Canadian Bioinformatics Workshops

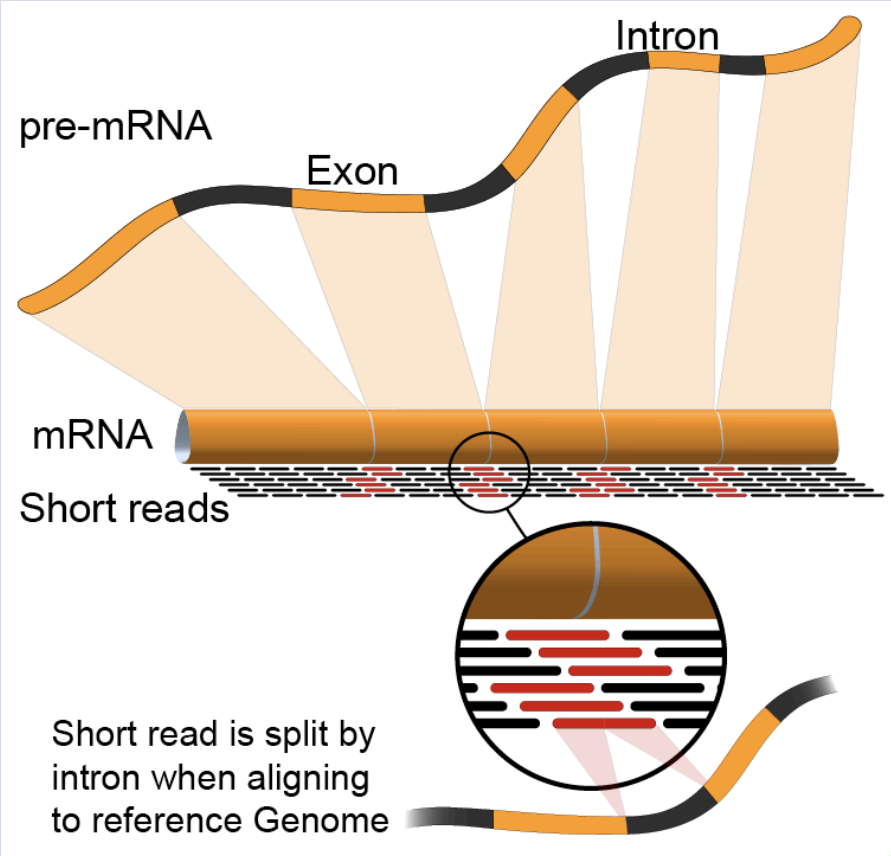
www.bioinformatics.ca

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



bioinformatics.ca

Learning objectives of the course

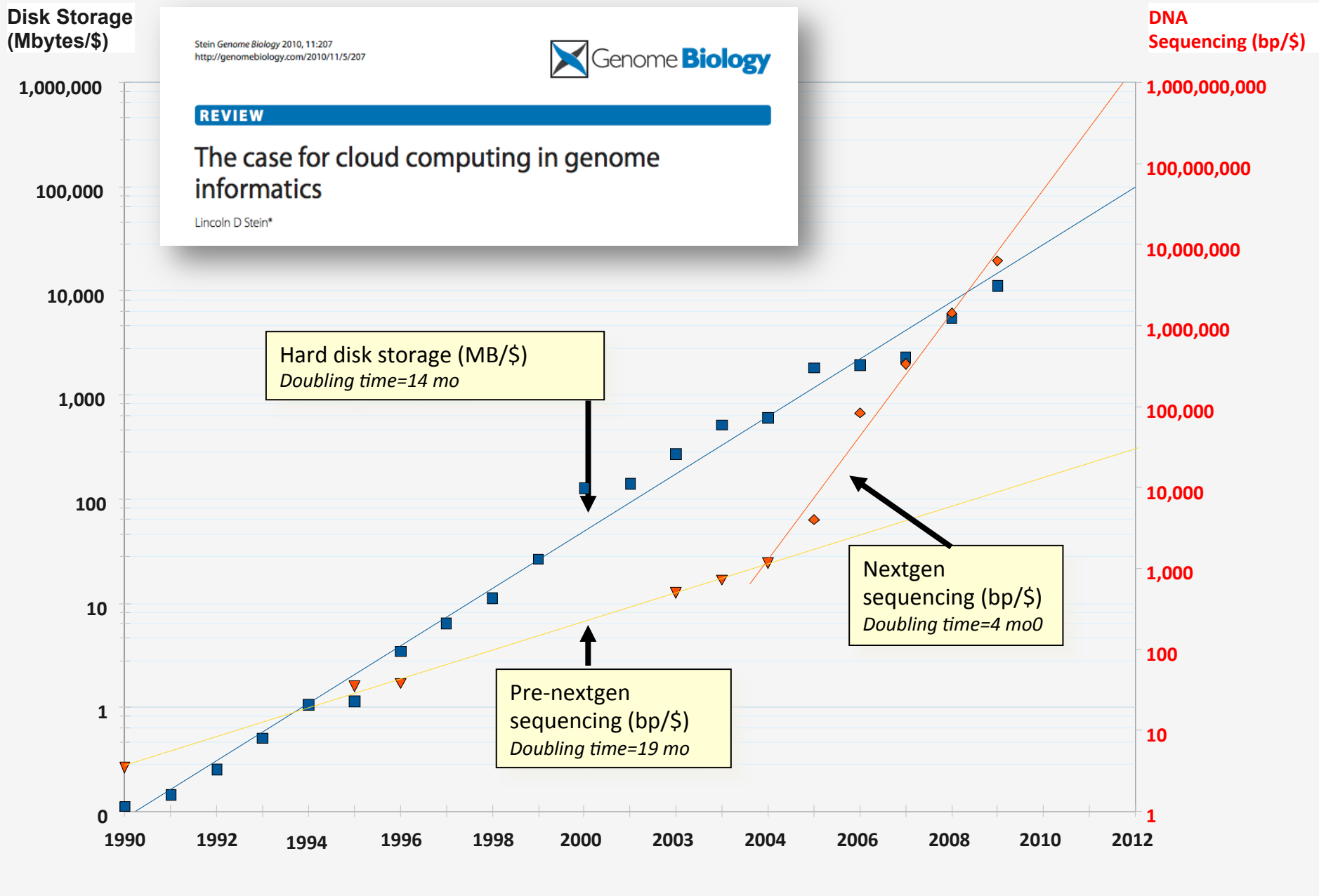
- **Module 0: Introduction to cloud computing**
- Module 1: Introduction to RNA sequencing
- Module 2: RNA-seq alignment and visualization
- Module 3: Expression and Differential Expression
- Module 4: Isoform discovery and alternative expression

- 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



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

Genome Biology

REVIEW

The case for cloud computing in genome informatics

Lincoln D Stein*

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:

amazon web services™

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/

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

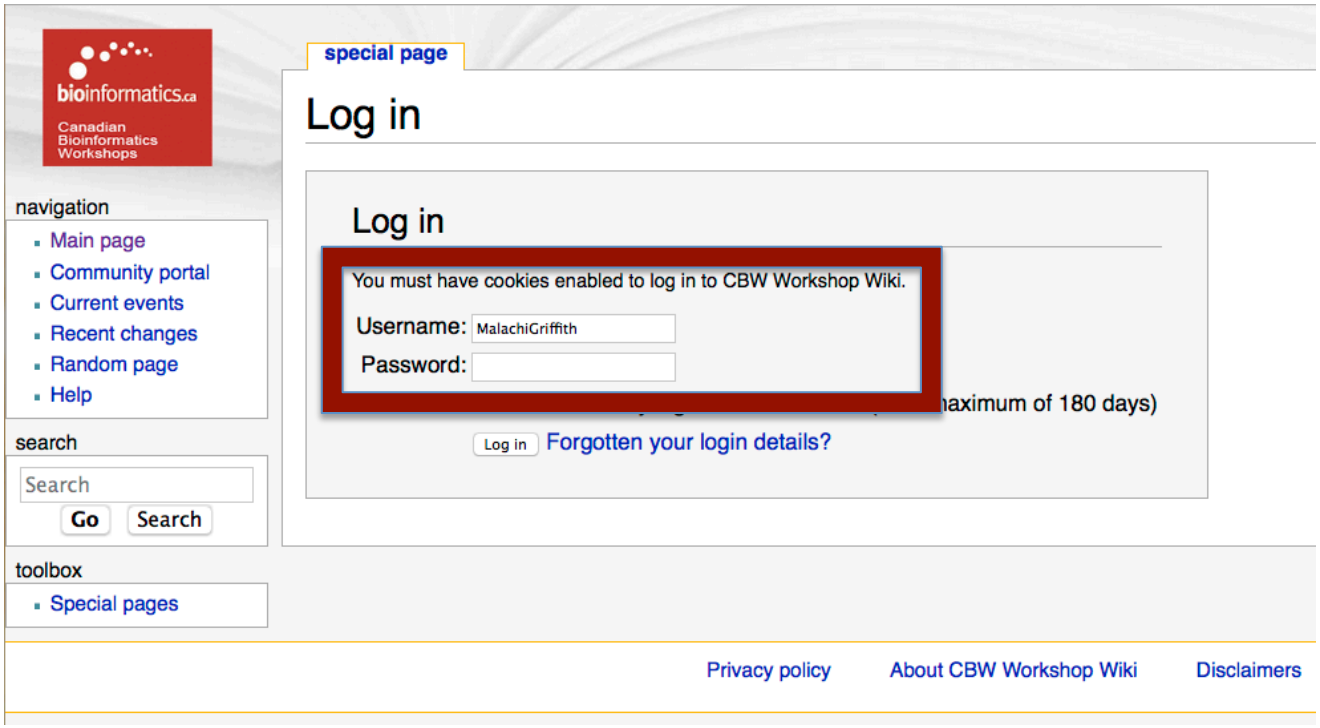
http://aws.amazon.com/console/

For this workshop: all on Wiki!

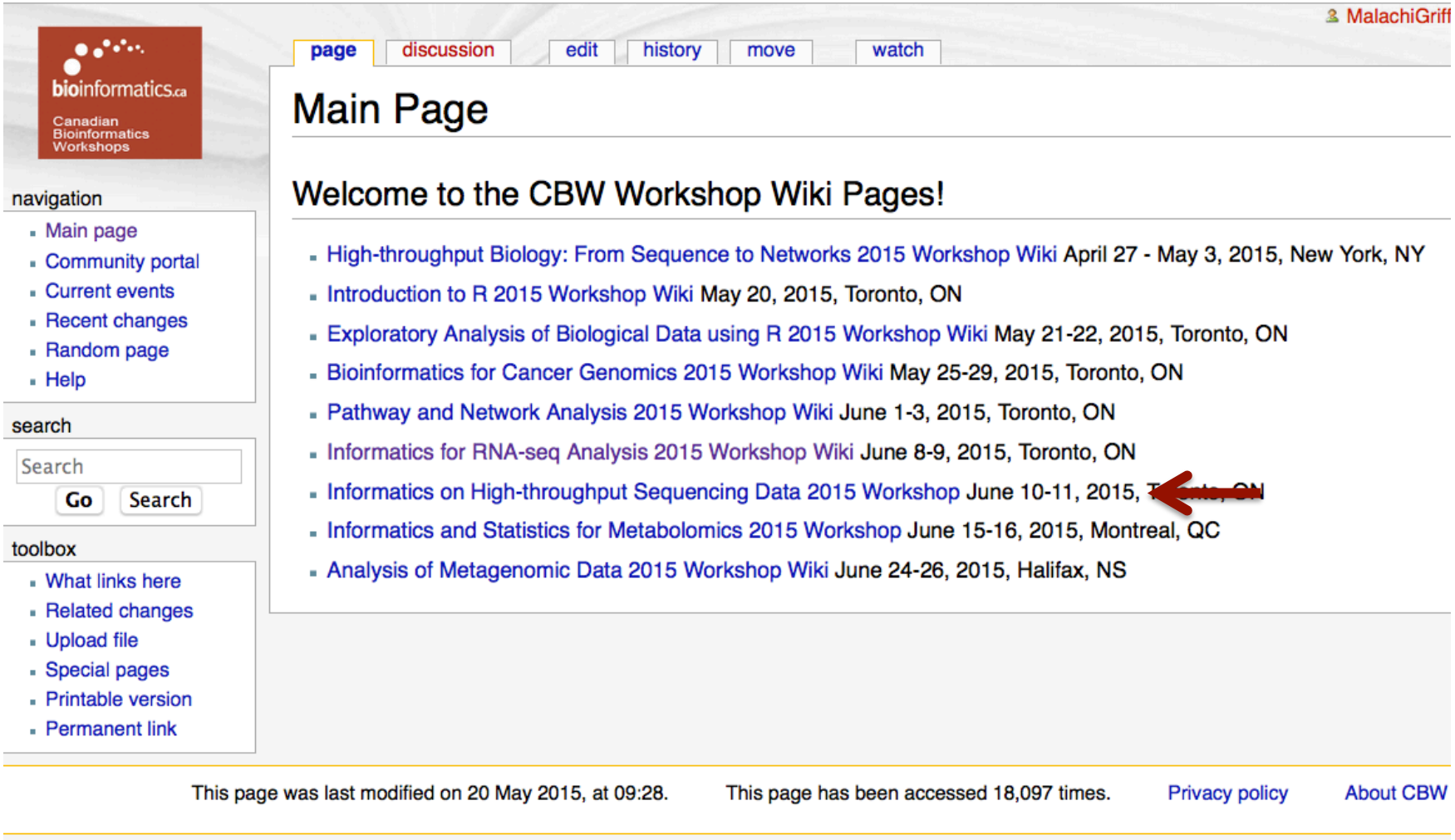
http://bioinformatics.ca/workshop_wiki/index.php/

Login: FirstnameLastname

Password: 'guest'

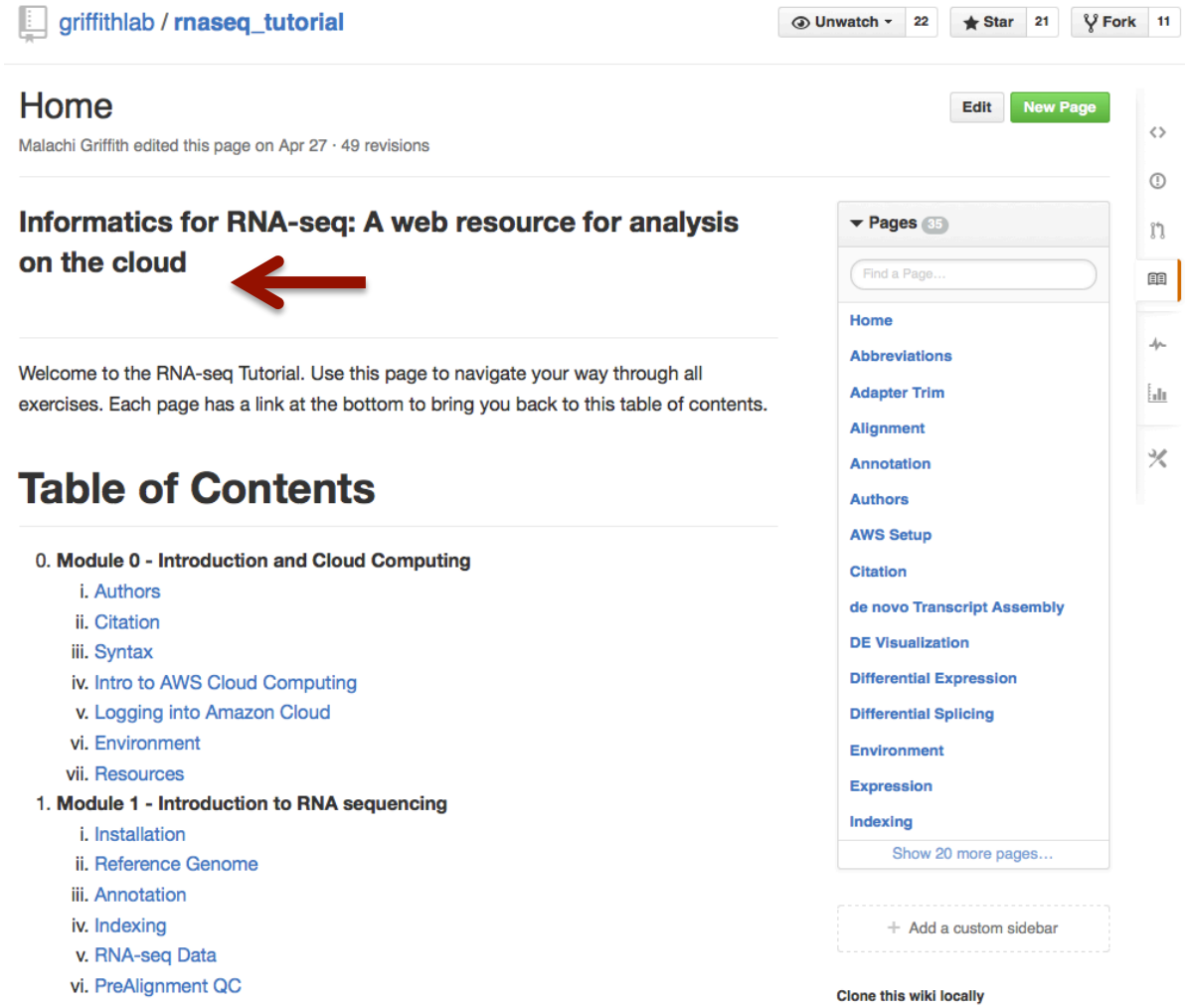


The main CBW Wiki



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

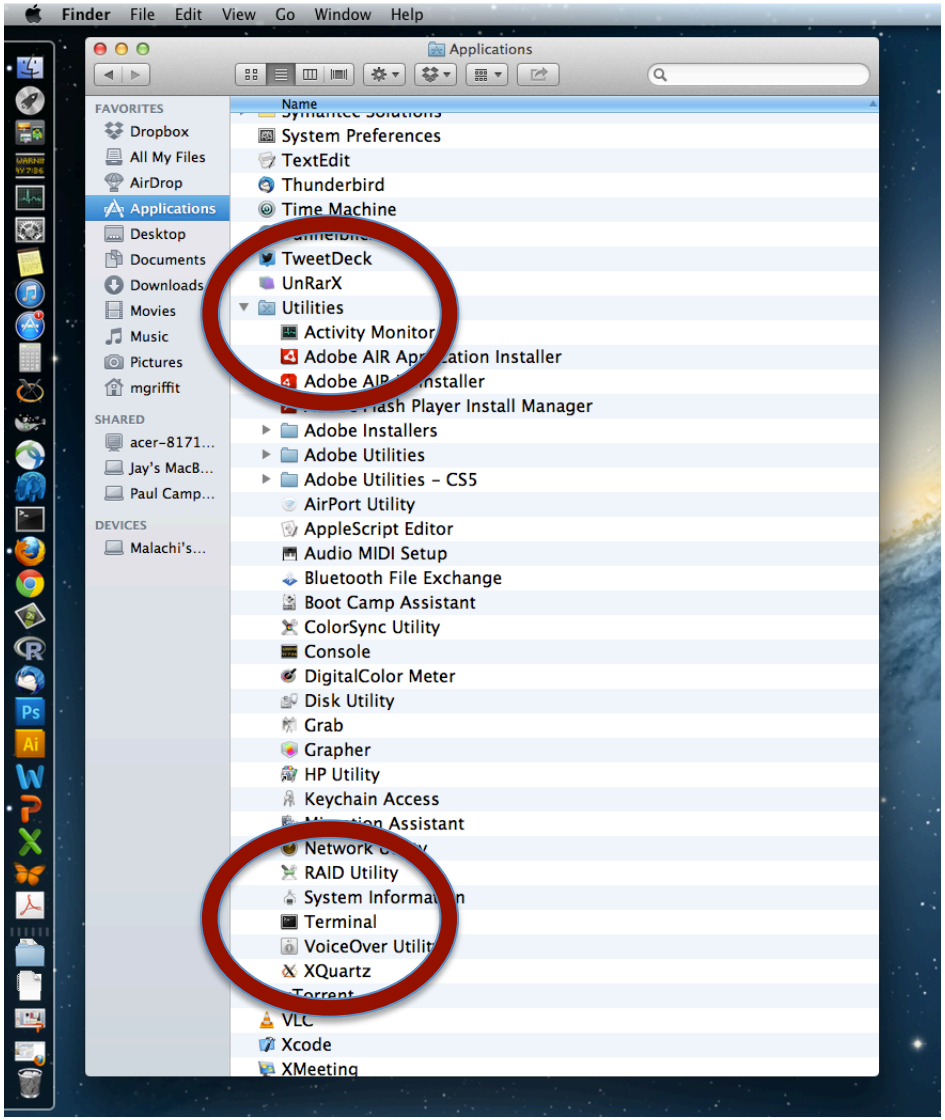
The RNA-seq wiki



http://www.rnaseq.wiki/

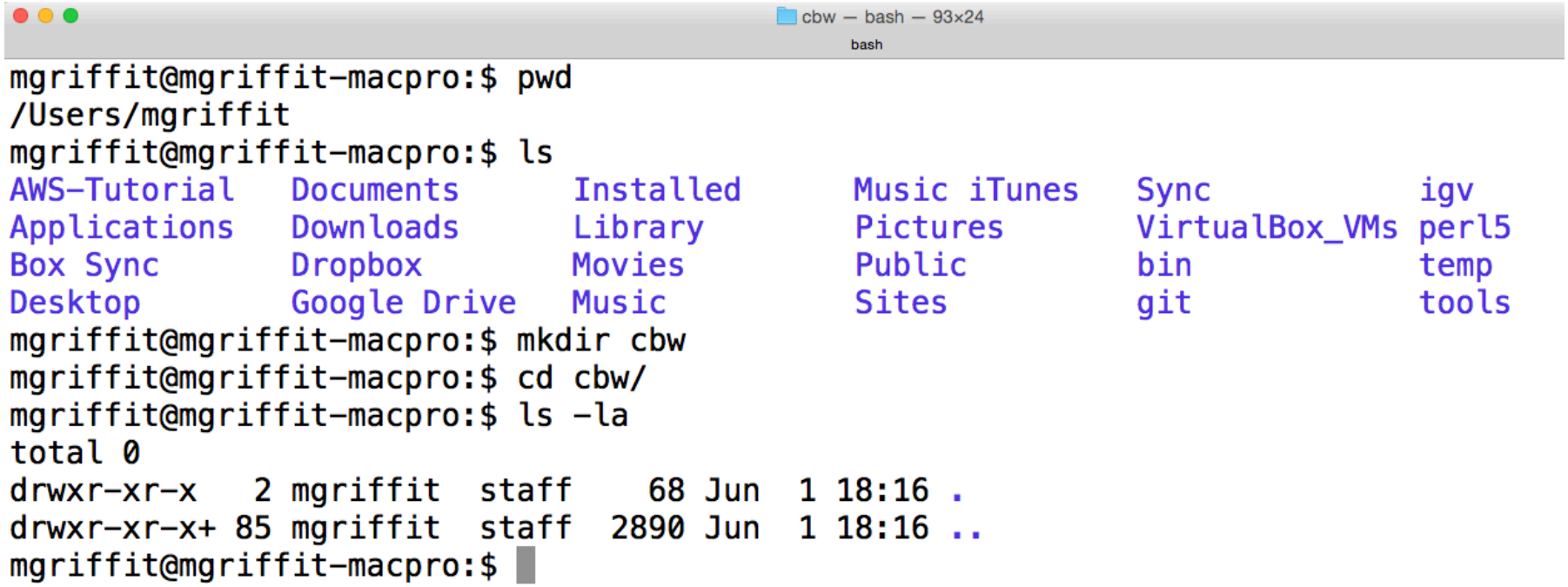
Macintosh users

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



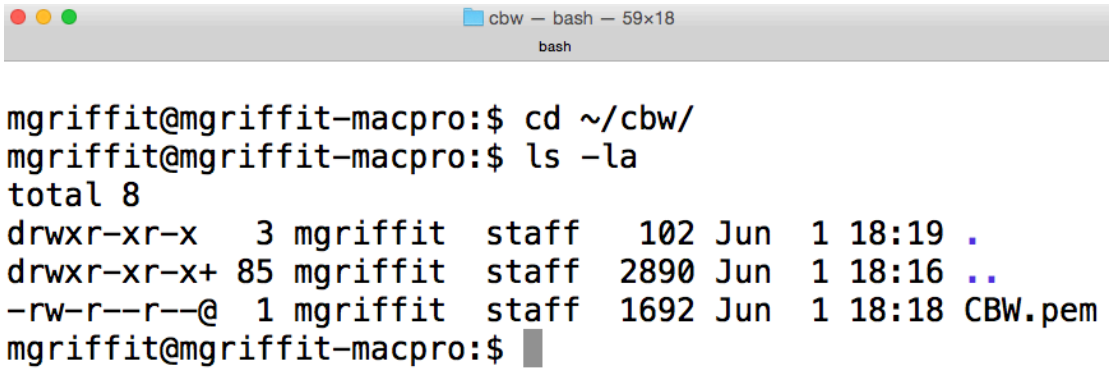
Obtaining your AWS 'key' file from the wiki

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.

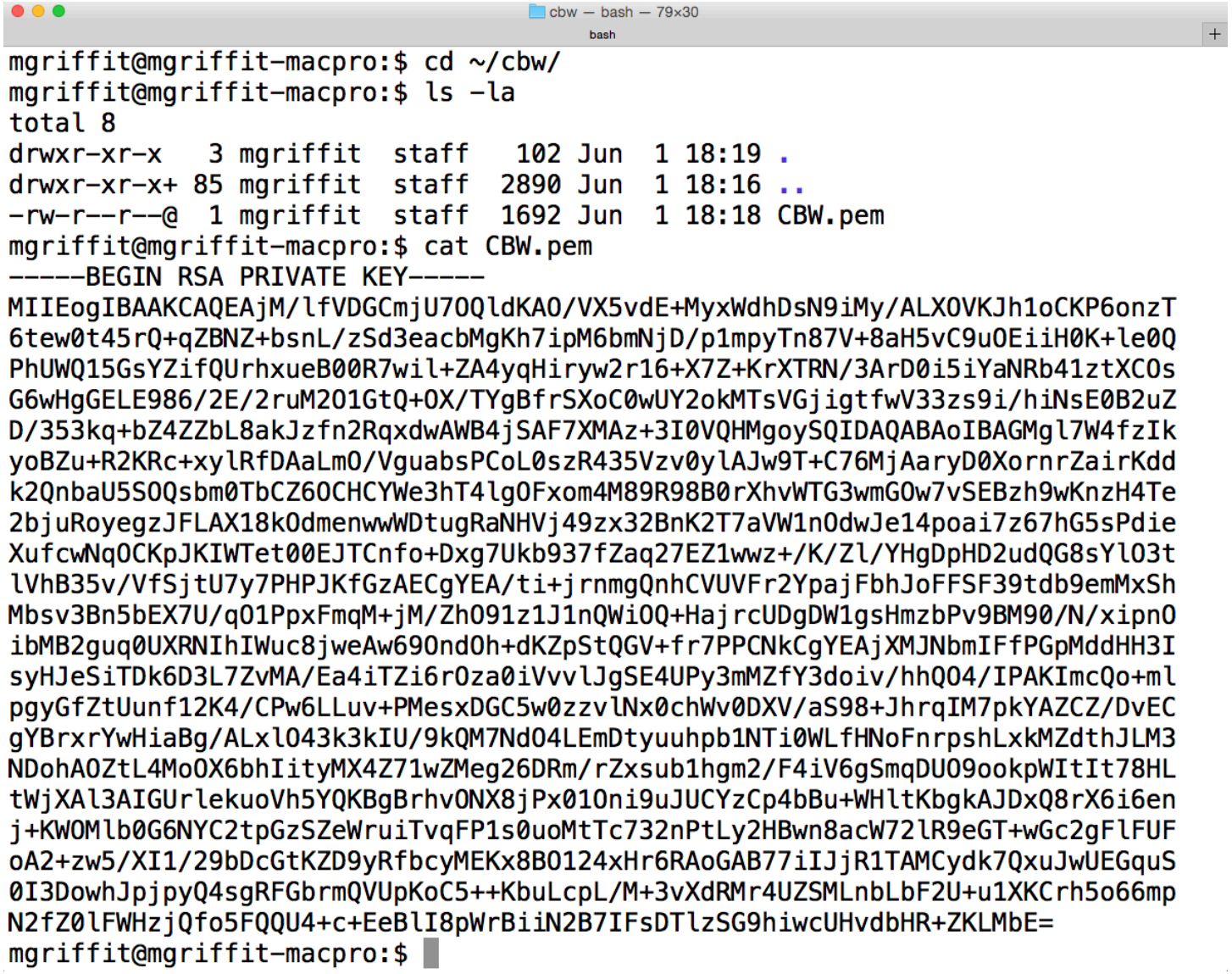
On Mac: Control+

On Windows



Save key file to your new 'cbw' directory

Viewing the ‘key’ file once downloaded



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:

Use your assigned student #

```
ssh -i CBW.pem ubuntu@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.)

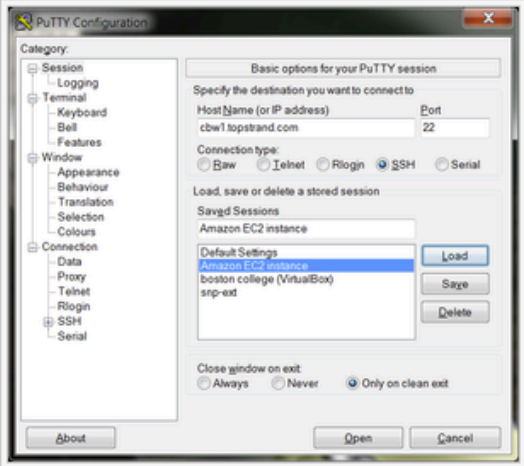
Logging into AWS (Windows)

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.)



Putty Basic Options

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