



Canadian Bioinformatics Workshops

www.bioinformatics.ca

This page is available in the following languages:

Afrikaans বাংলাৰাখী 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 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 srpski (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

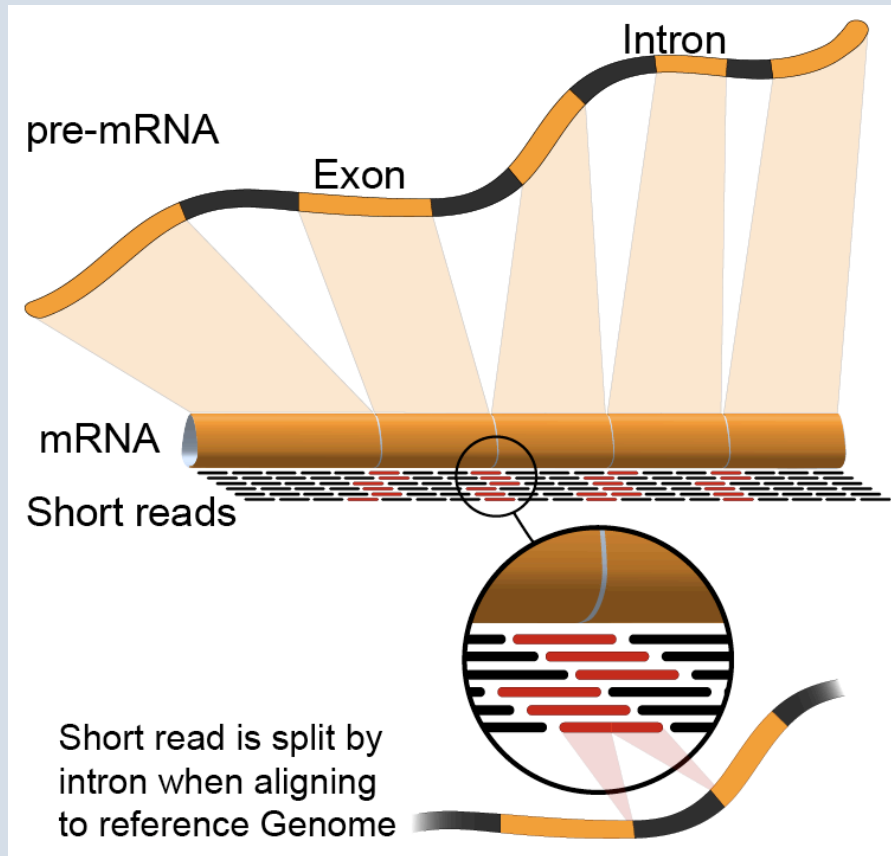
Introduction to cloud computing

(slides modified with permission from Francis Ouellette)

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

Disk Storage
(Mbytes/\$)

DNA
Sequencing (bp/\$)

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*

1,000,000

1,000,000,000

100,000

100,000,000

10,000

10,000,000

1,000

1,000,000

100

100,000

10

10,000

1

1,000

100

10

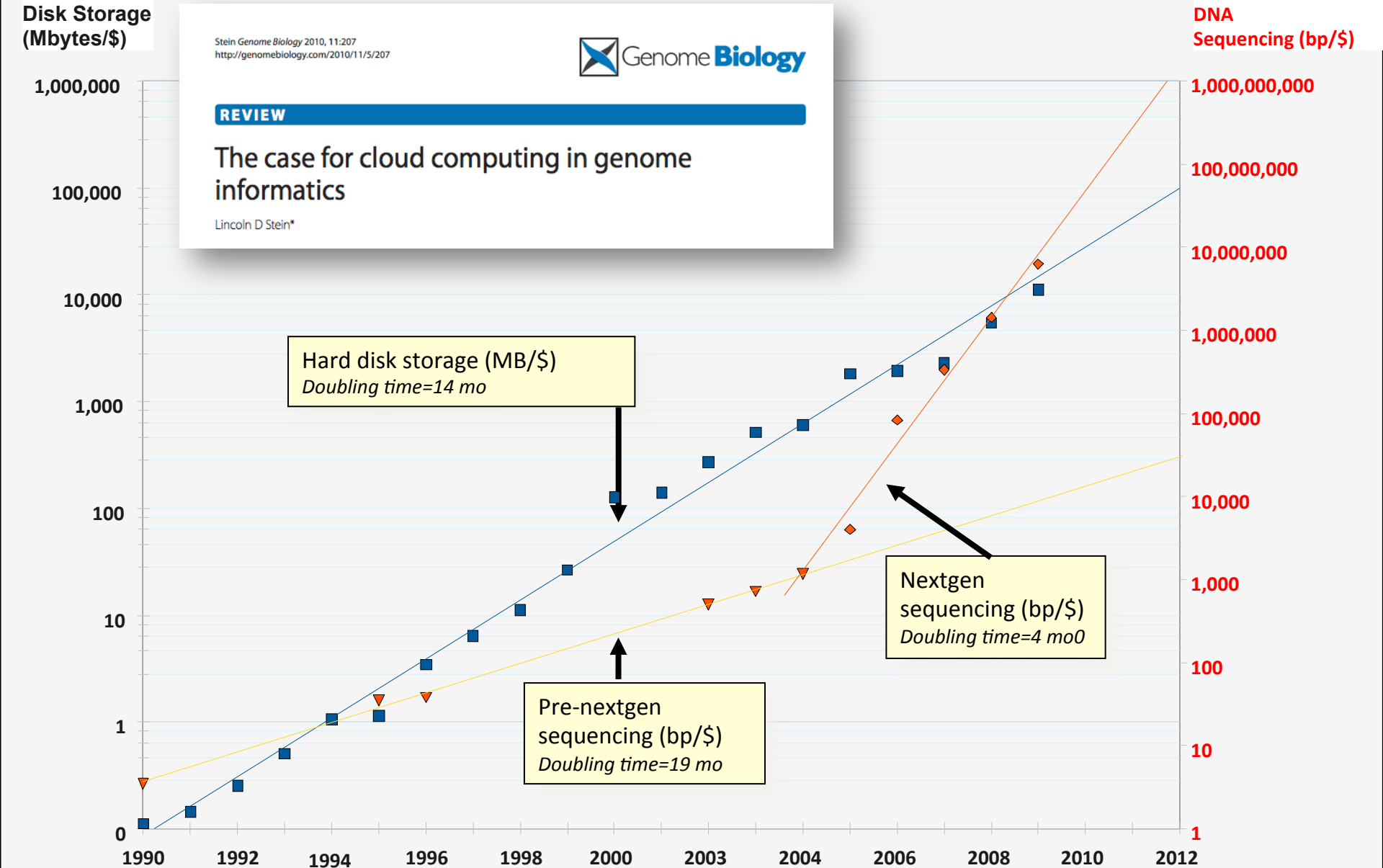
1

1990 1992 1994 1996 1998 2000 2003 2004 2006 2008 2010 2012

Hard disk storage (MB/\$)
Doubling time=14 mo

Nextgen
sequencing (bp/\$)
Doubling time=4 mo0

Pre-nextgen
sequencing (bp/\$)
Doubling time=19 mo



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

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

Search

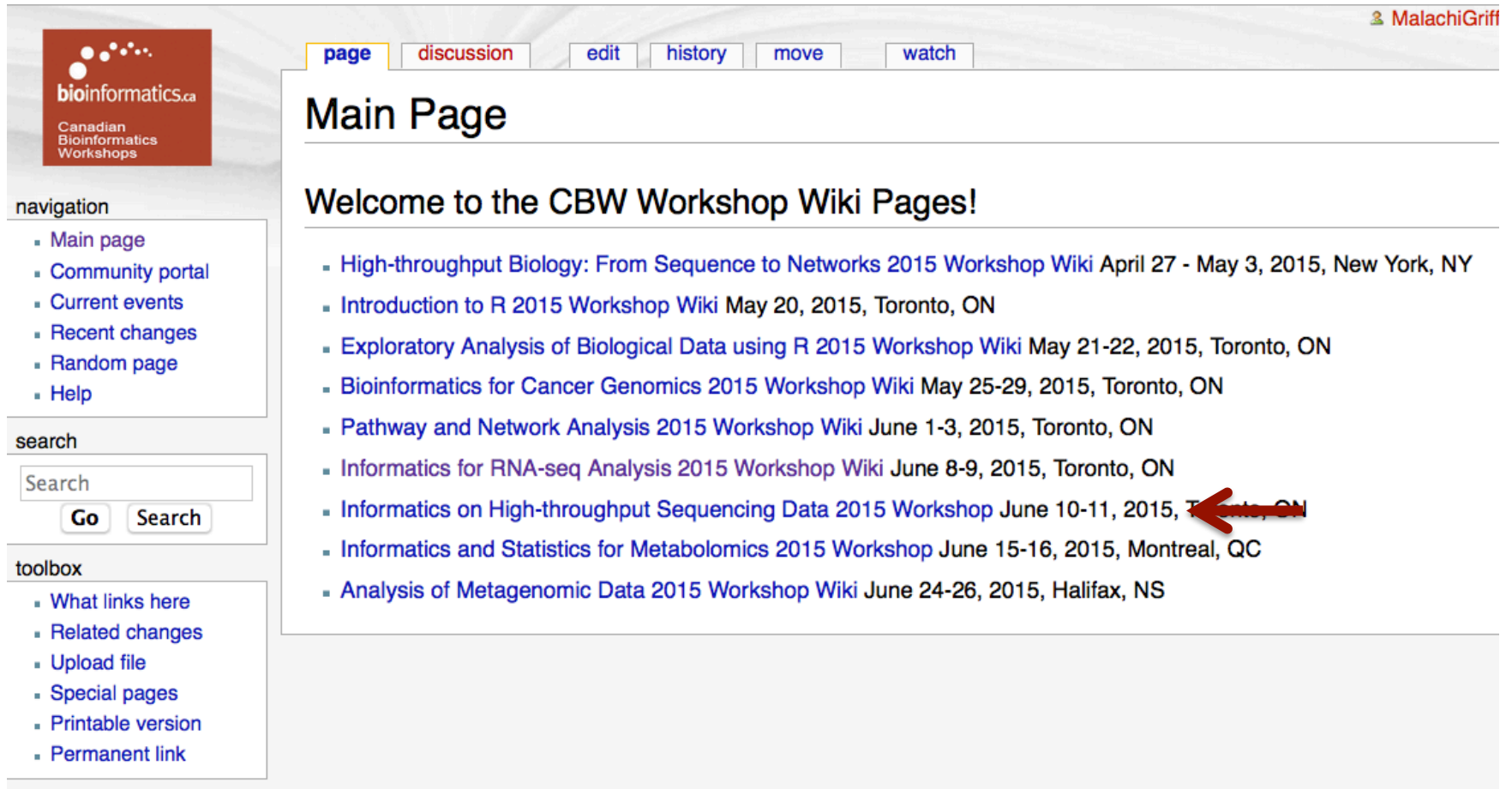
Go Search

toolbox

- [Special pages](#)

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

The CBW Wiki



The screenshot shows the main page of the CBW Wiki. At the top right, the user 'MalachiGriff' is logged in. Below the title, there are navigation tabs for 'page', 'discussion', 'edit', 'history', 'move', and 'watch'. The main heading is 'Main Page', followed by a welcome message: 'Welcome to the CBW Workshop Wiki Pages!'. A list of workshop events is displayed, including 'High-throughput Biology: From Sequence to Networks 2015 Workshop Wiki' and 'Informatics on High-throughput Sequencing Data 2015 Workshop'. A red arrow points to the 'Informatics on High-throughput Sequencing Data 2015 Workshop' entry. On the left side, there is a navigation menu with links like 'Main page', 'Community portal', and 'Current events'. Below that is a search box with 'Go' and 'Search' buttons. At the bottom left, a toolbox contains links for 'What links here', 'Related changes', 'Upload file', 'Special pages', 'Printable version', and 'Permanent link'.

bioinformatics.ca
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

Search

Go 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

- 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

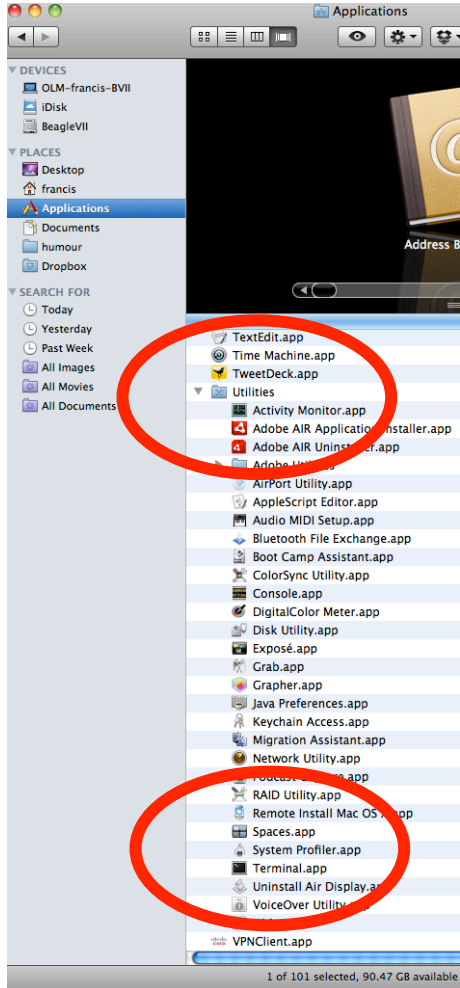
Clone this wiki locally

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

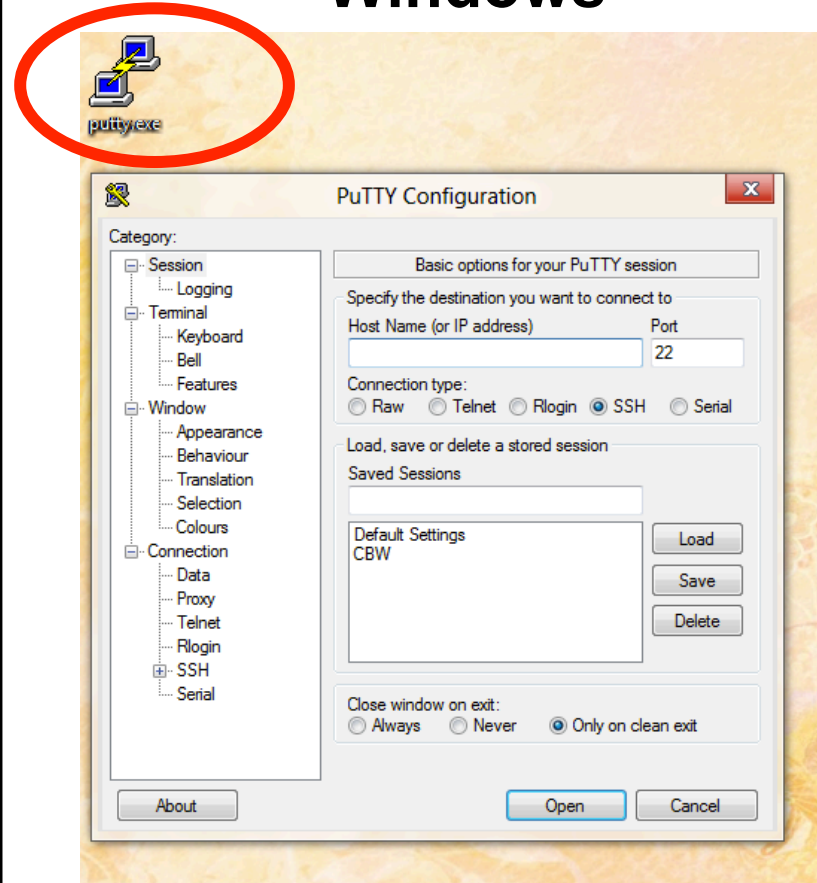
Logging into Amazon AWS

Opening a 'terminal session'

Mac



Windows

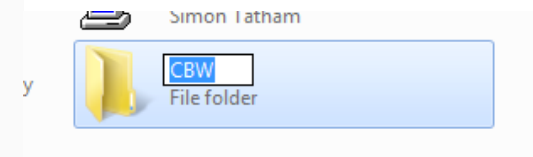
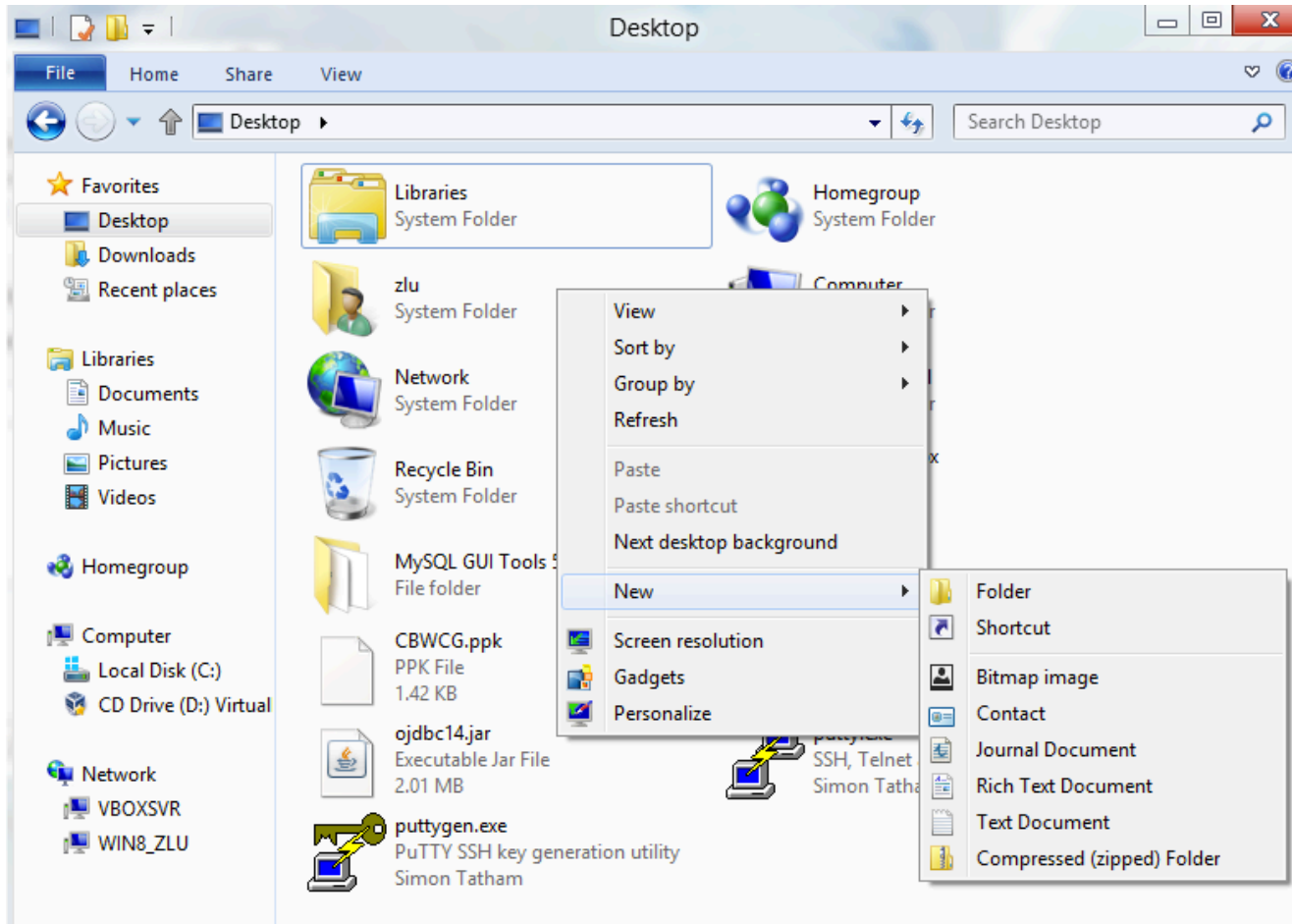


Creating a working directory (Mac/Linux)

```
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$ █
```

Creating a working directory (Windows)



Obtaining your AWS 'key' file from the CBW wiki

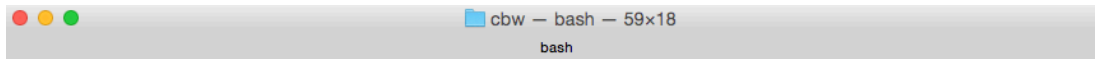
Logging into the Amazon cloud

[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



```
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

```
cbw - bash - 79x30
bash
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+8aH5vC9uOEiiH0K+le0Q
PhUWQ15GsYZifQRhxueB00R7wil+ZA4yqHiryw2r16+X7Z+KrXTRN/3ArD0i5iYaNRb41ztXC0s
G6wHgGELE986/2E/2ruM201GtQ+OX/TYgBfrSXoC0wUY2okMTsVGjigtfwV33zs9i/hiNsE0B2uZ
D/353kq+bZ4ZZbL8akJzfn2RqxdwAWB4jSAF7XMAz+3I0VQHMGoySQIDAQABAoIBAGMgl7W4fzIk
yoBZu+R2KRc+xyLrfdAaLm0/VguabsPCoL0szR435Vzv0ylAJw9T+C76MjAaryD0XornrZairKdd
k2QnbaU5S0Qsbm0TbCZ60CHCYWe3hT4lgOFxom4M89R98B0rXhvWTG3wmG0w7vSEBzh9wKnzH4Te
2bjuRoyegzJFLAX18k0dmenwWdttugRaNHVj49zx32BnK2T7aVW1n0dwJe14poai7z67hG5sPdie
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/F4iV6gSmqDU09oookPWItIt78HL
tWjXAl3AIGUrlekuoVh5YQKBgBrhvONX8jPx010ni9uJUCYzCp4bBu+WHltKbgbKAJDxQ8rX6i6en
j+KWOMlb0G6NYC2tpGzSZeWruITvqFP1s0uoMtTc732nPtLy2HBwn8acW72lR9eGT+wGc2gFLUF
oA2+zw5/XI1/29bDcGtKZD9yRfbcyMEKx8B0124xHr6RAoGAB77iIJjR1TAMCydk7QxuJwUEGquS
0I3DowhJpjpyQ4sgRFGbrmQVUpKoC5++KbuLcpL/M+3vXdRm4UZSMLnbLbF2U+u1XKCrh5o66mp
N2fZ0LFWhzjQfo5FQQU4+c+EeBlI8pWrBiIN2B7IFsDTLzSG9hiwcUHvdbHR+ZKLMbE=
mgriffit@mgriffit-macpro:~/cbw$
```

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>

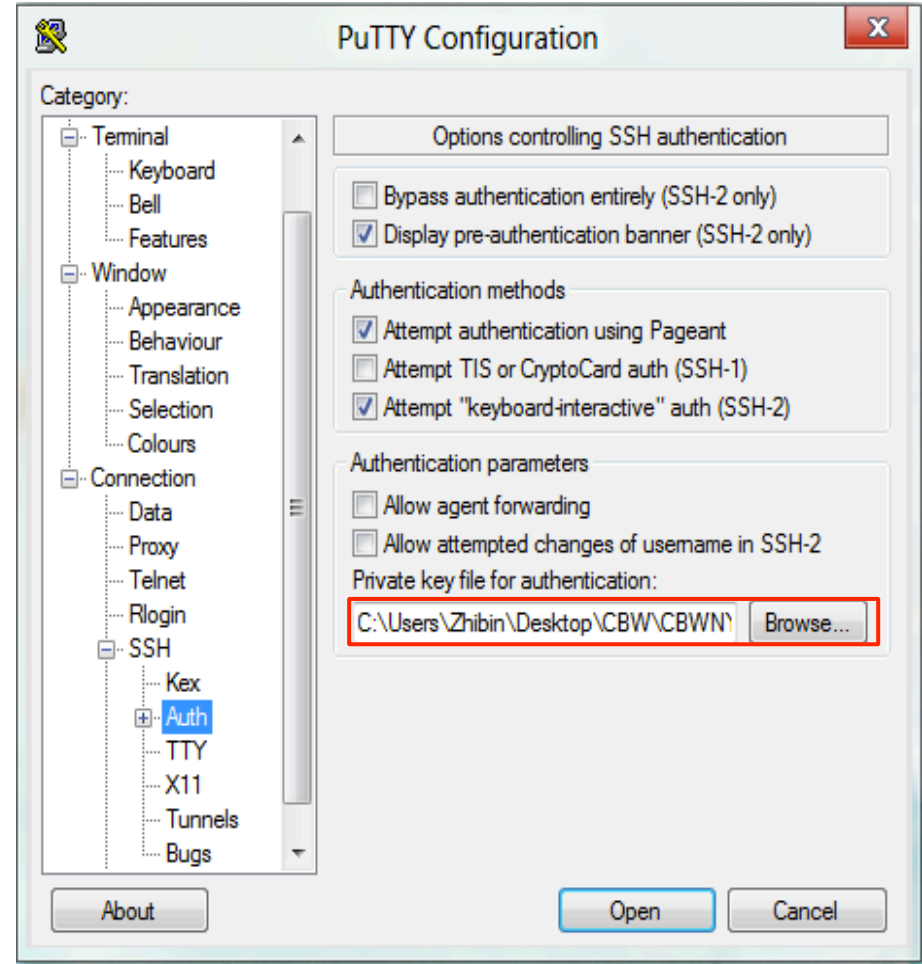
It is "rw" for the the file owner **only**

Logging into AWS

Mac/Linux

```
ssh -i CBWNY.pem
```

Windows

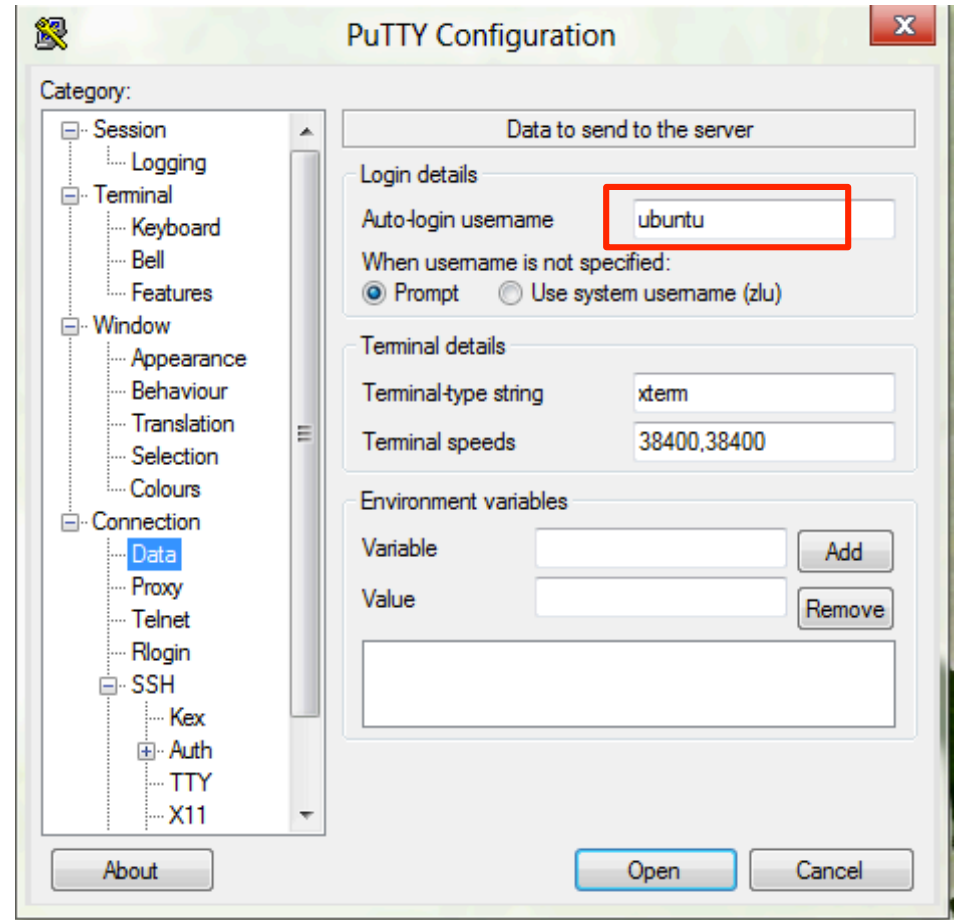


Logging into AWS

Mac/Linux

```
ssh -i CBWNY.pem ubuntu
```

Windows

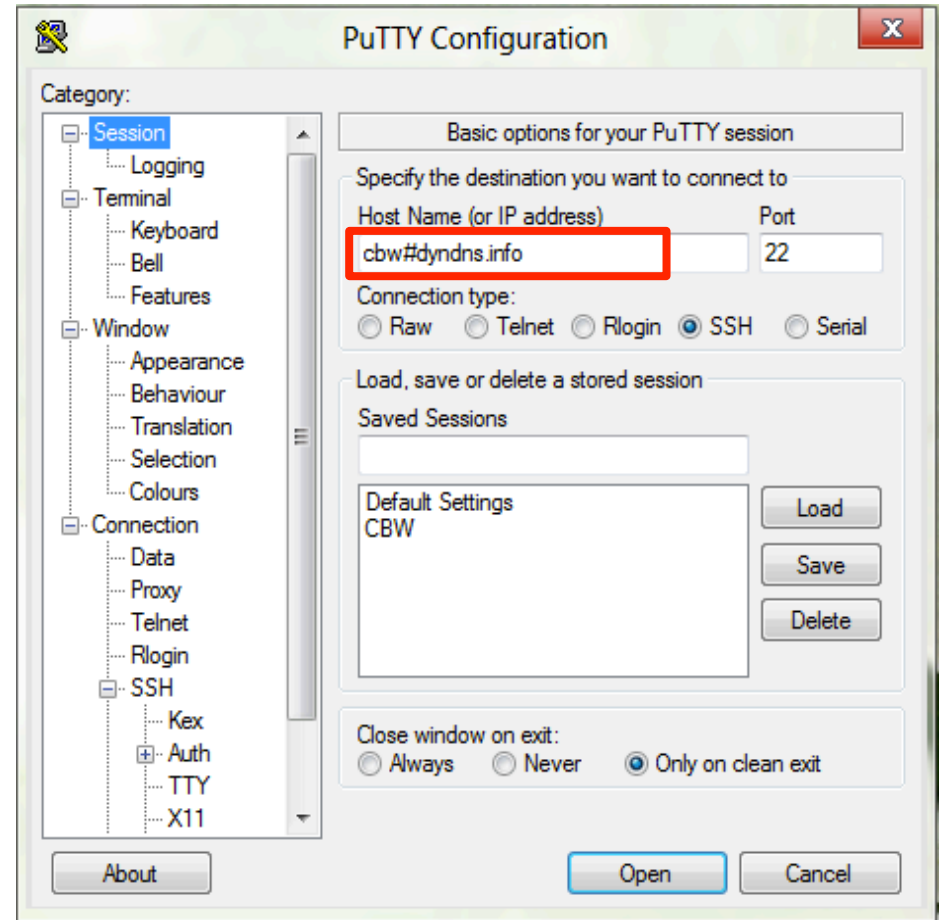


Logging into AWS

Mac/Linux

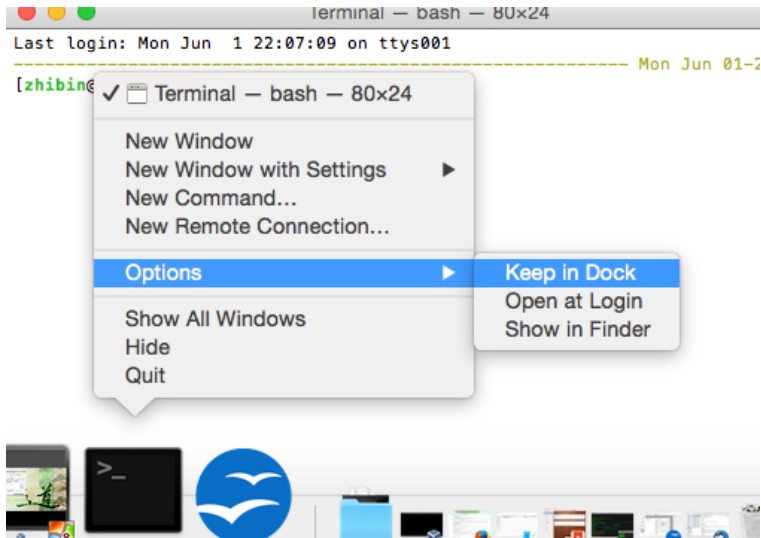
```
ssh -i CBWNY.pem  
ubuntu@cbw#.dyndns.info
```

Windows

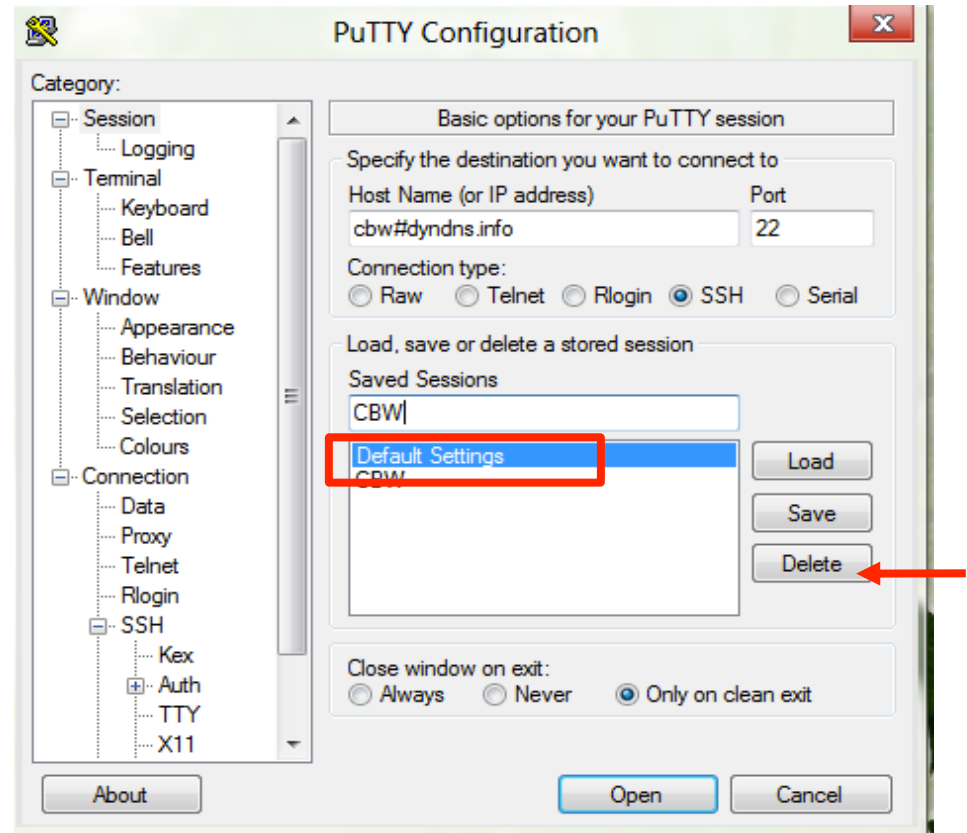


Logging into AWS

Mac/Linux

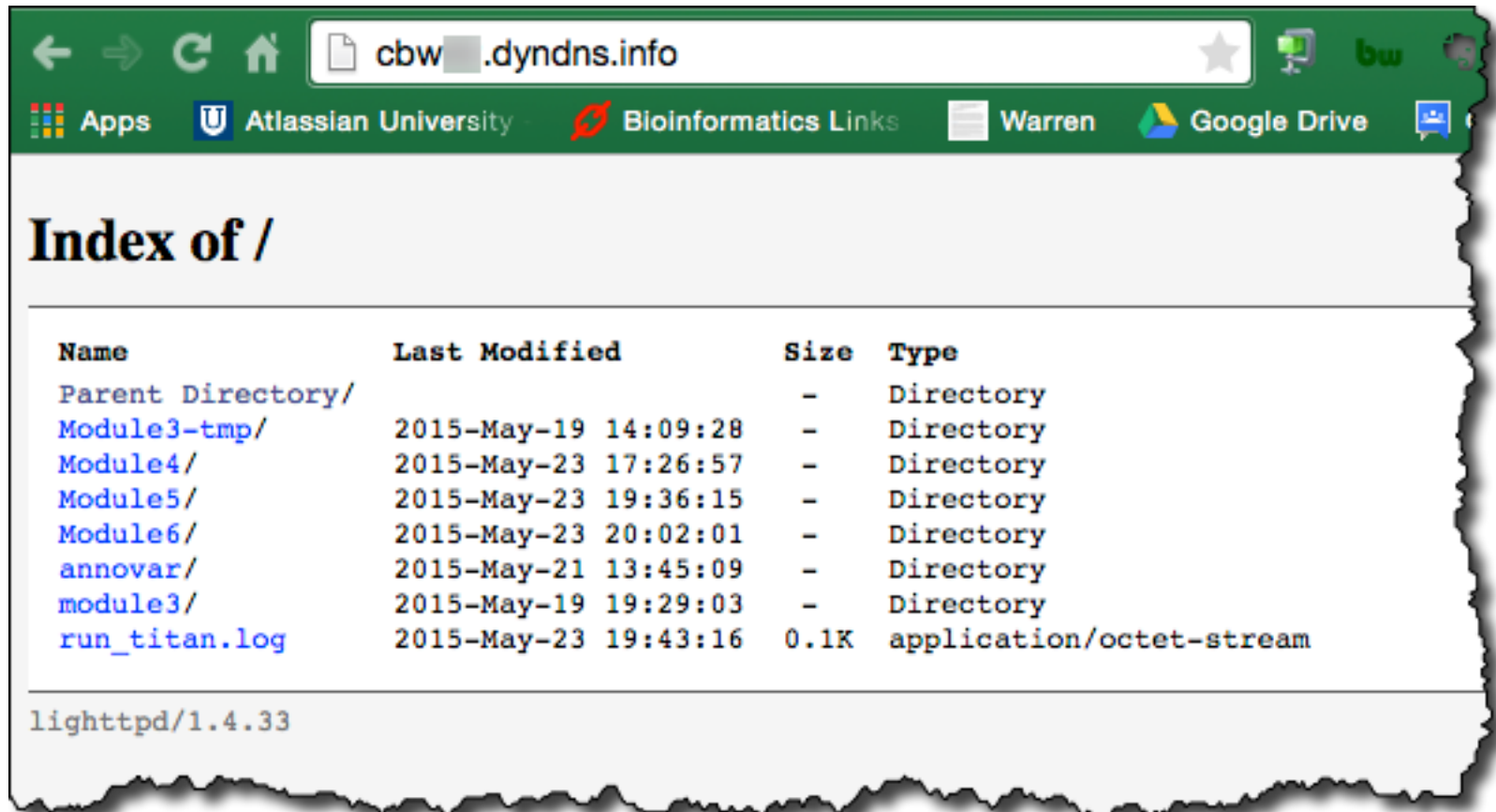


Windows



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

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



The screenshot shows a web browser window with a green address bar containing the URL `cbw#.dyndns.info`. The browser's toolbar includes navigation icons (back, forward, refresh, home) and several bookmarks: Apps, Atlassian University, Bioinformatics Links, Warren, and Google Drive. The main content area displays an "Index of /" directory listing. The listing is a table with columns for Name, Last Modified, Size, and Type. The entries include several directories (Module3-tmp/, Module4/, Module5/, Module6/, annovar/, module3/) and a file named `run_titan.log` with a size of 0.1K. At the bottom of the page, the text `lighttpd/1.4.33` is visible.

Name	Last Modified	Size	Type
Parent Directory/		-	Directory
Module3-tmp/	2015-May-19 14:09:28	-	Directory
Module4/	2015-May-23 17:26:57	-	Directory
Module5/	2015-May-23 19:36:15	-	Directory
Module6/	2015-May-23 20:02:01	-	Directory
annovar/	2015-May-21 13:45:09	-	Directory
module3/	2015-May-19 19:29:03	-	Directory
run_titan.log	2015-May-23 19:43:16	0.1K	application/octet-stream

lighttpd/1.4.33

<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](https://github.com/griffithlab/rnaseq_tutorial/wiki/Intro-to-AWS-Cloud-Computing)

We are on a Coffee Break & Networking Session

