Canadian Bioinformatics Workshops

www.bioinformatics.ca

Creative Commons

This page is available in the following languages:

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

creative commons

Attribution-Share Alike 2.5 Canada

You are free:

to Share — to copy, distribute and transmit the work

to Remix — to adapt the work

Free Cultural Works — **APPROVED FOR**

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

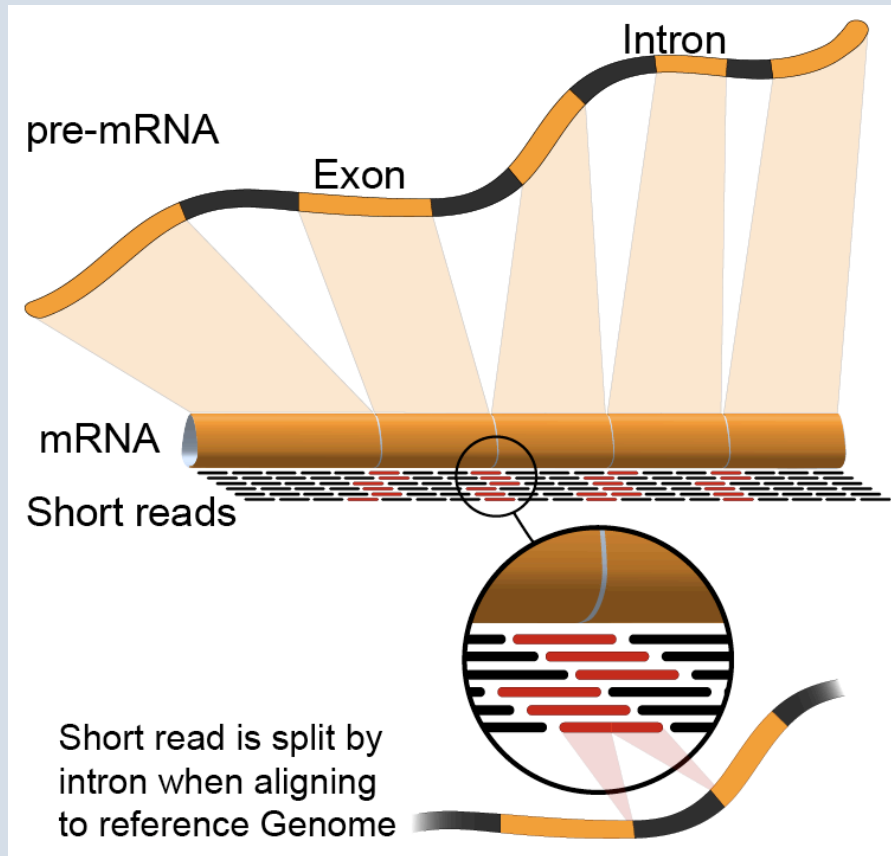Learn how to distribute your work using this licence

Module 0
Introduction to cloud computing
(slides modified with permission from Francis Ouellette)

Zhibin Lu, 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
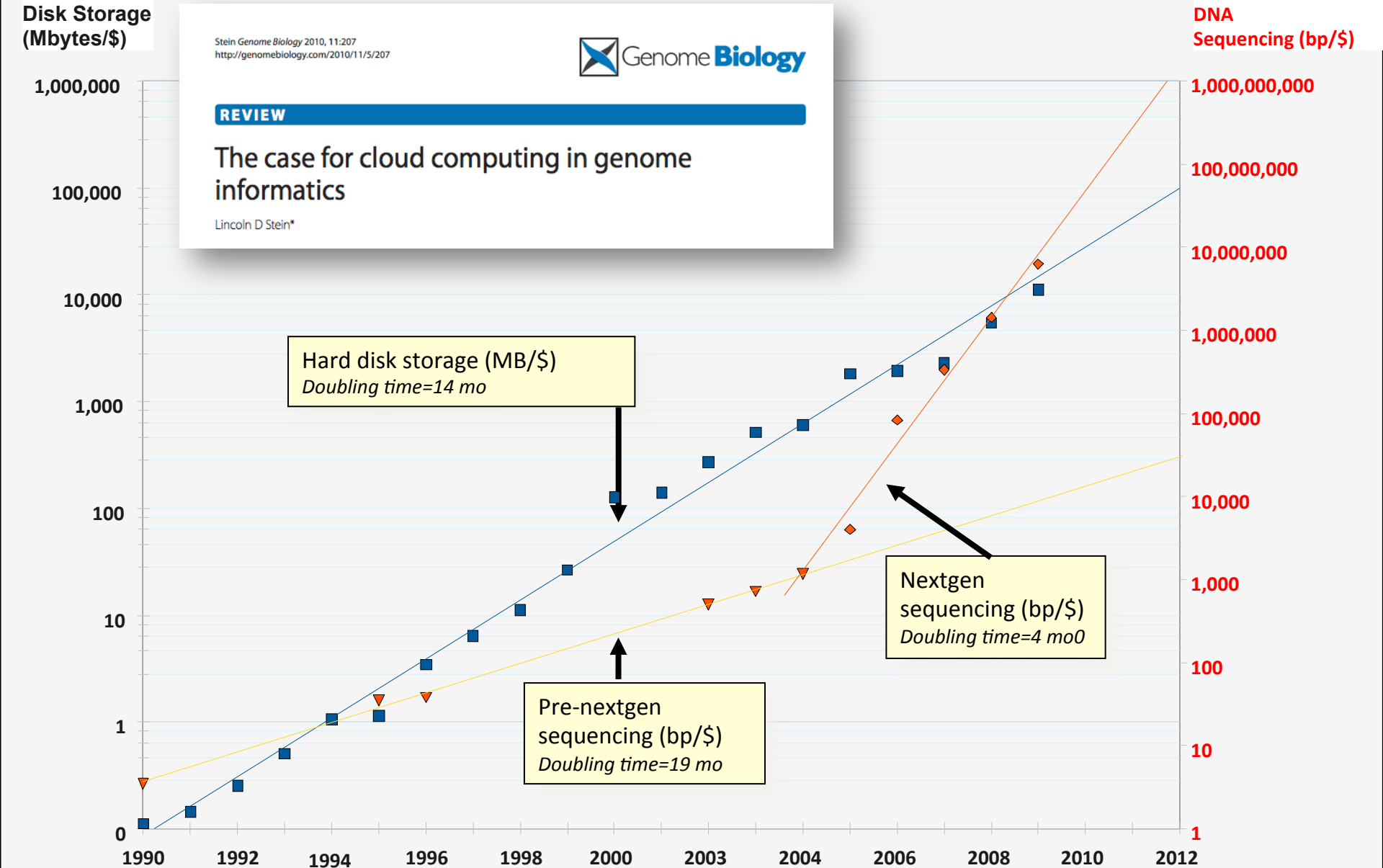
About DNA and computers

- 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/7PVAl

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/

bioinformatics.ca

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'

The CBW Wiki

The RNA-seq wiki

http://www.rnaseq.wiki/

Logging into Amazon AWS

bioinformatics.ca

Opening a 'terminal session'

Mac

Windows
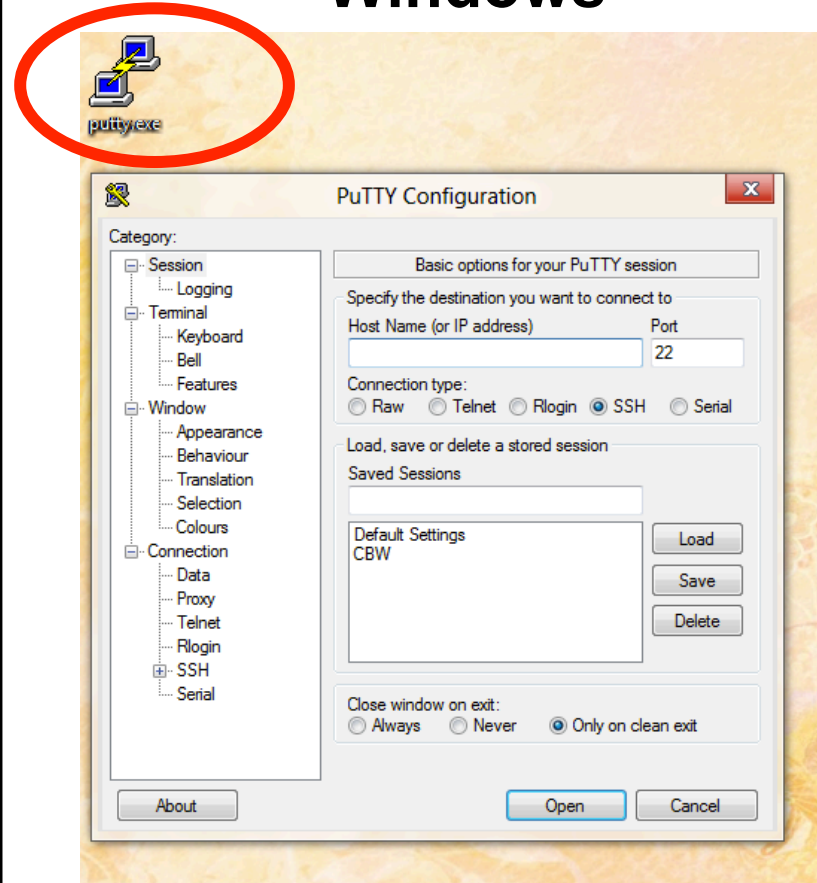
Creating a working directory (Mac/Linux)

bioinformatics.ca

Creating a working directory (Windows)
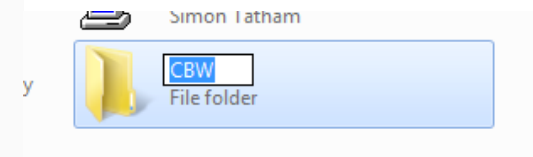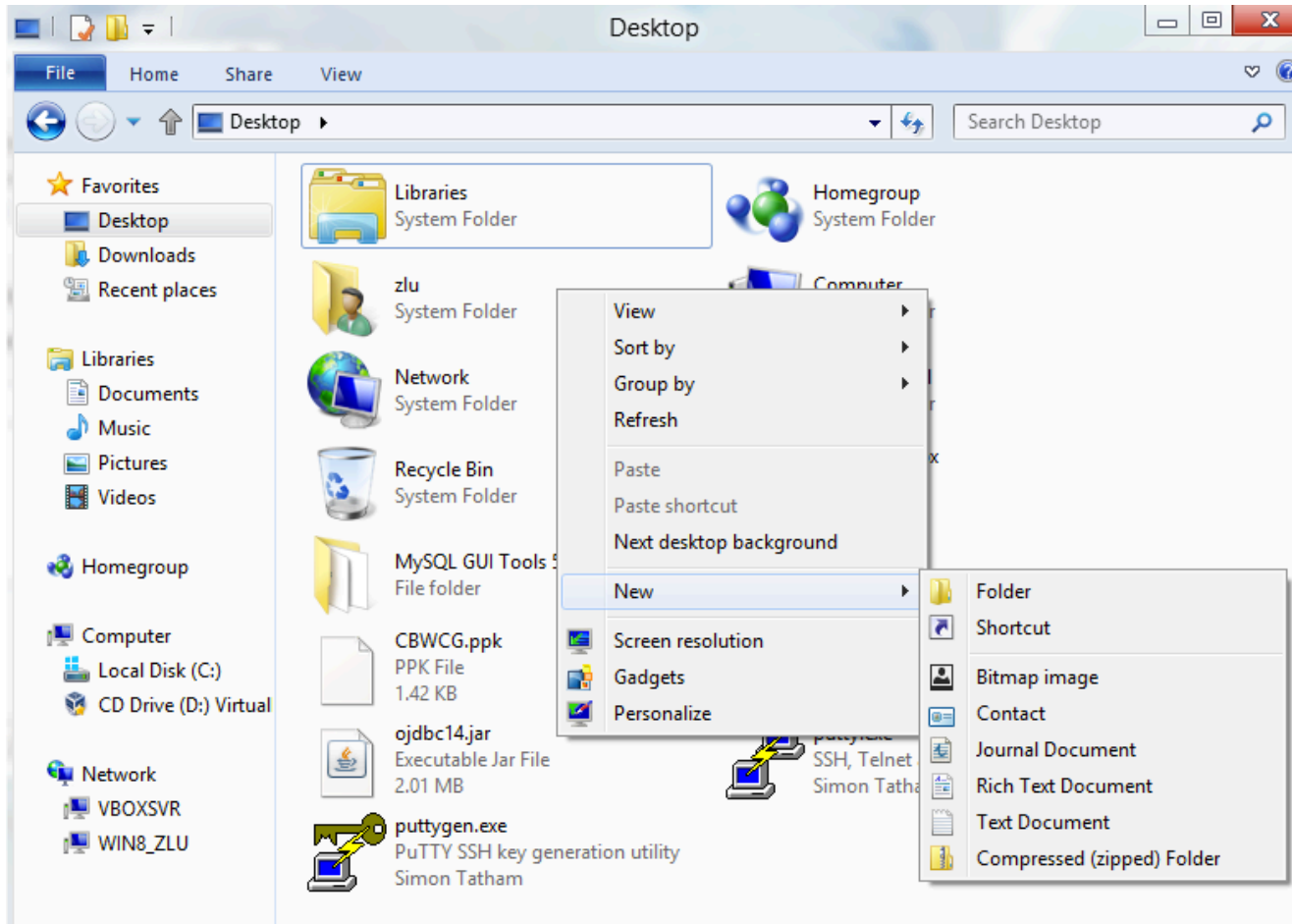
Obtaining your AWS 'key' file from the CBW wiki

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
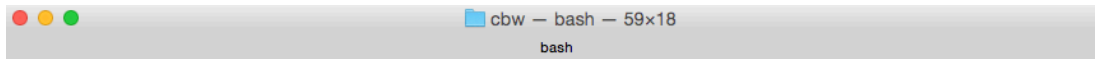
```
cbw — bash — 59×18
bash
mgriffit@mgriffit-macpro:$ cd ~/cbw/
mgriffit@mgriffit-macpro:$ ls -la
total 8
drwxr-xr-x   3 mgriffit  staff    102 Jun  1 18:19 .
drwxr-xr-x+ 85 mgriffit  staff   2890 Jun  1 18:16 ..
-rw-r--r--@  1 mgriffit  staff   1692 Jun  1 18:18 CBW.pem
mgriffit@mgriffit-macpro:$
```

Save key file to your new 'cbw' directory

Viewing the 'key' file once downloaded


```
mgriffit@mgriffit-macpro:$ cd ~/cbw/
mgriffit@mgriffit-macpro:$ 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:$ cat CBW.pem
-----BEGIN RSA PRIVATE KEY-----
MIIEogIBAAKCAQEAjM/lfVDGCmjU7OQldKAO/VX5vdE+MyxWdhDsN9iMy/ALXOVKJh1oCKP6onzT
6tew0t45rQ+qZBNZ+bsnL/zSd3eacbMgKh7ipM6bmNjD/p1mpyTn87V+8aH5vC9uOEiiH0K+le0Q
PhUWQ15GsYZifQUrhxueB00R7wil+ZA4yqHiryw2r16+X7Z+KrXTRN/3ArD0i5iYaNRb41ztXCOs
G6wHgGELE986/2E/2ruM2O1GtQ+OX/TYgBfrSXoC0wUY2okMTsVGjigtfwV33zs9i/hiNsE0B2uZ
D/353kq+bZ4ZZbL8akJzfn2RqxdwAWB4jSAF7XMAz+3I0VQHMgoySQIDAQABAoIBAGMgl7W4fzIk
yoBZu+R2KRc+xylRfDAaLmO/VguabsPCoL0szR435Vzv0ylAJw9T+C76MjAaryD0XornrZairKdd
k2QnbaU5SOQsbm0TbCZ6OCHCYWe3hT4lgOFxom4M89R98B0rXhvWTG3wmGOw7vSEBzh9wKnzH4Te
2bjuRoyegzJFLAX18kOdmenwwWDtugRaNHVj49zx32BnK2T7aVW1nOdwJe14poai7z67hG5sPdie
XufcwNqOCKpJKIWTet00EJTCnfo+Dxg7Ukb937fZaq27EZ1wwz+/K/Zl/YHgDpHD2udQG8sYlO3t
lVhB35v/VfSjtU7y7PHPJKfGzAECgYEA/ti+jrnmgQnhCVUVFr2YpajFbhJoFFSF39tdb9emMxSh
Mbsv3Bn5bEX7U/qO1PpxFmqM+jM/ZhO91z1J1nQWiOQ+HajrcUDgDW1gsHmzbPv9BM90/N/xipnO
ibMB2guq0UXRNIhIWuc8jweAw69OndOh+dKZpStQGV+fr7PPCNkCgYEAjXMJNbmIFfPGpMddHH3I
syHJeSiTDk6D3L7ZvMA/Ea4iTZi6rOza0iVvvlJgSE4UPy3mMZfY3doiv/hhQO4/IPAKImcQo+ml
pgyGfZtUunf12K4/CPw6LLuv+PMesxDGC5w0zzvlNx0chWv0DXV/aS98+JhrqIM7pkYAZCZ/DvEC
gYBrxrYwHiaBg/ALxl043k3kIU/9kQM7NdO4LEmDtyuuhpb1NTi0WLfHNoFnrpshLxkMZdthJLM3
NDohAOZtL4MoOX6bhIityMX4Z71wZMeg26DRm/rZxsub1hgm2/F4iV6gSmqDUO9ookpWItIt78HL
tWjXAl3AIGUrlekuoVh5YQKBgBrhvONX8jPx01Oni9uJUCYzCp4bBu+WHltKbgkAJDxQ8rX6i6en
j+KWOMlb0G6NYC2tpGzSZeWruiTvqFP1s0uoMtTc732nPtLy2HBwn8acW72lR9eGT+wGc2gFlFUF
oA2+zw5/XI1/29bDcGtKZD9yRfbcyMEKx8BO124xHr6RAoGAB77iIJjR1TAMCydk7QxuJwUEGquS
0I3DowhJpjpyQ4sgRFGbrmQVUpKoC5++KbuLcpL/M+3vXdRMr4UZSMLnbLbF2U+u1XKCrh5o66mp
N2fZ0lFWHzjQfo5FQQU4+c+EeBlI8pWrBiiN2B7IFsDTlzSG9hiwcUHvdbHR+ZKLMbE=
mgriffit@mgriffit-macpro:$ ▮
```

Changing file permissions of your 'key' file (Mac/Linux)

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 ...)
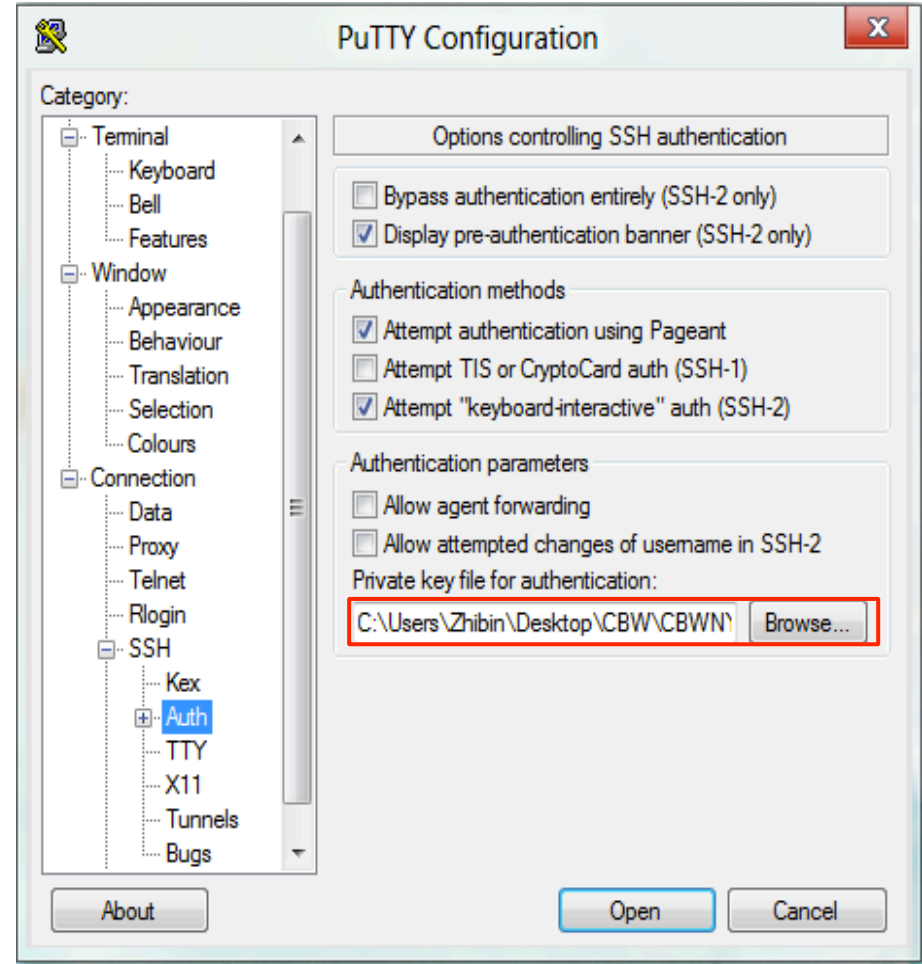
So, when you have:

chmod 600 <file name>

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

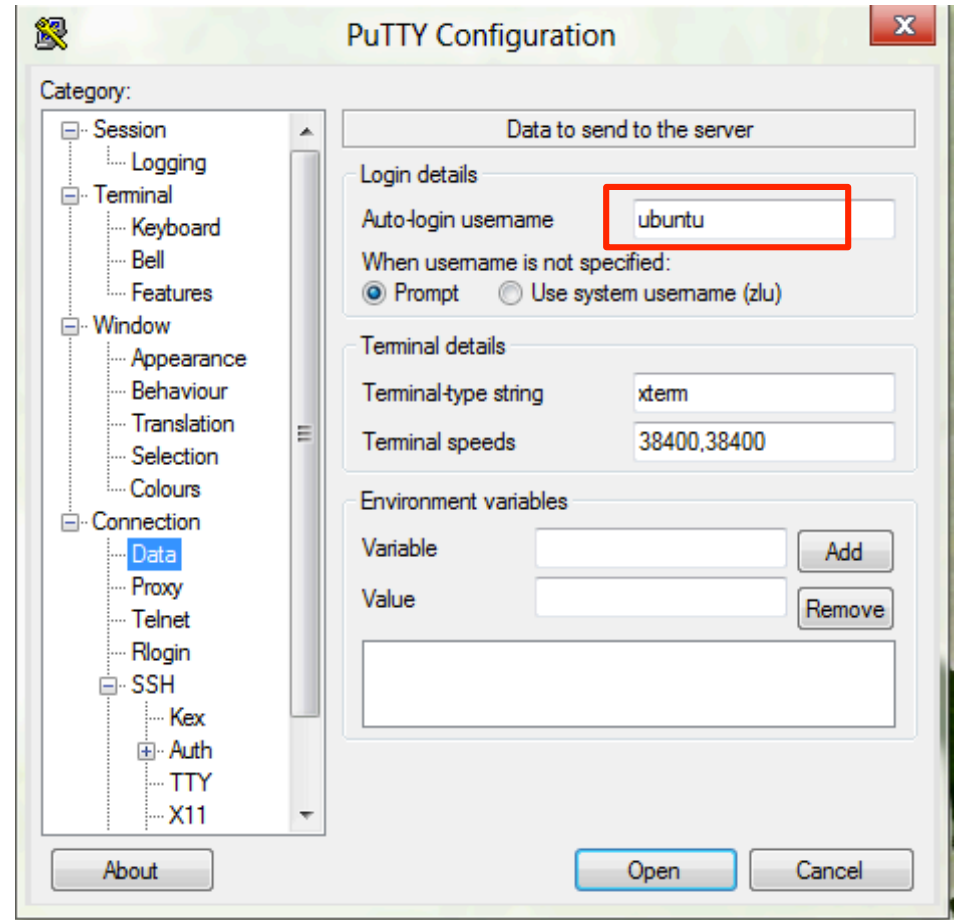# Logging into AWS

Mac/Linux

ssh -i CBWNY.pem

Windows

Logging into AWS

Mac/Linux
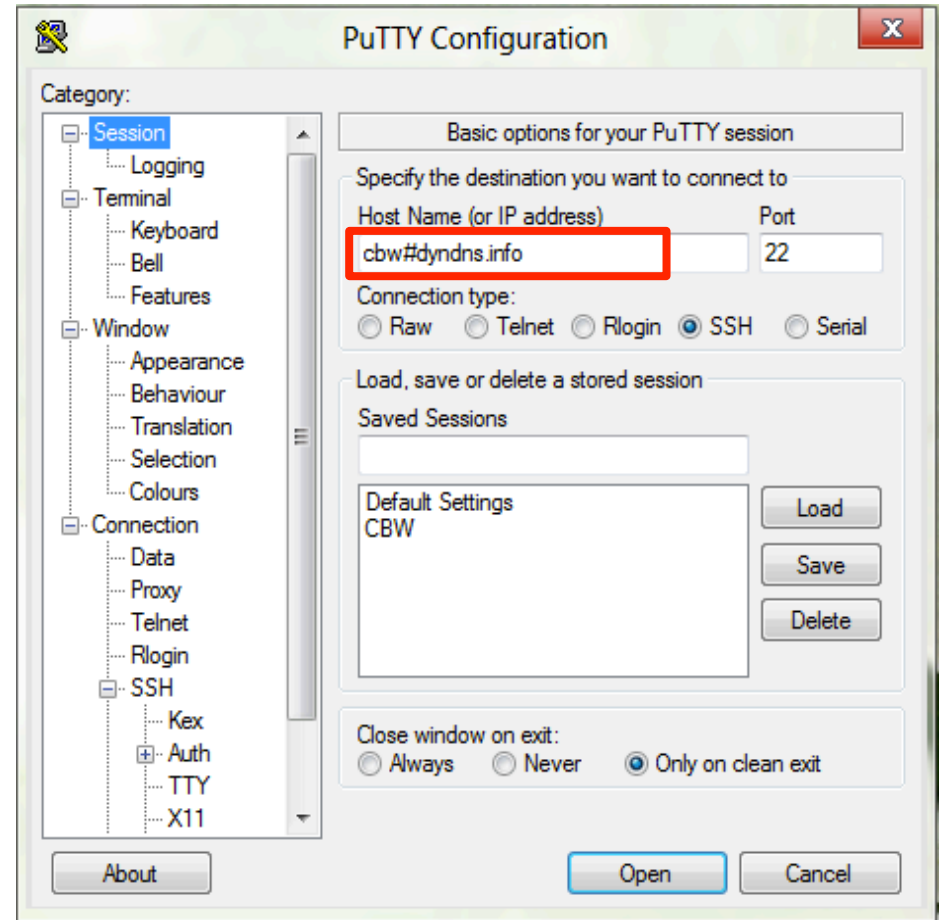
ssh -i CBWNY.pem ubuntu

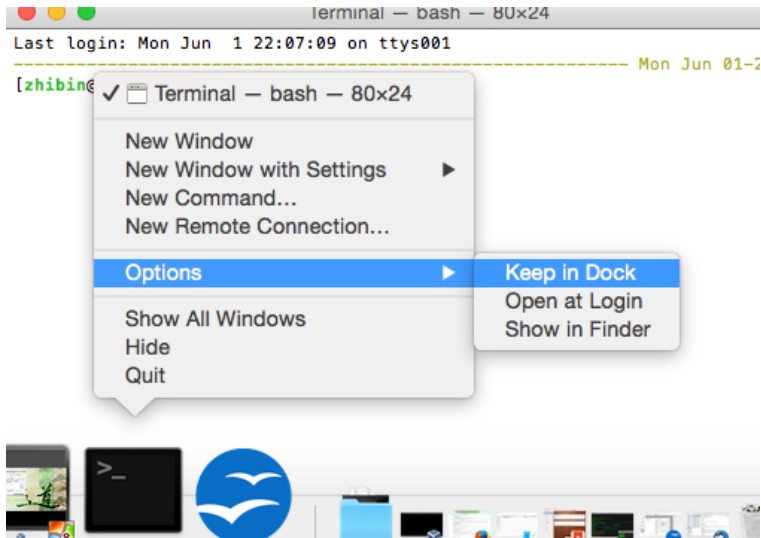## Windows

Logging into AWS

Mac/Linux

ssh -i CBWNY.pem
ubuntu@cbw#.dyndns.info
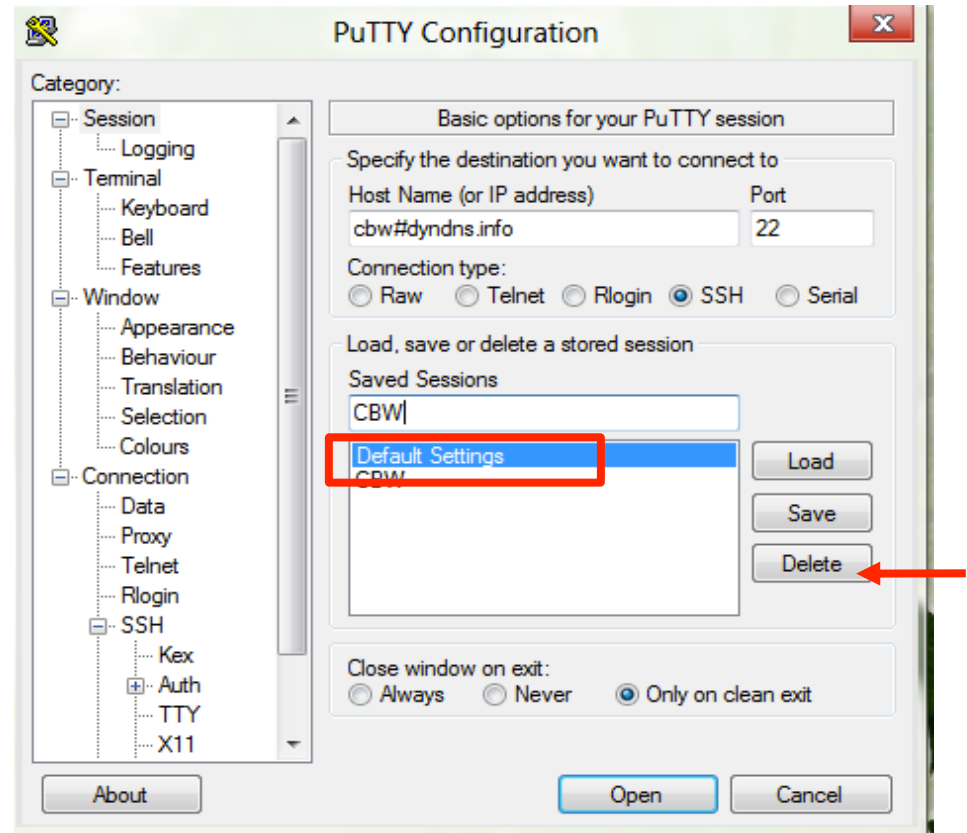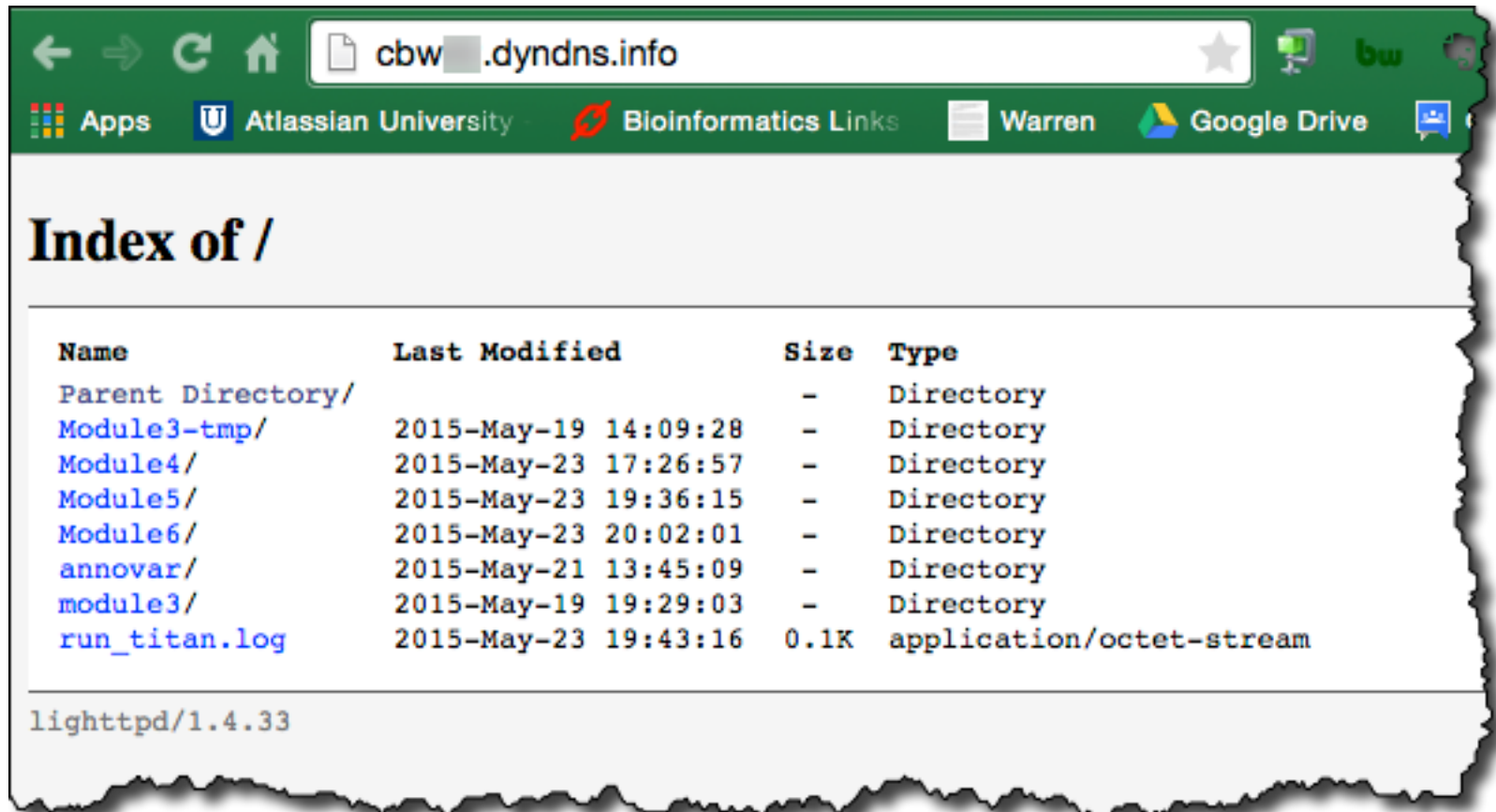
Windows

Logging into AWS

Mac/Linux

Windows

From now on, just double click 'CBW' to login

Copying files from AWS to your computer (using a web browser)

http://cbw#.dyndns.info/

Copying files from AWS to your computer (at the command line)

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

bioinformatics.ca