysis_of_introgression_with.snp_data/data $ vi ~/.bash_profile  
milamalinsky:/Sanger_work/tutorials/analysis_of_introgression_with.snp_data/data $ ls -lh ..  
total 152  
-rwxr--r--@  1 milamalinsky staff 73K 22 Nov 16:58 README.md  
drwxr-xr-x  24 milamalinsky staff 768B 22 Nov 16:23 data  
drwxr-xr-x  22 milamalinsky staff 768B 22 Nov 09:28 img  
drwxr-xr-x  15 milamalinsky staff 488B 16 Nov 16:58 res  
drwxr-xr-x  10 milamalinsky staff 320B 21 Nov 18:30 src  
milamalinsky:/Sanger_work/tutorials/analysis_of_introgression_with.snp_data/data $ ls -lh ..  
total 152  
-rwxr--r--@  1 milamalinsky staff 73K 22 Nov 16:58 README.md  
drwxr-xr-x  24 milamalinsky staff 768B 22 Nov 16:23 data  
drwxr-xr-x  22 milamalinsky staff 768B 22 Nov 09:28 img  
drwxr-xr-x  15 milamalinsky staff 488B 16 Nov 16:58 res  
drwxr-xr-x  10 milamalinsky staff 320B 21 Nov 18:30 src  
milamalinsky:/Sanger_work/tutorials/analysis_of_introgression_with.snp_data/data $
```



**Scripting:** Write down a sequence of commands to perform a task

In genomics, a task almost always takes minutes, sometimes hours - not fun to sit and wait this long for the next mouse-click.

# Why use something from the 1960s?

```
data — bash — 155x46
...np.data/data — bash ...c/pypopgen3 — bash ...c/pypopgen3 — bash ...c/pypopgen3 — bash ...c/pypopgen3 — bash ...
drwxr-xr-x 68 milamalinsky staff 2.1K 4 Mar 2013 report
-rw-r--r--@ 1 milamalinsky staff 4.0M 7 Sep 2012 report.zip
drwxr-xr-x 7 milamalinsky staff 84K 16 Jun 2014 rh1VariationSummary.png
-rw-r--r--@ 1 milamalinsky staff 224B 7 Jun 15:52 rondon_et_al
-rw-r--r--@ 1 milamalinsky staff 2.0K 1 May 2014 scaffolds_0_3749_samtools_var.raw.doubletons.txt
-rw-r--r-- 1 milamalinsky staff 849 18 Mar 2013 scaffolds_0_3749_samtools_var.raw.hets.txt
-rw-r--r-- 1 milamalinsky staff 1.8K 30 Jul 2012 sga-popcat.pl
drwxr-xr-x 5 milamalinsky staff 160B 27 Aug 2014 sga_git_direct
-rw-r--r-- 1 milamalinsky staff 9.6M 11 Feb 2014 sga_git_direct.zip
-rw-r--r--@ 1 milamalinsky staff 532B 15 Aug 2014 snpEff_assembly.pdf
-rw-r--r-- 1 milamalinsky staff 16K 18 Nov 2015 snpEff_genes.txt
-rw-r--r--@ 1 milamalinsky staff 336K 18 Nov 2015 snpEff_summary.html
-rw-r--r--@ 1 milamalinsky staff 336K 18 Nov 2017 species.tre.svg
-rw-r--r--@ 1 milamalinsky staff 481B 14 Apr 2013 tre_treology.tre
-rw-r--r--@ 1 milamalinsky staff 1.7K 11 Aug 2012 split_paired_from_one_line.pl
drwxr-xr-x 77 milamalinsky staff 2.4K 15 Mar 2015 stats_downloaded_from_farm
drwxr-xr-x 9 milamalinsky staff 288B 15 May 2013 travel
drwxr-xr-x 8 milamalinsky staff 256B 24 Jan 2014 treeall
drwxr-xr-x 6 milamalinsky staff 193B 24 Jan 2014 trio_assembler
drwxr-xr-x 65 milamalinsky staff 2.8K 6 Nov 2016 trio_assembly_data
drwxr-xr-x 22 milamalinsky staff 704B 16 Nov 16:59 tutorials
-rw-r--r-- 1 milamalinsky staff 551B 24 Apr 2013 unique_fixed11_scaffolds.txt
drwxr-xr-x 18 milamalinsky staff 488B 11 Apr 2018 variant_calls
drwxr-xr-x 5 milamalinsky staff 168B 25 Mar 2015 vcf_process
-rw-r--r-- 1 milamalinsky staff 2.1K 5 Oct 2015 x...
drwxr-xr-x 10 milamalinsky staff 320B 1 Jul 2015 yeast_data
milamalinsky:~/Sanger_work/tutorials/analysis_of_introgression_with.snp_data/data $ 
milamalinsky:~/Sanger_work/tutorials/analysis_of_introgression_with.snp_data/data $ ls -lh ..
milamalinsky:~/Sanger_work/tutorials/analysis_of_introgression_with.snp_data/data $ vi ~/.bash_profile
milamalinsky:~/Sanger_work/tutorials/analysis_of_introgression_with.snp_data/data $ vi ~/.bash_profile
milamalinsky:~/Sanger_work/tutorials/analysis_of_introgression_with.snp_data/data $ ls -lh ..
total 152
-rw-r--r--@ 1 milamalinsky staff 73K 22 Nov 16:58 README.md
drwxr-xr-x 24 milamalinsky staff 768B 22 Nov 16:23 data
drwxr-xr-x 22 milamalinsky staff 784B 22 Nov 09:28 img
drwxr-xr-x 15 milamalinsky staff 488B 16 Nov 16:58 res
drwxr-xr-x 10 milamalinsky staff 320B 21 Nov 18:30 src
milamalinsky:~/Sanger_work/tutorials/analysis_of_introgression_with.snp_data/data $ ls -lh ..
total 152
-rw-r--r--@ 1 milamalinsky staff 73K 22 Nov 16:58 README.md
drwxr-xr-x 24 milamalinsky staff 768B 22 Nov 16:23 data
drwxr-xr-x 22 milamalinsky staff 784B 22 Nov 09:28 img
drwxr-xr-x 15 milamalinsky staff 488B 16 Nov 16:58 res
drwxr-xr-x 10 milamalinsky staff 320B 21 Nov 18:30 src
milamalinsky:~/Sanger_work/tutorials/analysis_of_introgression_with.snp_data/data $
```

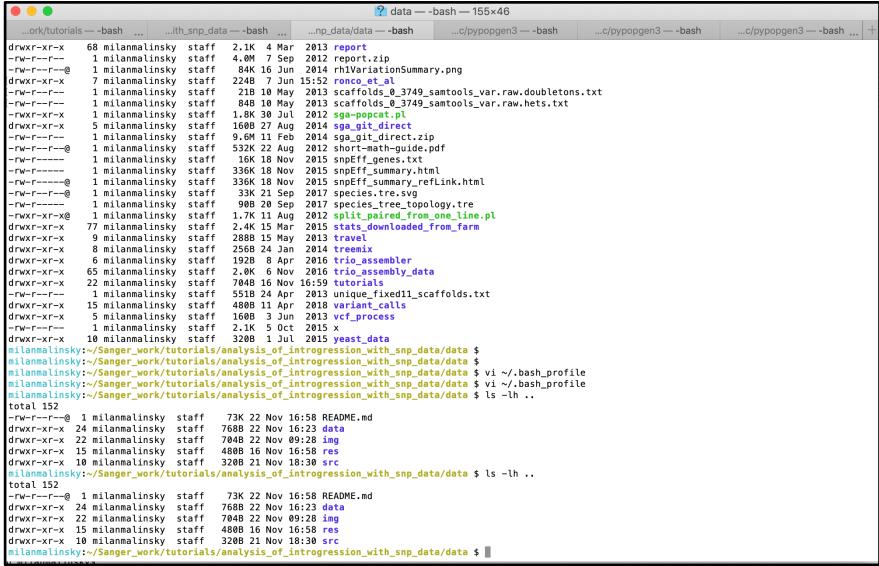


**Scripting:** Write down a sequence of commands to perform a task

In genomics, a task almost always takes minutes, sometimes hours - not fun to sit and wait this long for the next mouse-click.

**Easy remote access:** Running a real world genomics project on your own computer is impossible; you will usually access high performance compute facilities at your university.

# Why use something from the 1960s?



```
data -- bash -- 155x46
...np.data/data -- bash ...c/pypopgen3 -- bash ...c/pypopgen3 -- bash ...
drwxr-xr-x 68 milamalinsky staff 2.1K 4 Mar 2013 report
-rw-r--r--@ 1 milamalinsky staff 4.0M 7 Sep 2012 report.zip
drwxr-xr-x 7 milamalinsky staff 84K 16 Jun 2014 rh1VariationSummary.png
drwxr-xr-x 1 milamalinsky staff 224B 7 Jun 15:52 ronchet.al
drwxr-xr-x 1 milamalinsky staff 2.0M 1 May 2013 scaffolds_0_3749_samtools_var.raw.doubletons.txt
drwxr-xr-x 1 milamalinsky staff 849 10 May 2013 scaffolds_0_3749_samtools_var.raw.hets.txt
drwxr-xr-x 5 milamalinsky staff 160B 27 Aug 2014 sga.git_direct
drwxr-xr-x 1 milamalinsky staff 9.6M 11 Feb 2014 sga.git_direct.zip
drwxr-xr-x@ 1 milamalinsky staff 532B 23 Aug 2015 snpff_samtools.pdf
drwxr-xr-x 1 milamalinsky staff 16K 18 Nov 2015.snpff.genes.txt
drwxr-xr-x@ 1 milamalinsky staff 336K 18 Nov 2015.snpff.summary.html
drwxr-xr-x@ 1 milamalinsky staff 336K 18 Nov 2017.species.tre.svg
drwxr-xr-x 1 milamalinsky staff 4.0M 1 Apr 2018.snpff.phylogeny.tre
drwxr-xr-x@ 1 milamalinsky staff 1.7K 11 Aug 2012.split_paired_from_one_line.pl
drwxr-xr-x 77 milamalinsky staff 2.4K 15 Mar 2015.stats_downloaded_from_farm
drwxr-xr-x 9 milamalinsky staff 288B 15 May 2013.travel
drwxr-xr-x 8 milamalinsky staff 256B 24 Jan 2014.treeml
drwxr-xr-x 6 milamalinsky staff 198B 24 Jan 2014.trio_assembler
drwxr-xr-x 65 milamalinsky staff 2.8K 6 Nov 2016.trio_assembly.data
drwxr-xr-x 22 milamalinsky staff 704B 16 Nov 16:59.tutorials
drwxr-xr-x 1 milamalinsky staff 551B 24 Apr 2013.unique_fixed11_scaffolds.txt
drwxr-xr-x 18 milamalinsky staff 488B 11 Apr 2018.variant_calls
drwxr-xr-x 5 milamalinsky staff 160B 21 Mar 2015.vcf_process
drwxr-xr-x 1 milamalinsky staff 2.1K 5 Oct 2015.x
drwxr-xr-x 10 milamalinsky staff 320B 1 Jul 2015.yeast_data
milamalinsky:~/Sanger_work/tutorials/analysis_of_introgression_with.snp_data/data $ ls -l ..
total 152
drwxr-xr-x@ 1 milamalinsky staff 73K 22 Nov 16:58 README.md
drwxr-xr-x@ 24 milamalinsky staff 768B 22 Nov 16:23 data
drwxr-xr-x@ 22 milamalinsky staff 704B 22 Nov 09:28 img
drwxr-xr-x@ 15 milamalinsky staff 488B 16 Nov 16:58 res
drwxr-xr-x@ 10 milamalinsky staff 320B 21 Nov 18:30 src
milamalinsky:~/Sanger_work/tutorials/analysis_of_introgression_with.snp_data/data $ ls -lh ..
total 152
drwxr-xr-x@ 1 milamalinsky staff 73K 22 Nov 16:58 README.md
drwxr-xr-x@ 24 milamalinsky staff 768B 22 Nov 16:23 data
drwxr-xr-x@ 22 milamalinsky staff 704B 22 Nov 09:28 img
drwxr-xr-x@ 15 milamalinsky staff 488B 16 Nov 16:58 res
drwxr-xr-x@ 10 milamalinsky staff 320B 21 Nov 18:30 src
milamalinsky:~/Sanger_work/tutorials/analysis_of_introgression_with.snp_data/data $ 
```



**Scripting:** Write down a sequence of commands to perform a task

In genomics, a task almost always takes minutes, sometimes hours - not fun to sit and wait this long for the next mouse-click.

**Easy remote access:** Running a real world genomics project on your own computer is impossible; you will usually access high performance compute facilities at your university.

**GUI for many programs not available:** Genomics is a fast moving field and developing a graphical interface takes time and effort

# Why use something from the 1960s?

```

data — bash — 155x46
...np.data/data — bash ...c/pypopgen3 — bash ...c/pypopgen3 — bash ...c/pypopgen3 — bash ...
drwxr-xr-X 68 milamalinsky staff 2.1K 4 Mar 2013 report
-rw-r--r--@ 1 milamalinsky staff 4.0M 7 Sep 2012 report.zip
drwxr-xr-X 7 milamalinsky staff 84K 16 Jun 2014 rh1VariationSummary.png
drwxr-xr-X 1 milamalinsky staff 224B 7 Jun 15:52 ronchet.al
drwxr-xr-X 1 milamalinsky staff 517 27 May 2012 scaffolds_0_3749_samtools_var.raw.doubletons.txt
drwxr-xr-X 1 milamalinsky staff 849 10 Mar 2012 scaffolds_0_3749_samtools_var.raw.hets.txt
drwxr-xr-X 5 milamalinsky staff 160B 27 Aug 2014 sga_git_direct
drwxr-xr-X 1 milamalinsky staff 9.6M 11 Feb 2014 sga_git_direct.zip
drwxr-xr-X 1 milamalinsky staff 532B 18 Nov 2015 snpEff_snps.pdf
drwxr-xr-X 1 milamalinsky staff 16K 18 Nov 2015 snpEff_genes.txt
drwxr-xr-X@ 1 milamalinsky staff 336K 18 Nov 2015 snpEff_summary.html
drwxr-xr-X@ 1 milamalinsky staff 336K 18 Nov 2017 species.tre.svg
drwxr-xr-X 1 milamalinsky staff 4.1M 26 Jul 2012 split_paired_from_one_line.tre
drwxr-xr-X 77 milamalinsky staff 1.7K 11 Aug 2012 split_paired_from_one_line.pl
drwxr-xr-X 9 milamalinsky staff 2.4K 15 Mar 2015 stats_downloaded_from_farm
drwxr-xr-X 9 milamalinsky staff 288B 15 May 2013 travel
drwxr-xr-X 8 milamalinsky staff 256B 24 Jan 2014 treeall
drwxr-xr-X 6 milamalinsky staff 153B 16 Nov 2015 trio_assembler
drwxr-xr-X 65 milamalinsky staff 2.8K 6 Nov 2016 trio_assembly_data
drwxr-xr-X 22 milamalinsky staff 704B 16 Nov 16:59 tutorials
drwxr-xr-X 1 milamalinsky staff 551B 24 Apr 2013 unique_fixed11_scaffolds.txt
drwxr-xr-X 18 milamalinsky staff 488B 11 Apr 2018 variant_calls
drwxr-xr-X 5 milamalinsky staff 168B 23 May 2018 vcf_process
drwxr-xr-X 1 milamalinsky staff 2.1K 5 Oct 2015 x
drwxr-xr-X 10 milamalinsky staff 320B 1 Jul 2015 yeast_data
milamalinsky:~/Sanger_work/tutorials/analysis_of_introgression_with.snp_data/data $
milamalinsky:~/Sanger_work/tutorials/analysis_of_introgression_with.snp_data/data $ vi ~/.bash_profile
milamalinsky:~/Sanger_work/tutorials/analysis_of_introgression_with.snp_data/data $ vi ~/.bash_profile
milamalinsky:~/Sanger_work/tutorials/analysis_of_introgression_with.snp_data/data $ ls -lh ..
total 152
drwxr-xr-X@ 1 milamalinsky staff 73K 22 Nov 16:58 README.md
drwxr-xr-X 24 milamalinsky staff 768B 22 Nov 16:23 data
drwxr-xr-X 22 milamalinsky staff 784B 22 Nov 09:28 img
drwxr-xr-X 15 milamalinsky staff 488B 16 Nov 16:58 res
drwxr-xr-X 10 milamalinsky staff 320B 21 Nov 18:30 src
milamalinsky:~/Sanger_work/tutorials/analysis_of_introgression_with.snp_data/data $ ls -lh ..
total 152
drwxr-xr-X@ 1 milamalinsky staff 73K 22 Nov 16:58 README.md
drwxr-xr-X 24 milamalinsky staff 768B 22 Nov 16:23 data
drwxr-xr-X 22 milamalinsky staff 784B 22 Nov 09:28 img
drwxr-xr-X 15 milamalinsky staff 488B 16 Nov 16:58 res
drwxr-xr-X 10 milamalinsky staff 320B 21 Nov 18:30 src
milamalinsky:~/Sanger_work/tutorials/analysis_of_introgression_with.snp_data/data $ l

```



**Scripting:** Write down a sequence of commands to perform a task

In genomics, a task almost always takes minutes, sometimes hours - not fun to sit and wait this long for the next mouse-click.

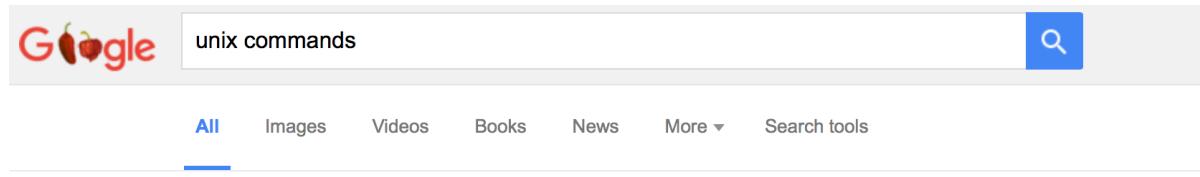
**Easy remote access:** Running a real world genomics project on your own computer is impossible; you will usually access high performance compute facilities at your university.

**GUI for many programs not available:** Genomics is a fast moving field and developing a graphical interface takes time and effort

**Powerful tools available in UNIX:** enabling you to work through large amounts of files, data, and tasks quickly in an automated (programmatic) way

# How to survive without ‘seeing’ your data, without pop-out menus, etc.?

## 1. Google: get a cheat-sheet with a list of possible commands:



About 20,000,000 results (0.28 seconds)

### Basic UNIX commands

[mally.stanford.edu/~sr/computing/basic-unix.html](http://mally.stanford.edu/~sr/computing/basic-unix.html) ▾

Note that some of these **commands** are different on non-Solaris machines - see SunOS

... But you can also edit the command line (see the guide to **More UNIX**).

[More UNIX Commands](#) - [Emacs](#) - [Pathnames](#)

## 2. Manual pages for each command:

`milanmalinsky:~ $ man ls`

```
milanmalinsky — less < man ls — 143x41
...torials -- bash ... ...p_data -- bash ... ...a/data -- bash ~ -- less < man ls ...pgen3 -- bash ...pgen3 -- bash ...pgen3 -- bash ...

LS(1) BSD General Commands Manual LS(1)

NAME
  ls -- list directory contents

SYNOPSIS
  ls [-ABCFGHLoprstuw@abcdefghijklmnopqrstuvwxyz] [file ...]

DESCRIPTION
  For each operand that names a file of a type other than directory, ls displays its name as well as any requested, associated information. For each operand that names a file of type directory, ls displays the names of files contained within that directory, as well as any requested, associated information.

  If no operands are given, the contents of the current directory are displayed. If more than one operand is given, non-directory operands are displayed first; directory and non-directory operands are sorted separately and in lexicographical order.

  The following options are available:

  -@      Display extended attribute keys and sizes in long (-l) output.

  -1      (The numeric digit `one'.) Force output to be one entry per line. This is the default when output is not to a terminal.
```

What do the cp, vi, and awk commands do? try: `man cp`, `man vi`, `man awk`

# How to survive without ‘seeing’ your data, without pop-out menus, etc.?

**3. Help:** more concise than manual pages; describes different ways to run the program:

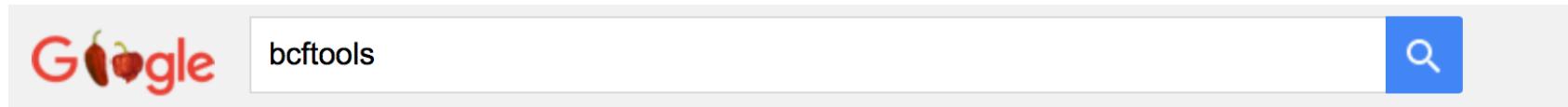
`milanmalinsky:~ $ Dsuite --help`  
or

`milanmalinsky:~ $ Dsuite -h`

**4. Genomics tools also sometimes have manual pages:**

`ubuntu@ip-10-144-40-7:~$ man bcftools`

**5. Google again:**



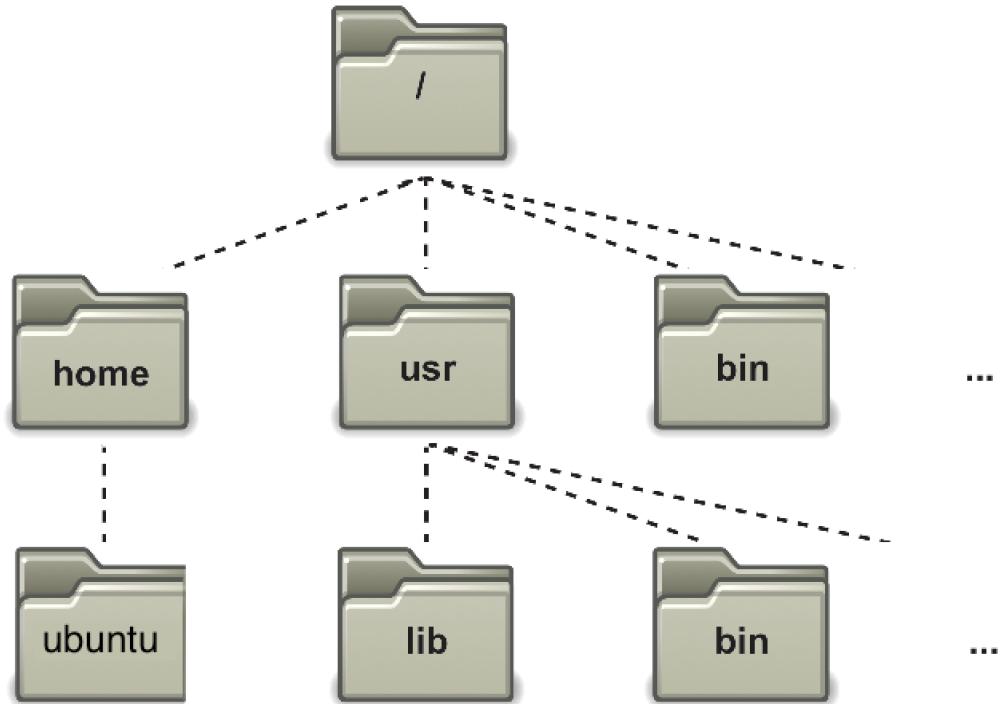
About 23,700 results (0.37 seconds)

## bcftools - Samtools

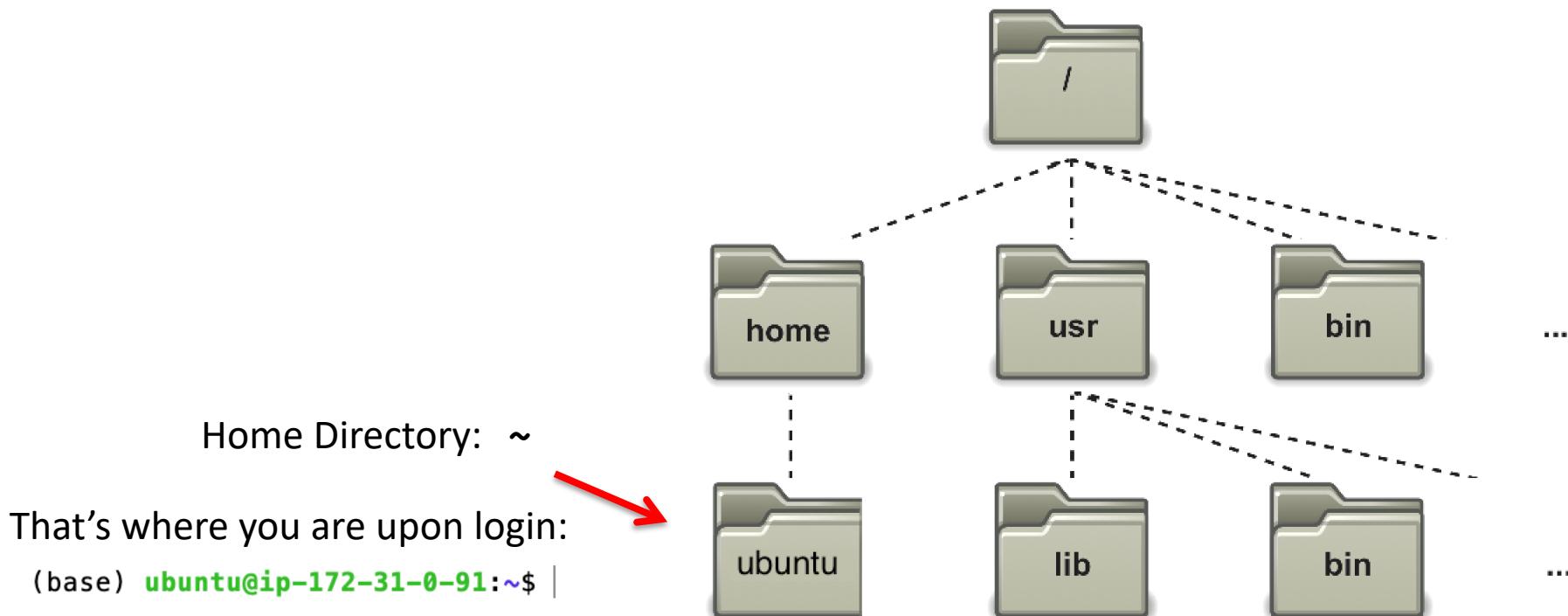
<https://samtools.github.io/bcftools/bcftools.html> ▾

Nov 16, 2015 - DESCRIPTION. BCFtools is a set of utilities that manipulate variant calls in the Variant Call Format (VCF) and its binary counterpart BCF.

# UNIX directory structure



# UNIX directory structure



# Many ‘programs’ are in the `/bin` folder

```
[milanmalinsky:~ $ ls /bin
[  chmod  date  echo  hostname  launchctl  ls  pax
bash  cp  dd  ed  kill  link  ln  mv  rm
cat  cd  df  expr  ksh  ln  mkdir  rmrmdir
                           mv  sleep  tcsh  wait4path
                           pwd  sh  stty  sync  test  zsh
                           sync  unlink  zsh
```



For example:

- `cp` – Copy
  - `ls` – List directory contents
  - `pwd` – Know where you are ('print working directory')
- Also maybe
- `cd` – Change directory

# A few ways to view a text file:

<b>less</b>	<b>head</b>	<b>tail</b>	<b>cat</b>
view a text file one screen-full at a time	view the top 10 lines of a file	view the bottom 10 lines of a file	print the whole file at once on screen
space-bar: scroll q: quit	<b>-n num</b> option controls the number of lines	<b>-n num</b> option controls the number of lines	

# Helpful features to save you typing:

## 1. Tab completion

- <tab> once to complete a ‘word’ uniquely
- <tab><tab> twice to show all possible completions

## 2. Up-arrow (history)

- Show previous commands

### Try:

1. cd /etc
2. ls
3. cd c<tab><tab>
4. cd cal<tab>
5. cd ~ (Move back to your folder – remember ~)
6. mkdir unix
7. cd unix

# Extracting example data into the new **unix** folder:

**Some example data are in the following compressed tar file:**  
**unixExampleData.tar.gz**

What is a tar archive?



**tar = tape archive**

When you download programs for UNIX, they often come in tar files.

Genomics data is normally stored compressed to save on disk space/costs.

# Extracting example data into the new **unix** folder:

The example data are in the following compressed **tar file**:

`~/unixExampleData.tar.gz`

<b>gzip/gunzip</b>	<b>tar -xvzf</b>	<b>tar -xvf</b>	<b>bgzip</b>
compress/decompress a file like <code>unixExampleData.tar.gz</code>	extract a gzipped tar archive	extract an uncompressed tar archive like <code>unixExampleData.tar</code>	another compression algorithm you will come across in genomics

# Extracting example data into the new **unix** folder:

The example data are in the following compressed **tar file**:

**unixExampleData.tar.gz**

```
(base) ubuntu@ip-172-31-0-91:~$ cd unix/
(base) ubuntu@ip-172-31-0-91:~/unix$ wget https://www.dropbox.com/s/i5uuxobrc0kjndi/unixExampleData.tar.gz

(base) ubuntu@ip-172-31-0-91:~/unix$ ls -lh
total 3.8M
-rw-rw-r-- 1 ubuntu ubuntu 3.8M Nov 23 11:52 unixExampleData.tar.gz
(base) ubuntu@ip-172-31-0-91:~/unix$ tar -xvzf unixExampleData.tar.gz
exampleVariants.vcf.gz
examplesGenomeSeqence.fastq.gz
(base) ubuntu@ip-172-31-0-91:~/unix$ ls -lh
total 7.5M
-rw-r----- 1 ubuntu ubuntu 3.6M Jan 22 2016 examplesGenomeSeqence.fastq.gz
-rw-r----- 1 ubuntu ubuntu 174K Jan 23 2016 exampleVariants.vcf.gz
-rw-rw-r-- 1 ubuntu ubuntu 3.8M Nov 23 11:52 unixExampleData.tar.gz
```

The **unixExampleData** archive contains two compressed files:

- 1) **examplesGenomeSeqence.fastq.gz**
  - output from a genome sequencer - Illumina type
- 2) **exampleVariants.vcf.gz**
  - a file with variants (differences) from the reference genome

# Working with the sequence file:

The FASTQ format:

---

```
@HS22_09582:8:1105:9492:70247#84/1
GAGAATCTCATCCACATCACAGGCTATTTGGCCCCAGCCAGGCAGCGGGGGTAAATCCTCTTCATGCCTGATCCACCCCTGGCATGCATCTACTGAT
+
B?CFDFGEFFEGFGFFIFEEGICCDFFGJFGFGEFFEHHFEGHEEGFGFG?GFFFGGGDGFGEFFHFBGGFFEGFDFFHGEFGGBFFFDFFGEEFCE
```

1. Decompress the file
2. Have a look at the file
  - `less examplesGenomeSeqence.fastq`
3. Count the number of reads
  - `wc -l examplesGenomeSeqence.fastq`
  - `grep "@HS" examplesGenomeSeqence.fastq`
  - `grep -c "@HS" examplesGenomeSeqence.fastq`
  - `grep -v "@HS" examplesGenomeSeqence.fastq`
  - `grep -v -c "@HS" examplesGenomeSeqence.fastq`
4. Print and count reads with undetermined bases in them:
  - `grep "N" examplesGenomeSeqence.fastq`
  - `grep -c "N" examplesGenomeSeqence.fastq`

# Working with the VCF variants file:

The VCF file (without a header) is a text file with <tab> separated columns:

- columns 1-8: information about a variant (location, alleles, quality scores, filtering, etc.)
- columns 9 onwards: information about the genotypes (variants) present in each individual

# Working with the VCF variants file:

The VCF file (without a header) is a text file with <tab> separated columns:

- columns 1-8: information about a variant (location, alleles, quality scores, filtering, etc.)
- columns 9 onwards: information about the genotypes (variants) present in each individual

There is a lot of information, especially in the INFO and FORMAT columns.

The formal specification is here:

<https://samtools.github.io/hts-specs/VCFv4.2.pdf>

# Working with the VCF variants file:

The VCF file (without a header) is a text file with <tab> separated columns:

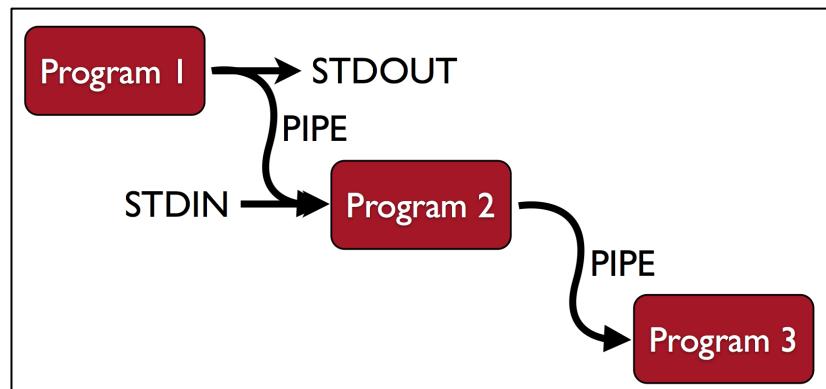
- columns 1-8: information about a variant (location, alleles, quality scores, filtering, etc.)
- columns 9 onwards: information about the genotypes (variants) present in each individual

We will use one of UNIX's cool features: **Pipes**

So far we have been doing this:



Now we are going to learn to link multiple UNIX programs:



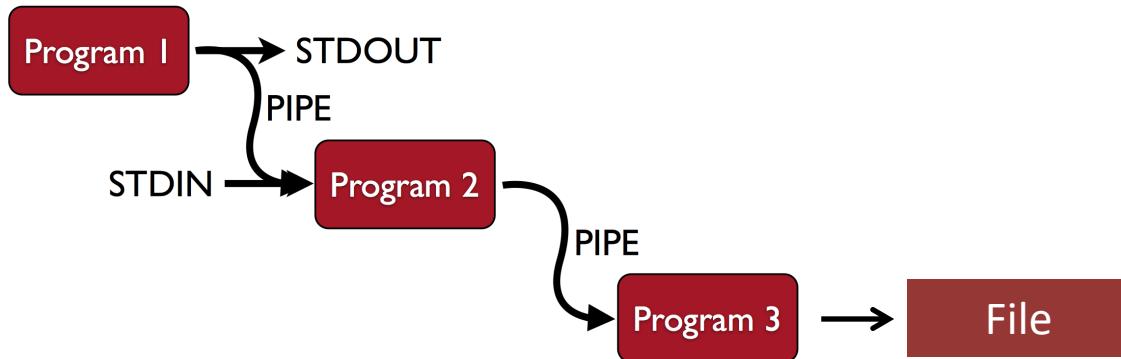
# Working with the VCF variants file:

The VCF file (without a header) is a text file with <tab> separated columns:

- columns 1-8: information about a variant (location, alleles, quality scores, filtering, etc.)
- columns 9 onwards: information about the genotypes (variants) present in each individual

1. Decompress the file
2. Have a look at the file
  - `less exampleVariants.vcf`
  - `less -S exampleVariants.vcf`
3. Find out how many chromosomes there are:
  - `cut -f 1 exampleVariants.vcf`
  - `cut -f 1 exampleVariants.vcf | uniq`
  - Make sure you know what `cut -f` does! Try `cut -f 2`
  - `man uniq` (see what `uniq` does)
4. Find how many variants have a “T” as the reference allele (fourth column):
  - `cut -f 4 exampleVariants.vcf`
  - `cut -f 4 exampleVariants.vcf | grep "T"`
  - `cut -f 4 exampleVariants.vcf | grep -o "T"`
  - `cut -f 4 exampleVariants.vcf | grep -o "T" | wc -l`
  - `man grep` (see what the `-o` option does)

# Working with the VCF variants file:



5. Capture the genotype information for the first three individuals into a different file:
  - `cut -f 10-12 exampleVariants.vcf > threeGenotypes.txt`
6. Remove the newly created file:
  - `rm threeGenotypes.txt`



!!!There is **NO** undo button, no trash can!!!!

**Once deleted a file is lost forever (or at least you'd need professional data recovery)**

- `rm -i exampleVariants.vcf` (then press **n** and enter)

```
(base) ubuntu@ip-172-31-0-91:~/unix$ rm -i exampleVariants.vcf
rm: remove regular file 'exampleVariants.vcf'? █
```

# Try again:

Do you still remember how to uncompress the **tar** file?

