



Canadian Bioinformatics Workshops

www.bioinformatics.ca

This page is available in the following languages:

Afrikaans বাংলা Català Dansk Deutsch Ελληνικά English (GB) English (US) Esperanto
 Castellano Castellano (AR) Español (CL) Castellano (CO) Español (Ecuador) Castellano (MX) Castellano (PE)
 Euskara Suomi français français (CA) Galego עברית hrvatski Magyar Italiano 日本語 한국어 Macedonian Malayu
 Nederlands Norsk Sesotho sa Leboa polski Português română slovenski jezik српски (latinica) Sotho svenska
 中文 華語 (台灣) isiZulu



Attribution-Share Alike 2.5 Canada

You are free:



to Share — to copy, distribute and transmit the work



to Remix — to adapt the work



Under the following conditions:



Attribution. You must attribute the work in the manner specified by the author or licensor (but not in any way that suggests that they endorse you or your use of the work).



Share Alike. If you alter, transform, or build upon this work, you may distribute the resulting work only under the same or similar licence to this one.

- For any reuse or distribution, you must make clear to others the licence terms of this work.
- Any of the above conditions can be waived if you get permission from the copyright holder.
- The author's moral rights are retained in this licence.

[Disclaimer](#)

Your fair dealing and other rights are in no way affected by the above.

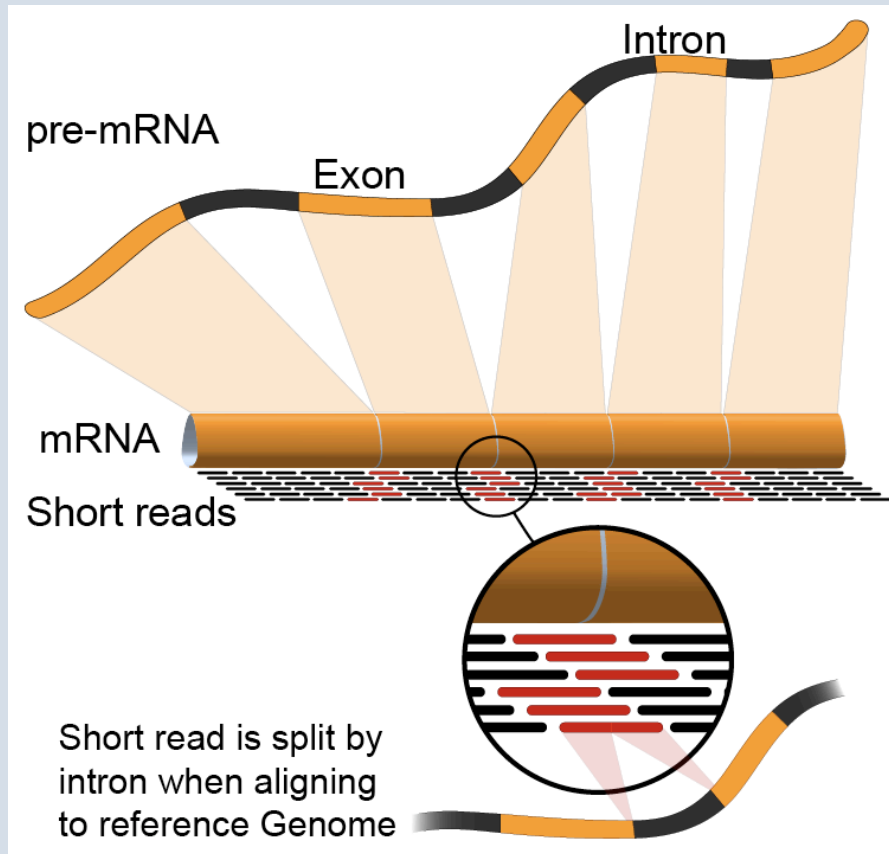
This is a human-readable summary of the Legal Code (the full licence) available in the following languages:
[English](#) [French](#)

Module 0

Introduction to cloud computing

(slides modified with permission from Francis Ouellette)

Malachi Griffith & Obi Griffith
Informatics for RNA-seq Analysis
June 8-9, 2015



Learning objectives of the course

- **Module 0: Introduction to cloud computing**
- Module 1: Introduction to RNA sequencing
- Module 2: RNA-seq alignment and visualization
- Module 3: Expression and Differential Expression
- Module 4: Isoform discovery and alternative expression
- Tutorials
 - Provide a working example of an RNA-seq analysis pipeline
 - Run in a ‘reasonable’ amount of time with modest computer resources
 - Self contained, self explanatory, portable

Learning Objectives of module 0

- Introduction to cloud computing
- Use of the wiki(s) in this workshop
- How to log into the cloud

Disk Capacity vs Sequencing Capacity, 1990-2012

Disk Storage
(Mbytes/\$)

Stein Genome Biology 2010, 11:207
<http://genomebiology.com/2010/11/5/207>

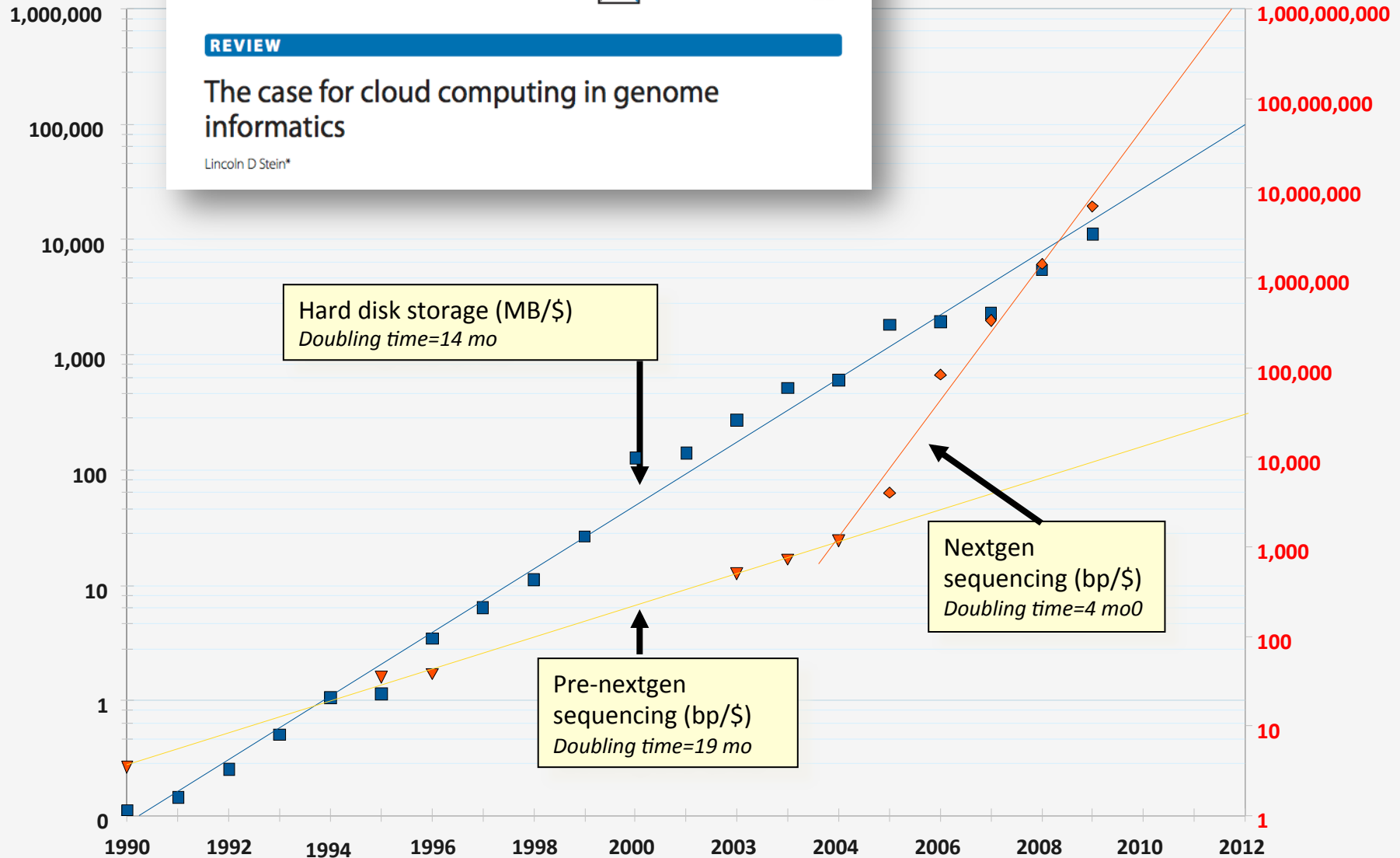


REVIEW

The case for cloud computing in genome informatics

Lincoln D Stein*

DNA
Sequencing (bp/\$)



About DNA and computers

- We'll hit the \$1000 genome during 2015-?, then need to think about the \$100 genome.
- The doubling time of sequencing has been ~5-6 months.
- The doubling time of storage and network bandwidth is ~12 months.
- The doubling time of CPU speed is ~18 months.
- The cost of sequencing a base pair will eventually equal the cost of storing a base pair

What is the general biomedical scientist to do?

- Lots of data
- Poor IT infrastructure in many labs
- Where do they go?
- Write more grants?
- Get bigger hardware?

Amazon Web Services (AWS)

- Infinite storage (scalable): S3 (simple storage service)
- Compute per hour: EC2 (elastic cloud computing)
- Ready when you are High Performance Computing
- Multiple football fields of HPC throughout the world
- HPC are expanded at one container at a time:



<http://goo.gl/7PVAI>



Some of the challenges of cloud computing:

- Not cheap!
- Getting files to and from there
- Not the best solution for everybody
- Standardization
- PHI: personal health information & security concerns
- In the USA: HIPAA act, PSQIA act, HITECH act, Patriot act, CLIA and CAP programs, etc.
 - <http://www.biostars.org/p/70204/>

Some of the advantages of cloud computing:

- We received a grant from Amazon, so supported by 'AWS in Education grant award'.
- There are better ways of transferring large files, and now AWS makes it free to upload files.
- A number of datasets exist on AWS (e.g. 1000 genome data).
- Many useful bioinformatics AMI's (Amazon Machine Images) exist on AWS: e.g. cloudbiolinux & CloudMan (Galaxy) – now one for this course!
- Many flavors of cloud available, not just AWS



In this workshop:

- Some tools (data) are
 - on your computer
 - on the web
 - on the cloud.
- You will become efficient at traversing these various spaces, and finding resources you need, and using what is best for you.
- There are different ways of using the cloud:
 1. Command line (like your own very powerful Unix box)
 2. With a web-browser (e.g. Galaxy): not in this workshop

Things we have set up:

- Loaded data files to an ftp server
- We brought up an Ubuntu (Linux) instance, and loaded a whole bunch of software for NGS analysis.
- We then cloned this, and made separate instances for everybody in the class.
- We've simplified the security: you basically all have the same login and file access, and opened ports. In your own world you would be more secure.

Amazon AWS Management Console – quick walkthrough

<http://aws.amazon.com/console/>

For this workshop: all on Wiki!

http://bioinformatics.ca/workshop_wiki/index.php/

Login: FirstnameLastname

Password: 'guest'

special page

Log in

You must have cookies enabled to log in to CBW Workshop Wiki.

Username:

Password:

[Forgotten your login details?](#)

(maximum of 180 days)

navigation

- [Main page](#)
- [Community portal](#)
- [Current events](#)
- [Recent changes](#)
- [Random page](#)
- [Help](#)


search

toolbox

- [Special pages](#)

[Privacy policy](#) [About CBW Workshop Wiki](#) [Disclaimers](#)

The main CBW Wiki



Canadian
Bioinformatics
Workshops

MalachiGriff

page discussion edit history move watch

Main Page

Welcome to the CBW Workshop Wiki Pages!

- [High-throughput Biology: From Sequence to Networks 2015 Workshop Wiki](#) April 27 - May 3, 2015, New York, NY
- [Introduction to R 2015 Workshop Wiki](#) May 20, 2015, Toronto, ON
- [Exploratory Analysis of Biological Data using R 2015 Workshop Wiki](#) May 21-22, 2015, Toronto, ON
- [Bioinformatics for Cancer Genomics 2015 Workshop Wiki](#) May 25-29, 2015, Toronto, ON
- [Pathway and Network Analysis 2015 Workshop Wiki](#) June 1-3, 2015, Toronto, ON
- [Informatics for RNA-seq Analysis 2015 Workshop Wiki](#) June 8-9, 2015, Toronto, ON
- [Informatics on High-throughput Sequencing Data 2015 Workshop](#) June 10-11, 2015, Toronto, ON
- [Informatics and Statistics for Metabolomics 2015 Workshop](#) June 15-16, 2015, Montreal, QC
- [Analysis of Metagenomic Data 2015 Workshop Wiki](#) June 24-26, 2015, Halifax, NS

navigation

- [Main page](#)
- [Community portal](#)
- [Current events](#)
- [Recent changes](#)
- [Random page](#)
- [Help](#)

search

toolbox

- [What links here](#)
- [Related changes](#)
- [Upload file](#)
- [Special pages](#)
- [Printable version](#)
- [Permanent link](#)

This page was last modified on 20 May 2015, at 09:28. This page has been accessed 18,097 times. [Privacy policy](#) [About CBW](#)

http://bioinformatics.ca/workshop_wiki/index.php/

The RNA-seq wiki

griffithlab / rnaseq_tutorial

Unwatch

22

Star

21

Fork

11

Home

Malachi Griffith edited this page on Apr 27 · 49 revisions

Edit

New Page

Informatics for RNA-seq: A web resource for analysis on the cloud



Welcome to the RNA-seq Tutorial. Use this page to navigate your way through all exercises. Each page has a link at the bottom to bring you back to this table of contents.

Table of Contents

0. Module 0 - Introduction and Cloud Computing

- i. [Authors](#)
- ii. [Citation](#)
- iii. [Syntax](#)
- iv. [Intro to AWS Cloud Computing](#)
- v. [Logging into Amazon Cloud](#)
- vi. [Environment](#)
- vii. [Resources](#)

1. Module 1 - Introduction to RNA sequencing

- i. [Installation](#)
- ii. [Reference Genome](#)
- iii. [Annotation](#)
- iv. [Indexing](#)
- v. [RNA-seq Data](#)
- vi. [PreAlignment QC](#)

Pages 35

Find a Page...

[Home](#)

[Abbreviations](#)

[Adapter Trim](#)

[Alignment](#)

[Annotation](#)

[Authors](#)

[AWS Setup](#)

[Citation](#)

[de novo Transcript Assembly](#)

[DE Visualization](#)

[Differential Expression](#)

[Differential Splicing](#)

[Environment](#)

[Expression](#)

[Indexing](#)

Show 20 more pages...

+ Add a custom sidebar

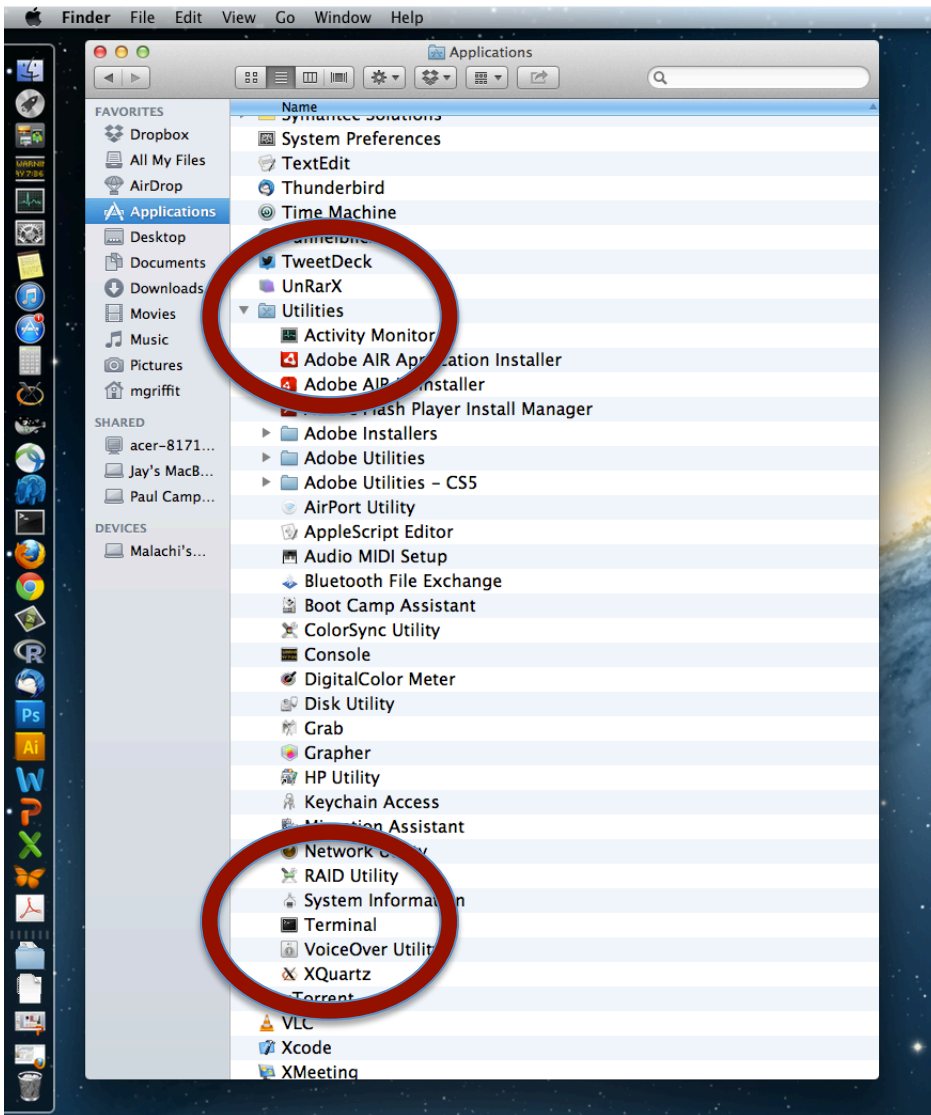
Clone this wiki locally

<http://www.rnaseq.wiki/>

Macintosh users



Opening a 'terminal session'



```
[mgriffit@Malachis-MacBook-Pro-2:~]$ pwd
/Users/mgriffit
[mgriffit@Malachis-MacBook-Pro-2:~]$ ls -l
total 16
drwxr-xr-x  2 mgriffit  staff   68 Jan  8 23:27 Applications
drwx-----+ 8 mgriffit  staff  272 Jun  1 15:58 Desktop
drwx-----+ 7 mgriffit  staff  238 Apr 17 22:00 Documents
drwx-----+ 12 mgriffit  staff  408 Jun  1 15:38 Downloads
drwx-----@ 12 mgriffit  staff  408 Jun  1 15:13 Dropbox
drwx-----@ 51 mgriffit  staff 1734 Feb 27 10:00 Library
drwx-----+ 4 mgriffit  staff  136 May 30 2012 Movies
drwx-----+ 4 mgriffit  staff  136 Mar  9 2012 Music
drwx-----+ 8 mgriffit  staff  272 Dec  9 21:39 Pictures
drwxr-xr-x+  5 mgriffit  staff  170 Mar  5 2012 Public
-rwxr-xr-x  1 mgriffit  staff  594 Mar  6 2012 addSSHkey.sh
drwxr-xr-x 16 mgriffit  staff  544 Jan 17 21:42 backup
drwxr-xr-x  5 mgriffit  staff  170 May  7 18:32 git
drwxr-xr-x  6 mgriffit  staff  204 Mar  6 2012 igv
-rwxr-xr-x  1 mgriffit  staff  552 Dec  3 21:53 mac2unix
drwxr-xr-x  3 mgriffit  staff  102 Jun  1 15:55 notes
drwxr-xr-x 95 mgriffit  staff 3230 Dec 29 23:42 temp
drwxr-xr-x  3 mgriffit  staff  102 Oct 27 2012 workspace
[mgriffit@Malachis-MacBook-Pro-2:~]$ mkdir cbw
[mgriffit@Malachis-MacBook-Pro-2:~]$ cd cbw
[mgriffit@Malachis-MacBook-Pro-2:cbw]$ ls -la
total 0
drwxr-xr-x  2 mgriffit  staff   68 Jun  1 16:03 .
drwxr-xr-x+ 64 mgriffit  staff 2176 Jun  1 16:03 ..
[mgriffit@Malachis-MacBook-Pro-2:cbw]$
```

Creating a working directory on your mac

```
cbw — bash — 93x24
bash

mgriffit@mgriffit-macpro:~$ pwd
/Users/mgriffit
mgriffit@mgriffit-macpro:~$ ls
AWS-Tutorial    Documents      Installed      Music  iTunes      Sync          igv
Applications    Downloads      Library        Pictures  VirtualBox_VMs  perl5
Box Sync        Dropbox        Movies         Public    bin           temp
Desktop         Google Drive   Music          Sites     git           tools

mgriffit@mgriffit-macpro:~$ mkdir cbw
mgriffit@mgriffit-macpro:~$ cd cbw/
mgriffit@mgriffit-macpro:~/cbw$ ls -la
total 0
drwxr-xr-x  2 mgriffit  staff   68 Jun  1 18:16 .
drwxr-xr-x+ 85 mgriffit  staff 2890 Jun  1 18:16 ..
mgriffit@mgriffit-macpro:~/cbw$
```

Obtaining your AWS 'key' file from the wiki

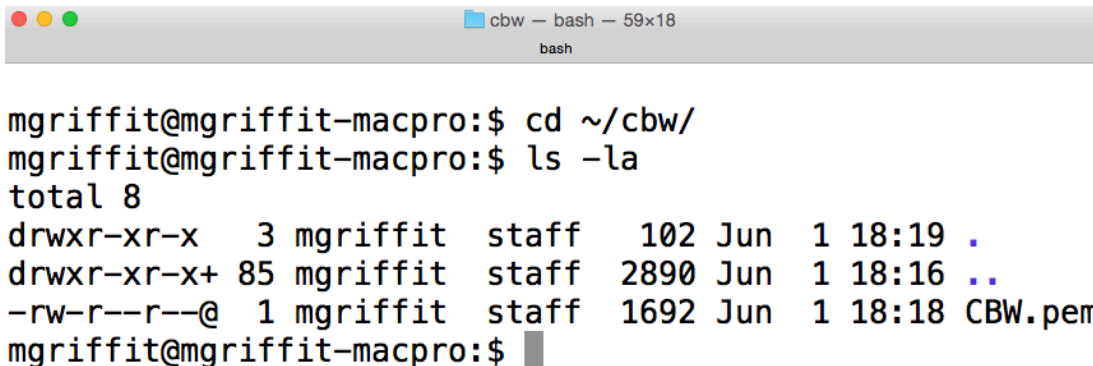
Logging into the Amazon cloud

[\[edit\]](#)

- This will ONLY occur once we are in the classroom as it costs to have these servers running. Instructions will be provided in class.
- We have set up 40 instances on the Amazon cloud - one for each student. In order to log in to your instance, you will need a security certificate.
 - If you plan on using **Linux** or **Mac OS X**, please download this [CBW.pem](#).
 - If you plan on using **Windows** (with Putty and Winscp), please download this [CBW.ppk](#).

On Mac:
Control+

On
Windows



A terminal window titled 'cbw — bash — 59x18' with 'bash' in the subtitle. The window shows a user named 'mgriffit' at a 'mgriffit-macpro' machine. The user has navigated to the directory '~/cbw/' and run the command 'ls -la'. The output shows three files: a dot file, a '..' file, and a file named 'CBW.pem' with permissions '-rw-r--r--@'.

```
mgriffit@mgriffit-macpro:~$ cd ~/cbw/
mgriffit@mgriffit-macpro:~/cbw$ ls -la
total 8
drwxr-xr-x  3 mgriffit  staff   102 Jun  1 18:19 .
drwxr-xr-x+ 85 mgriffit  staff  2890 Jun  1 18:16 ..
-rw-r--r--@ 1 mgriffit  staff   1692 Jun  1 18:18 CBW.pem
mgriffit@mgriffit-macpro:~/cbw$
```

**Save key file to
your new 'cbw'
directory**

Viewing the 'key' file once downloaded

```
mgriffit@mgriffit-macpro:~$ cd ~/cbw/
mgriffit@mgriffit-macpro:~/cbw$ ls -la
total 8
drwxr-xr-x  3 mgriffit  staff   102 Jun  1  18:19 .
drwxr-xr-x+ 85 mgriffit  staff  2890 Jun  1  18:16 ..
-rw-r--r--@ 1 mgriffit  staff  1692 Jun  1  18:18 CBW.pem
mgriffit@mgriffit-macpro:~/cbw$ cat CBW.pem
-----BEGIN RSA PRIVATE KEY-----
MIIeogIBAAKCAQEAjM/lfVDGcmjU70QldKA0/VX5vdE+MyxWdhDsN9iMy/ALX0VKJh1oCKP6onzT
6tew0t45rQ+qZBNZ+bsnL/zSd3eacbMgKh7ipM6bmNjD/p1mpyTn87V+8aH5vC9u0EiiH0K+le0Q
PhUWQ15GsYZifQRhxueB00R7wil+ZA4yqHiryw2r16+X7Z+KrXTRN/3ArD0i5iYaNRb41ztXC0s
G6wHgGELE986/2E/2ruM201GtQ+OX/TYgBfrSXoC0wUY2okMTsVGjigtfwV33zs9i/hiNsE0B2uZ
D/353kq+bZ4ZZbL8akJzfn2RqxdwAWB4jSAF7XMAz+3I0VQHMgoySQIDAQABAoIBAGMgl7W4fzIk
yoBZu+R2KRc+xyLRfDAaLm0/VguabsPCoL0szR435Vzv0ylAJw9T+C76MjAaryD0XornrZairKdd
k2QnbaU5S0Qsbm0TbCZ60CHCYWe3hT4lgOFxom4M89R98B0rXhvwTG3wmG0w7vSEBzh9wKnzH4Te
2bjuRoyegzJFLAX18k0dmenwwWdtugRaNHVj49zx32BnK2T7aVW1n0dwJe14poai7z67hG5sPdie
XufcwNq0CKpJKIWTet00EJTCnfo+Dxg7Ukb937fZaq27EZ1wwz+/K/ZL/YHgDpHD2udQG8sYl03t
lVhB35v/VfSjtU7y7PHPJKfGzAECgYEA/ti+jrnmgQnhCVUVFr2YpajFbhJoFFSF39tdb9emMxSh
Mbsv3Bn5bEX7U/q01PpxFmqM+jM/Zh091z1J1nQWi0Q+HajrcUDgDW1gsHmzbPv9BM90/N/xipn0
ibMB2guq0UXRNIhIWuc8jweAw690nd0h+dKZpStQGV+fr7PPCNkCgYEAjXMJNbmIFfPGpMddHH3I
syHJeSiTDk6D3L7ZvMA/Ea4iTZi6r0za0iVvvlJgSE4UPy3mMZfY3doiv/hhQ04/IPAKImcQo+mL
pgyGfZtUunf12K4/CPw6LLuv+PMesxDGC5w0zzvLNx0chWv0DXV/aS98+JhrqIM7pkYAZCZ/DvEC
gYBrxrYwHiaBg/ALxl043k3kIU/9kQM7Nd04LEmDtyuuhpb1NTi0WLfHNoFnrpshLxkMZdthJLM3
NDohA0ZtL4Mo0X6bhIityMX4Z71wZMeg26DRm/rZxsub1hgm2/F4iV6gSmqDU09ookpWItIt78HL
tWjXAl3AIGUrlekuoVh5YQKBgBrhvONX8jPx010ni9uJUCYzCp4bBu+WHltKbgkAJDxQ8rX6i6en
j+KWOMlb0G6NYC2tpGzSZeWruITvqFP1s0uoMtTc732nPtLy2HBwn8acW72lR9eGT+wGc2gFLFUF
oA2+zw5/XI1/29bDcGtKZD9yRfbcyMEKx8B0124xHr6RAoGAB77iIJjR1TAMCydk7QxuJwUEGquS
0I3DowhJpjpYQ4sgRFGbrmQVUpKoC5++KbuLcpL/M+3vXdRMr4UZSMLnbLbF2U+u1XKCrh5o66mp
N2fZ0lFWHzjQfo5FQQU4+c+EeBlI8pWrBiN2B7IFsDTLzSG9hiwcUHvdbHR+ZKLMbE=
mgriffit@mgriffit-macpro:~/cbw$
```


Changing file permissions of your 'key' file

ls -l (long listing)

```
drwx-----+ 67 francis staff 2278 22 May 21:25 ../  
-rw-r--r--@ 1 francis staff 1696 22 May 21:31 CBW.pem
```

rwX : owner

rwX : group

rwX: world

r read (4)

w write (2)

x execute (1)

Which ever way you add these 3 numbers, you know which integers were used (6 is always 4+2, 5 is 4+1, 4 is by itself, 0 is none of them etc ...)

So, when you have:

chmod 600 <file name>

It is "rw" for the the file owner **only**

Logging into AWS (Mac)

Logging in with ssh (Mac/Linux)

[\[edit\]](#)

Logging in

- Make sure the permissions on your certificate are secure. Use `chmod` on your downloaded certificate:

```
chmod 600 CBW.pem
```

- To log in to the node, use the `-i` command line argument to specify your certificate:

```
ssh -i CBW.pem ubuntu@cbw##.dyndns.info
```

Use your assigned student #

(where # is your assigned student number. Your student number is the number on the participant list. If your number less than 10, please add 0 in front of it.)

Logging into AWS (Windows)

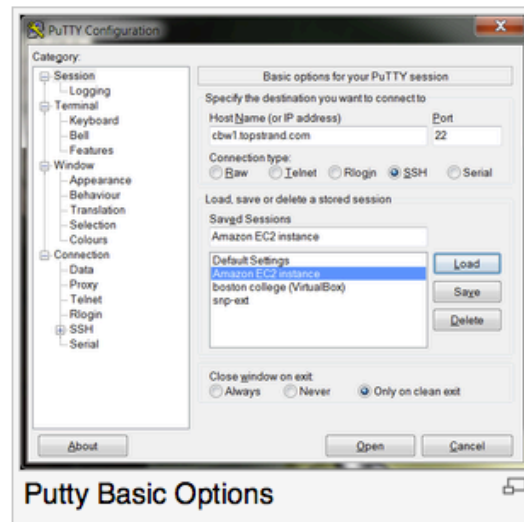
Logging in with Putty (Windows)

[edit]

Logging in

To configure Putty, start Putty and do the following:

- Fill in the "Host name" field with **cbw##.dyndns.info** (where # is your assigned student number. Your student number is the number on the participant list. If your number less than 10, please add 0 in front of it.)



Follow complete instructions on Wiki:

http://bioinformatics.ca/workshop_wiki/index.php/Informatics_for_RNA-seq_Analysis_2015_Workshop_Wiki

Copying files from AWS to your computer

Copying files to your computer

- To copy files from a node, use scp in a similar fashion (in this case to copy a file called nice_alignments.bam):

```
scp -i CBW.pem ubuntu@cbw##.dyndns.info:workspace/nice_alignments.bam .
```

- Everything created in your workspace on the cloud is also available by a web server on your cloud instance. Simply go to the following in your browser:

```
http://cbw##.dyndns.info/
```

So, at this point:

- Your laptop is ready for the workshop
- If it is not, you know where to get the information you need
- You know how to use the wiki for this workshop
- You know where all of the lectures are
- You have read all of the pre-lecture material
- If not, you know where the papers are, and you are a speed reader
- You know how to login to AWS

A much more detailed tutorial on AWS cloud computing...

- https://github.com/griffithlab/rnaseq_tutorial/wiki/Intro-to-AWS-Cloud-Computing

We are on a Coffee Break & Networking Session

