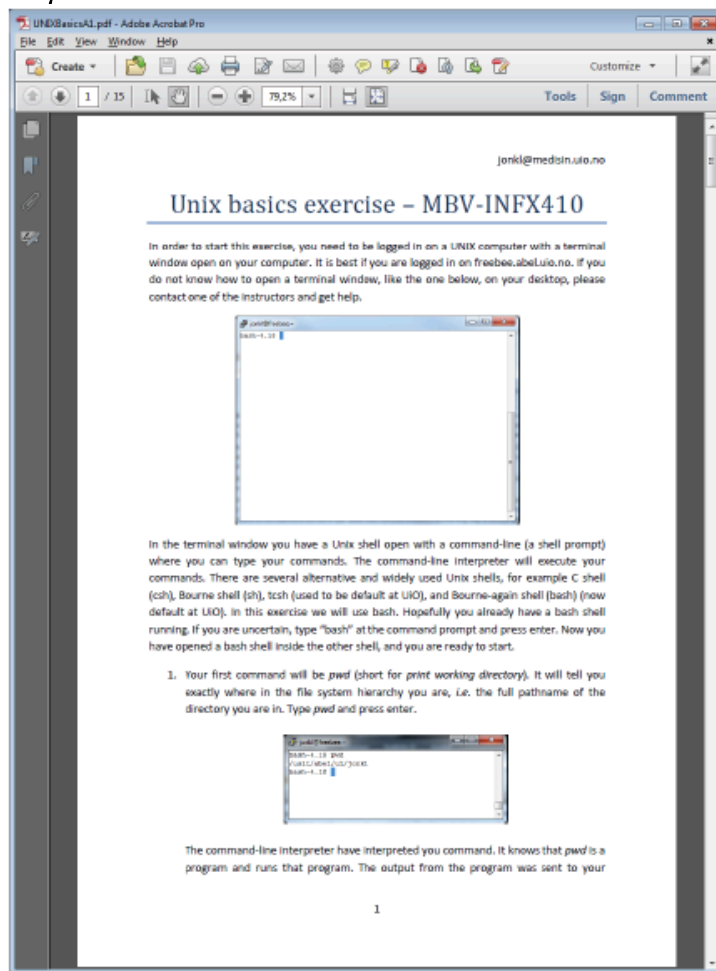


November 16. More Unix
Bioinformatics for Molecular Biology





Very important to be able to use
basic Unix

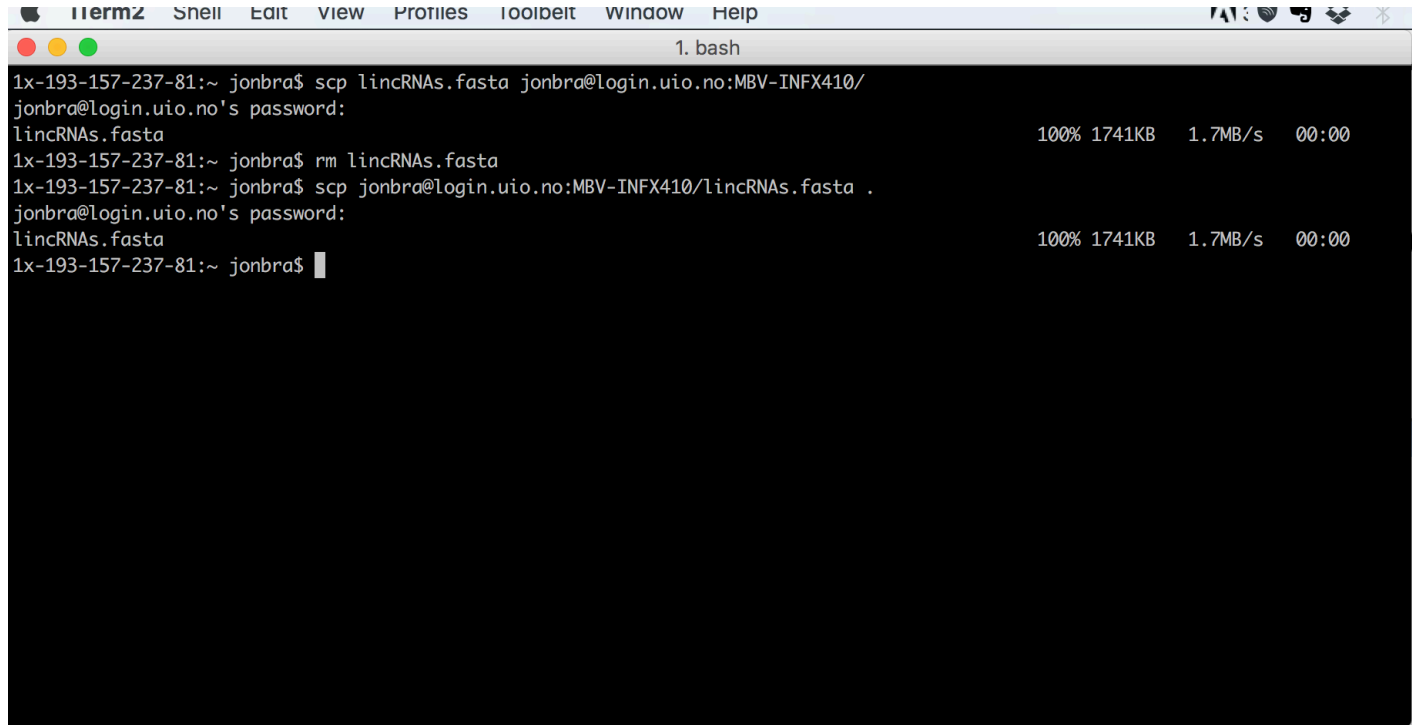
Make your own Unix scripts is
not very central

Most likely you will instead use
Python for writing small
programs/scripts

More Unix stuff

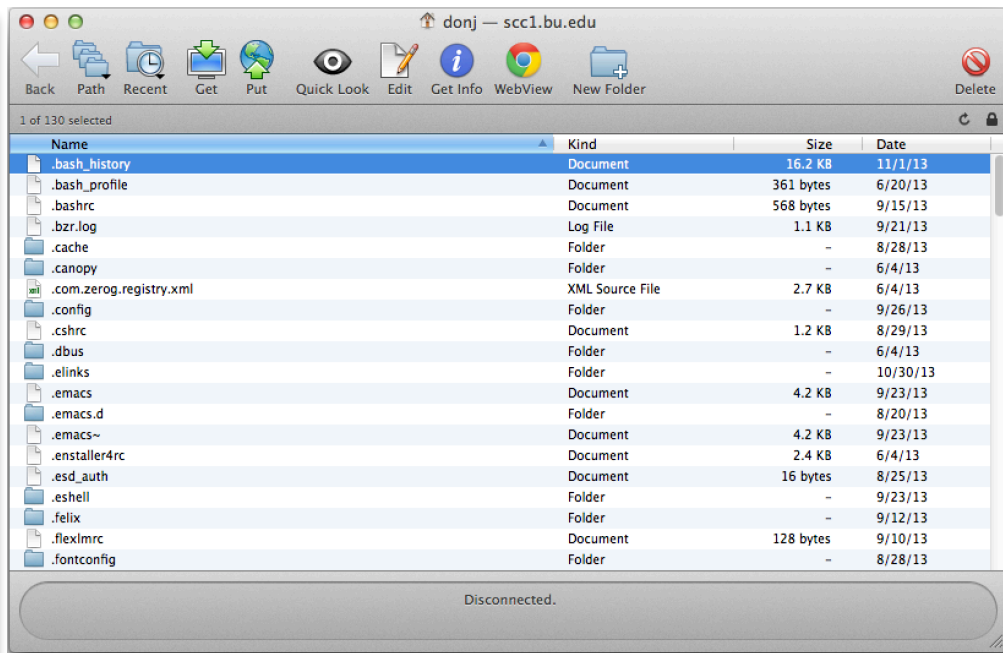
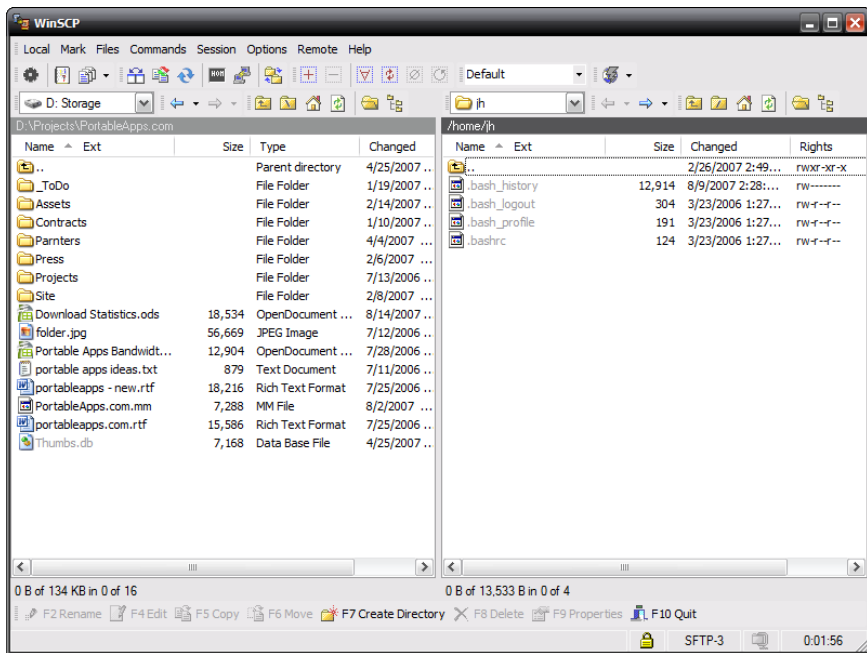
- You ***must*** be able to move files between laptop and a Unix server such as freebee.abel.uio.no or login.uio.no!
- That is, between the harddisk on your laptop and your UiO home area (M:)
- Most likely you use "UiO Programkiosk"
 - Or WebDrive, Cyberduck etc.

Copy files to and from UiO home area (Terminal/Cygwin etc.)



```
Term2  Shell  Edit  View  Profiles  Toolbelt  Window  Help
1. bash
1x-193-157-237-81:~ jonbra$ scp lincRNAs.fasta jonbra@login.uio.no:MBV-INF410/
jonbra@login.uio.no's password:
lincRNAs.fasta                                     100% 1741KB   1.7MB/s   00:00
1x-193-157-237-81:~ jonbra$ rm lincRNAs.fasta
1x-193-157-237-81:~ jonbra$ scp jonbra@login.uio.no:MBV-INF410/lincRNAs.fasta .
jonbra@login.uio.no's password:
lincRNAs.fasta                                     100% 1741KB   1.7MB/s   00:00
1x-193-157-237-81:~ jonbra$
```

Copy files to and from UiO home area (graphical)



Moving files with the mv command

```
iTerm2  Shell  Edit  View  Profiles  Toolbelt  Window  Help
2. jonbra@ulrik:~/MBV-INF410 (ssh)

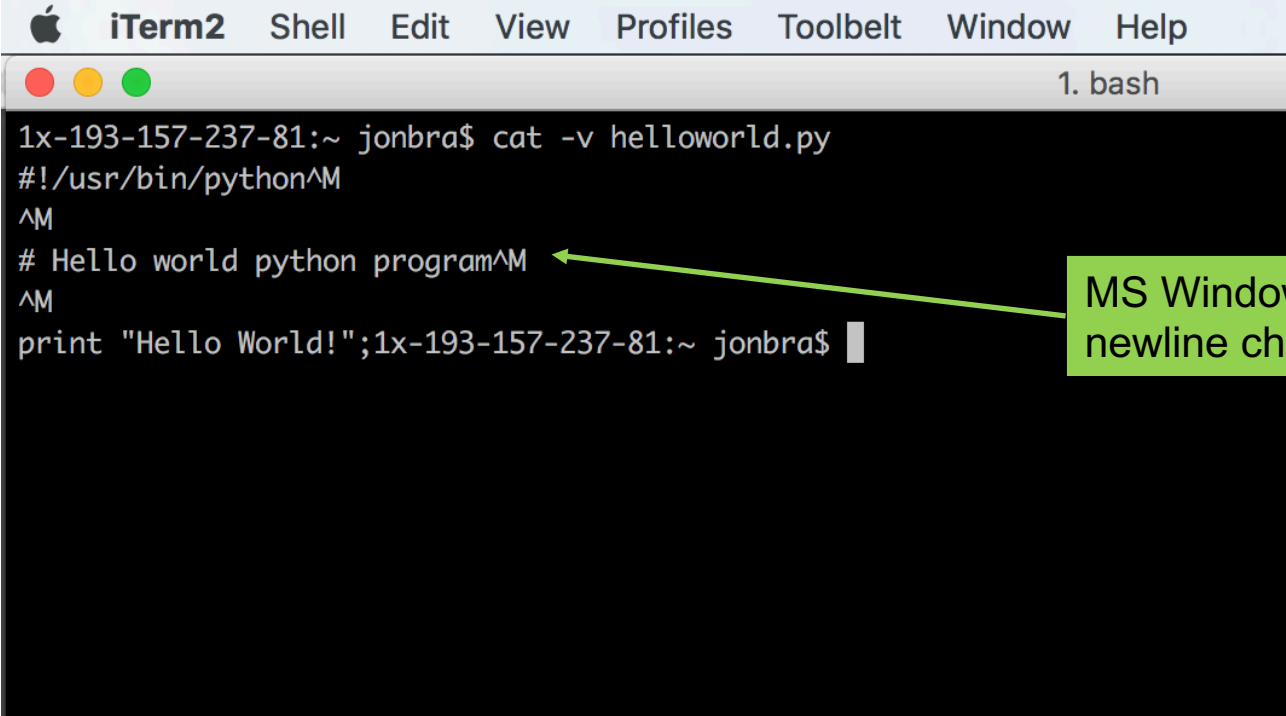
-bash-4.1$ ls -lrt
total 204403
drwxr-xr-x 5 jonbra users      7 Jul 14 21:32 sratoolkit.2.7.0-ubuntu64
-rw-r--r-- 1 jonbra users 63707890 Jul 14 21:33 sratoolkit.tar.gz
-rw-r--r-- 1 jonbra users   297 Oct 11 13:41 getRawData.sh
-rw-r--r-- 1 jonbra users   546 Oct 11 13:51 sra.slurm
-rw-r--r-- 1 jonbra users   730 Oct 11 14:03 slurm-15702476.base
-rw-r--r-- 1 jonbra bio   72051869 Oct 14 23:39 gbbct100.seq.gz.1
-rw-r--r-- 1 jonbra bio   72051869 Oct 14 23:39 gbbct100.seq.gz
-rw-r--r-- 1 jonbra users 1496824 Nov  7 10:00 genes.fpk_tracking
drwxr-xr-x 2 jonbra bio      1 Nov 15 09:36 Ensembl-demo
drwxr-xr-x 2 jonbra bio      0 Nov 15 15:15 mvtest
-rw-r--r-- 1 jonbra bio      0 Nov 15 15:15 file1.txt
-bash-4.1$ mv file1.txt mvtest/
-bash-4.1$ cd mvtest/
-bash-4.1$ ls
file1.txt
-bash-4.1$ mv file1.txt file2.txt
-bash-4.1$ ls
file2.txt
-bash-4.1$
```

Move file1.txt to directory mvtest

Pro tip:
mv command can also rename files!
("moves" file to a new filename)

Non-printing characters

Use `cat -v` to see “invisible characters”. Typically End of Line (EOL) characters



```
iTerm2  Shell  Edit  View  Profiles  Toolbelt  Window  Help
1. bash
1x-193-157-237-81:~ jonbra$ cat -v helloworld.py
#!/usr/bin/python^M
^M
# Hello world python program^M
^M
print "Hello World!";1x-193-157-237-81:~ jonbra$
```

MS Windows
newline character

NB! dos2unix might
have to be installed.
Already installed on
freebee

Non-printing characters

EOLs differ between Windows and Unix (including Mac) languages

Use dos2unix command to go from MS Windows to Unix

Use unix2dos command to do the reverse



```
2. jonbra@ulrik:~/MBV-INF410 (ssh)
-bash-4.1$ dos2unix helloworld.py
dos2unix: converting file helloworld.py to UNIX format ...
-bash-4.1$ cat -v helloworld.py
#!/usr/bin/python

# Hello world python program

print "Hello World!";-bash-4.1$
```



```
jonki@freebee: ~/MBV-INF410_Python
bash-4.1$ cat 10verts.fasta
>gi|4505471|ref|NP_002519.1| endonuclease III-like protein 1 [Homo sapiens]
MCSPEQESGNTALGARMLTRSRSLGPGAGPRGCREEPGLRRREAAAARKSHSPVKRPRKAQRLRVAYEGSDSEKGEAE
PLKVPVWEPQDQQQLVNIIRAMRNKKDAPVDHLGTEHCYDSSAPPKVRRYQVL LSLMLSSQTKDQVTAGAMQRLRARGLT
VDSILQITDAILGKLIYPVGFWRSKVKYIKQTSAILQQHYGGDIPASVAELVALPGVGPKMAHLAMAVANGTVSGIAVDI
HVHRIANRLNWKATKATKSPEETRAJALEEENLPRELNHEINGLLVGFSGQQTCLFVHPCHACLNQALCFAAQGL
>gi|6912520|ref|
MTPLVSRLSRLWAIMR
YHLFRDVAEVTAFRGS
SASLEEVENQLWAGLGY
CRVRAIGADPSSSTLVS
SGSPDVEECAPNTGQC
FPSVTIWEPEQLQRKA
AAVSTAMKKVFRVYQG
>gi|227908769|re
MNSGVRMVTRSRSRAT
QQQLANIRIMRSKKDA
GRLIYPVGFWRANKVKY
KKMIKIPEETIRKNLEE
>gi|227330621|re
MKKLQASVRSCHKQPD
DLFWRMLAKEEANSR
ARKVVEELGGHMPRTA
VDFARPQDFNQAMEL
MGVANFPRKASRRPFR
FAIRLQHLGEVHIYS
VCPSSSRKKPSLGQV
>gi|114051958|re
MNAAGVRMVVTRARSR
QPQDWRQQLDNIRTM
DDSTLGLIYFVGFW
RLRWTKKATKSPEETR
>gi|281485563|re
MKKSRAAVGNASGRK
KKRDLFWRLVEDEV
QEGARKVVEELGGMF
QQIVDPARPGDFNQAMELGAIVCTPKRPLCSHCVFQNLCRARQVREREQLSASQSLPGNCDVEECAPNTGQCPLCAPPT
EPFWDQTLGVTNFPKASRKPPRECSAICVLEQPKALGGAHILLVQRPNSGLLAGLWEFPSSVSVNAEASGQHQAALLQE
LQSWVGPLEPDLRLQHLGQVHTFPHIMTYQVYSLALEEHTPVTIVPPGARWLTREDHTAAVSTAMKKVFRMYEGQPG
TCKGSKRSQVAILSKRKKFSPGQQVLESFFWEHVPTDAPSLNTAAQ
>gi|118601744|ref|NP_001073043.1| endonuclease III-like protein 1 [Gallus gallus]
MCAAAPRGGGGAARRLGAATAGSRVPSAAPRYSRRTRRVPIAYEAEKPESEPGKWEPEENWQQQLERIREMRHRDAPVD
EMGVDKCYDTISAPPQVMRYQVLLSLMLSSQTKDQVTSAAMLRLRQGLTVDSILQMDDATLGQIYFVGFWRNKVKYIKQ
TTAILKQHYGGDIPGTVEELVKLPGVGPKMAHLAMNIAWNSVSGIAVDTHVHRITNRLKWKVKESTRYFEETRVLEDWLP
RDLWREINNLVGFSGQQTCLFVNPRCKECLNQDIPAAKRF
>gi|513197809|ref|XP_422433.3| PREDICTED: A/G-specific adenine DNA glycosylase isoform X5 [Gallus gallus]
MGGAARVARRSVKVRAGGHEVGPGLGSPAIALRTHRRCCDPTVPVPSRQGLPLDHMCISSTVPSRSMFYAACSPGMITK
AGGTFPGRMLQLSNMLTGGRMQLGLLVLASERQVWARERSCGDEDEGECWVGCCSSWNQQHGDGACCEKWHWHLCT
DPSVLLAVNVSEIMLQQTQVATVIDYNNMMQKNFTLQALAAASLEEVENLNAGLGYYSRGKRLQEAARKVVSSELAGRMP
RTAEDLQRLLPVGGRYTAGAIASISFGQATGVVDGNVIRVLRLCAIGADTSSLAVIDCLDWMANTLVDRSRPGDFNQAL
```

Everything on a single line!

NB! There's no trash can on Unix!
Your files are immediately deleted!

Delete files with the `rm` command

```
iTerm2  Shell  Edit  View  Profiles  Too
-bash-4.1$ ls
file1.txt  file2.txt  file3.txt
-bash-4.1$ rm file1.txt
-bash-4.1$ ls
file2.txt  file3.txt
-bash-4.1$ rm file*
-bash-4.1$ ls
-bash-4.1$
```

Delete empty directories with `rm -r`

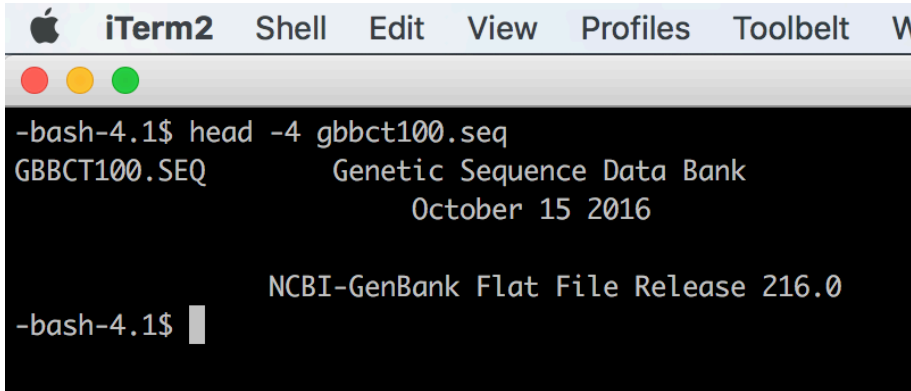
Delete directories AND
files within with `rm -rf` Use with
caution!!

```
iTerm2  Shell  Edit  View  Profiles  Toolbelt  Window  Help
2. jonbra@u
-bash-4.1$ ls -lrt
total 63677
drwxr-xr-x 5 jonbra users      7 Jul 14 21:32 sratoolkit.2.7.0-ubuntu64
-rw-r--r-- 1 jonbra users 63707890 Jul 14 21:33 sratoolkit.tar.gz
-rw-r--r-- 1 jonbra users    297 Oct 11 13:41 getRawData.sh
-rw-r--r-- 1 jonbra users    546 Oct 11 13:51 sra.slurm
-rw-r--r-- 1 jonbra users    730 Oct 11 14:03 slurm-15702476.base
-rw-r--r-- 1 jonbra users 1496824 Nov  7 10:00 genes.fpk_tracking
drwxr-xr-x 2 jonbra bio        1 Nov 15 09:36 Ensembl-demo
drwxr-xr-x 2 jonbra bio        0 Nov 15 14:44 deleteme
-bash-4.1$ rm -r deleteme/
```

Viewing files

Pro tip:
Use the different
options to tweak the output

- `cat` displays the entire content on the screen
- `less` displays the content on the screen one file/line at a time (press "q" to exit).
- `head/tail` displays the first or last lines of a file



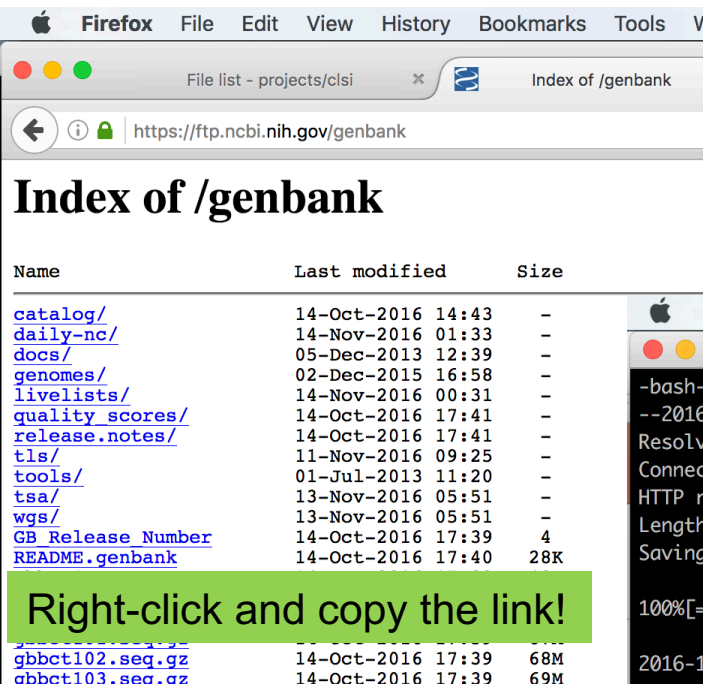
```
iTerm2  Shell  Edit  View  Profiles  Toolbelt  W
-bash-4.1$ head -4 gbbct100.seq
GBBCT100.SEQ      Genetic Sequence Data Bank
                  October 15 2016

                  NCBI-GenBank Flat File Release 216.0
-bash-4.1$
```



```
iTerm2  Shell  Edit  View  Profiles  Toolbelt  Window  Help
2. jonbra@ulrik:~/M
-bash-4.1$ tail -9 gbbct100.seq
6542881 cctttgcgcg cgatcaccac gatataccag ccagccaggg ttctctggtt atggcggaac
6542941 gatttcgcga tcaggcggtt gaggcgattg cgctggacgg cgagcttgac gttcttcttg
6543001 ccgatcacca ggcccaggcg ggggtgatcg agaccgttct cgcgcgccag cagcaggacg
6543061 tgcttgccgg ggaccttgcc gatcggagag tcgaagactg cgctgaattg ccgggctgtc
6543121 agtagacgct tgtcccgttc gaagtccga ctcaccaccg gtaccggata aatcagacgg
6543181 tcagacgctt acggcctttg gcgcgacgac gcgacagAAC ctgacggccg ttcttggtgg
6543241 ccatacgggc gcggaacccg tggacgcgag cgcgcttgag ggtgctgggt tggaaagtac
6543301 gtttcgatgat tcggtacctg gggtgacgac ttgaggtcgc agtgaccgcc
//
-bash-4.1$
```

Downloading files from the internet



Firefox File Edit View History Bookmarks Tools V

File list - projects/clsi Index of /genbank

https://ftp.ncbi.nih.gov/genbank

Index of /genbank

Name	Last modified	Size
catalog/	14-Oct-2016 14:43	-
daily-nc/	14-Nov-2016 01:33	-
docs/	05-Dec-2013 12:39	-
genomes/	02-Dec-2015 16:58	-
livelists/	14-Nov-2016 00:31	-
quality_scores/	14-Oct-2016 17:41	-
release.notes/	14-Oct-2016 17:41	-
tls/	11-Nov-2016 09:25	-
tools/	01-Jul-2013 11:20	-
tsa/	13-Nov-2016 05:51	-
wgs/	13-Nov-2016 05:51	-
GB Release Number	14-Oct-2016 17:39	4
README.genbank	14-Oct-2016 17:40	28K
gbbct102.seq.gz	14-Oct-2016 17:39	68M
gbbct103.seq.gz	14-Oct-2016 17:39	69M

Right-click and copy the link!

The wget command should work on freebee
Other options are curl or ftp



```
iTerm2 Shell Edit View Profiles Toolbelt Window Help
2. jonbra@ulrik:~/MBV-INF410 (ssh)

-bash-4.1$ wget https://ftp.ncbi.nih.gov/genbank/gbbct100.seq.gz
--2016-11-15 14:59:18-- https://ftp.ncbi.nih.gov/genbank/gbbct100.seq.gz
Resolving ftp.ncbi.nih.gov... 130.14.250.13, 2607:f220:41e:250::7
Connecting to ftp.ncbi.nih.gov|130.14.250.13|:443... connected.
HTTP request sent, awaiting response... 200 OK
Length: 72051869 (69M) [application/x-gzip]
Saving to: `gbbct100.seq.gz'

100%[=====>] 72,051,869 19.8M/s in 5.8s

2016-11-15 14:59:25 (11.8 MB/s) - `gbbct100.seq.gz' saved [72051869/72051869]

-bash-4.1$
```

Back up your data!

- Don't store only on your laptop
- Use M: on UiO machines
- Same as ~/ in Unix environment
- ~/pc/Desktop on login or freebee is the same as M: on UiO machines