Canadian Bioinformatics Workshops

www.bioinformatics.ca

In collaboration with

Cold Spring Harbor Laboratory

&

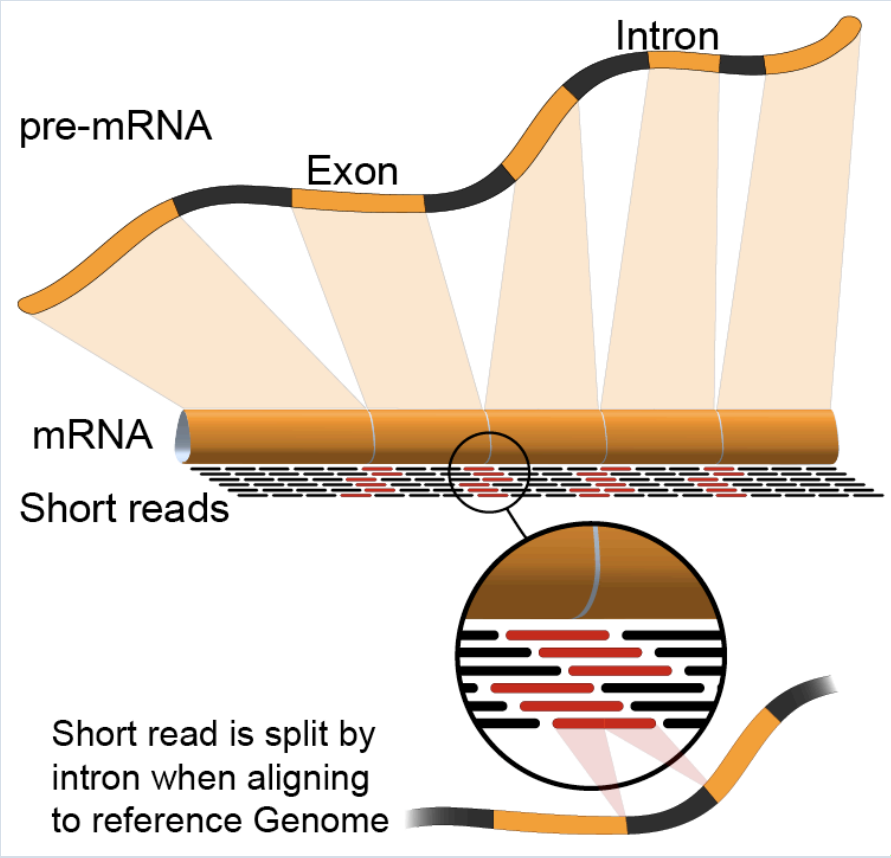
New York Genome Center

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

Malachi Griffith & Obi Griffith

High-throughput Biology: From Sequence to Networks

April 27-May 3, 2015



bioinformatics.ca

CSH Cold Spring Harbor Laboratory 1890 125 2015

NEW YORK GENOME CENTER SM

Learning objectives of the course

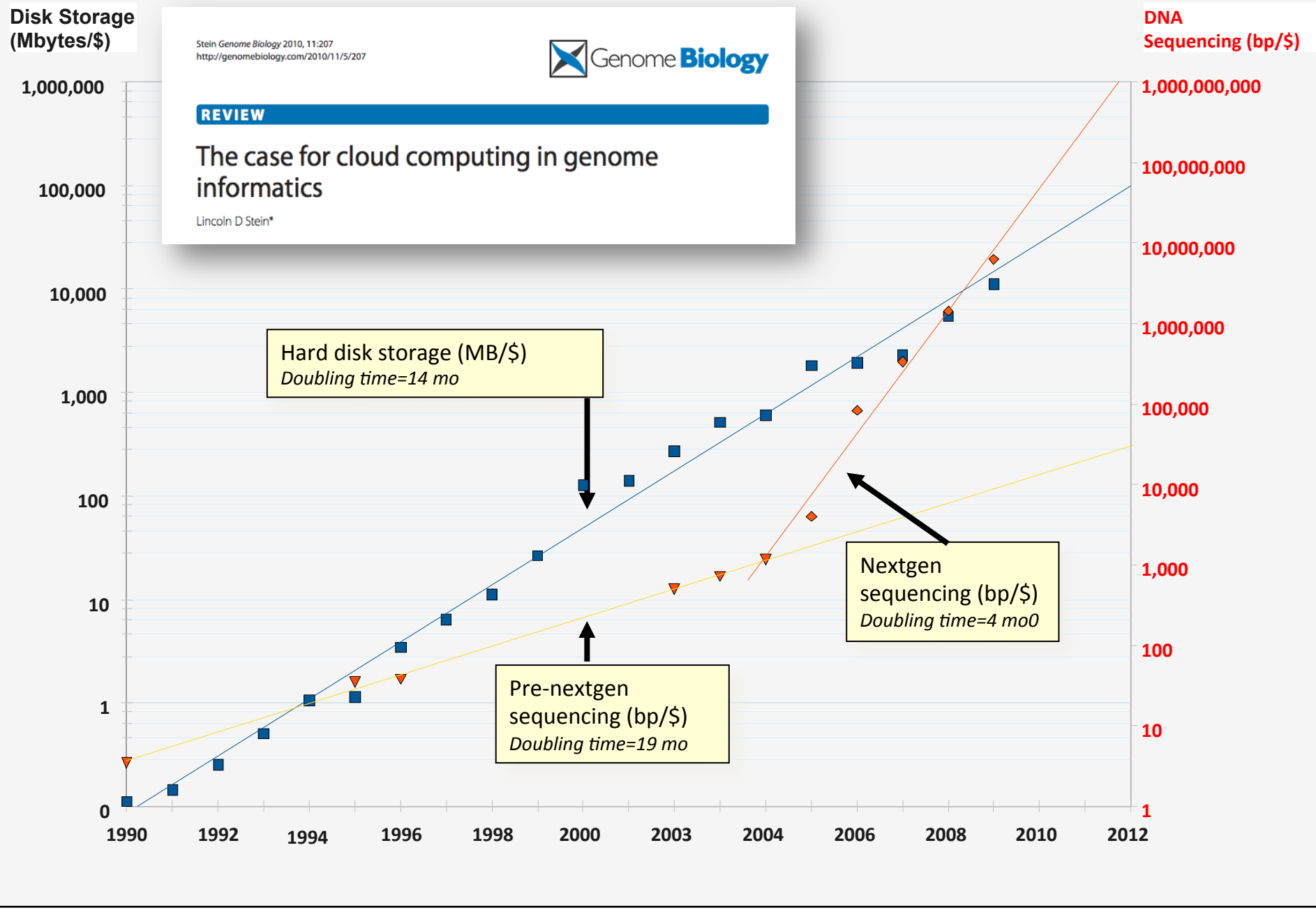
- **Module 0: Introduction to cloud computing**
- Module 1: Introduction to RNA sequencing
- Module 2: RNA-seq alignment and visualization
- Module 3: Expression and Differential Expression
- Module 4: Isoform discovery and alternative expression

- 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 1

- Introduction to cloud computing
- Use of the wiki 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 2014-?, 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

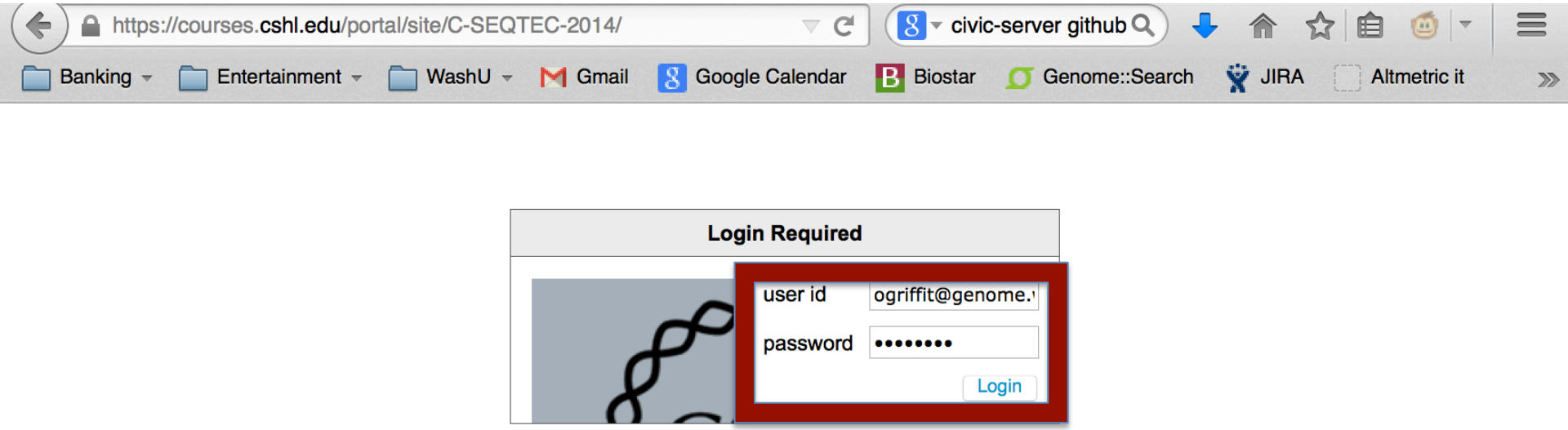
http://aws.amazon.com/console/

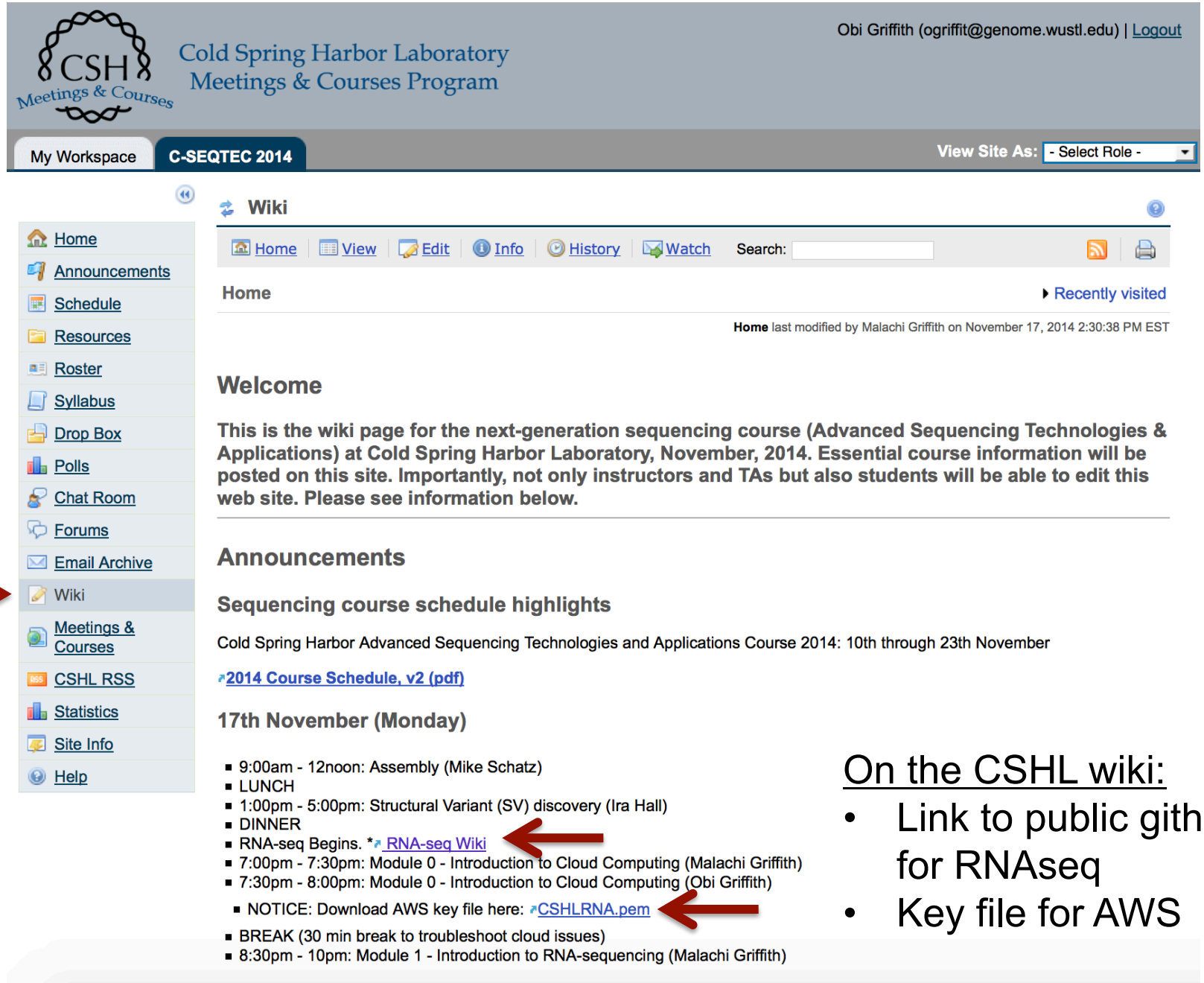
For this workshop: all on Wiki!

https://courses.cshl.edu/portal/site/C-SEQTEC-2014/

Login: email address

Password: password



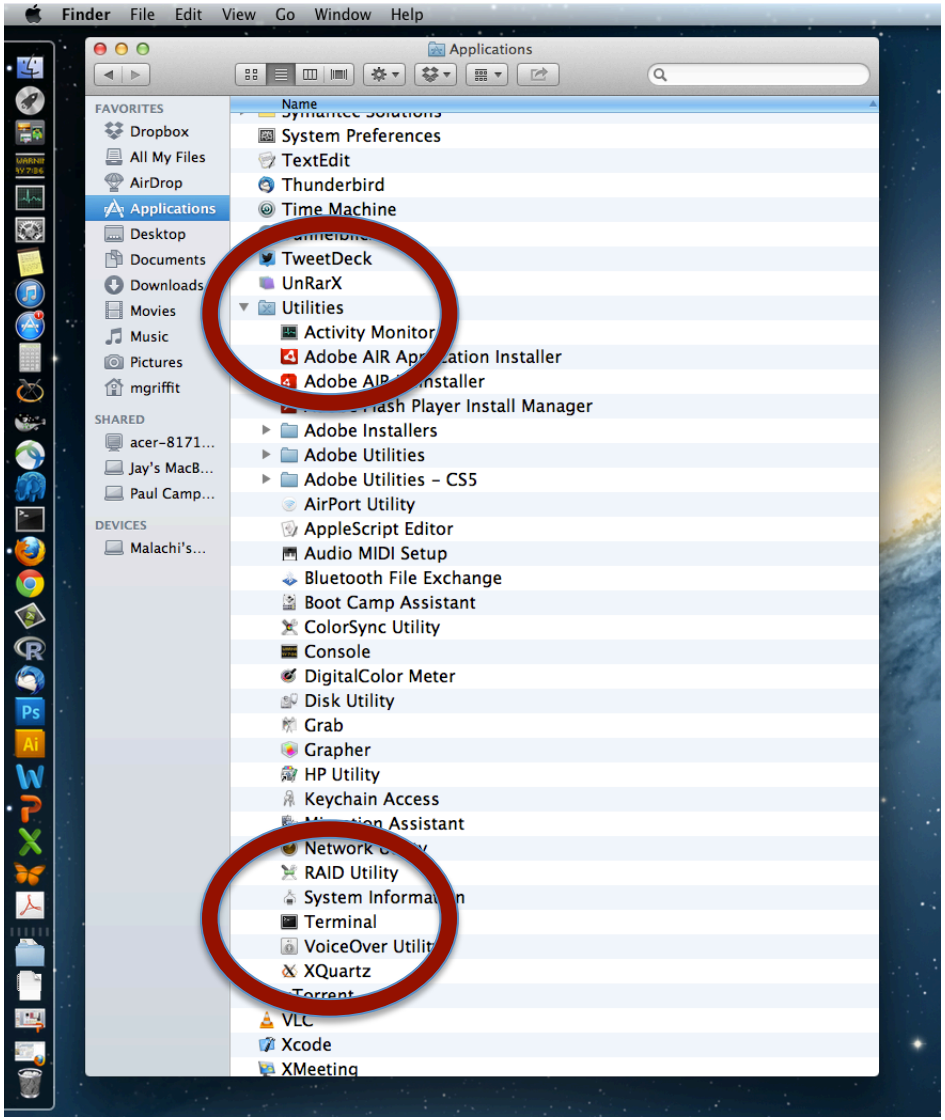


On the CSHL wiki:

- Link to public github wiki for RNAseq
- Key file for AWS

Macintosh users

Opening a ‘terminal session’

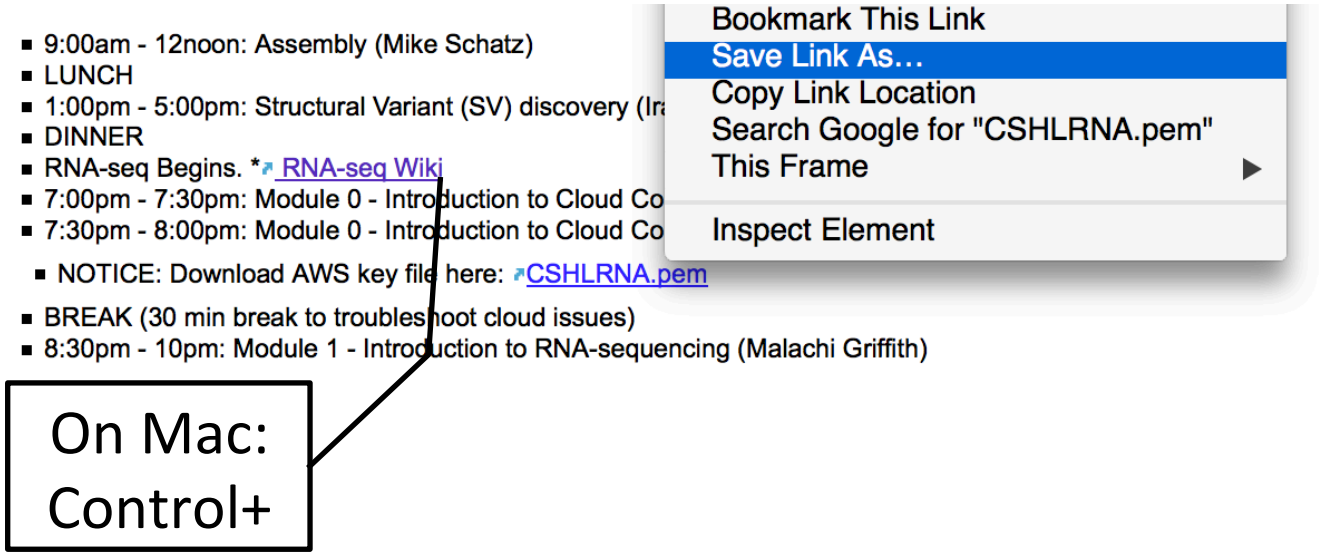


```
[mgriffit@Malachis-MacBook-Pro-2:~]$ pwd
/Users/mgriffit
[mgriffit@Malachis-MacBook-Pro-2:~]$ ls -l
total 16
drwxr-xr-x   2 mgriffit  staff    68 Jan  8 23:27 Applications
drwx------+  8 mgriffit  staff   272 Jun  1 15:58 Desktop
drwx------+  7 mgriffit  staff   238 Apr 17 22:00 Documents
drwx------+ 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

```
vpn-10-1-24-10:~ ogriffit$ pwd
/Users/ogriffit
vpn-10-1-24-10:~ ogriffit$ ls -l
total 0
drwxr-xr-x    2 ogriffit  staff     68 May 16  2013 Applications
drwxr-xr-x  331 ogriffit  staff  11254 Nov 17 14:59 Attachments
drwx------    5 ogriffit  staff    170 Sep 29 16:32 Box Sync
drwx------+  18 ogriffit  staff    612 Nov 12 12:34 Desktop
drwx------+   7 ogriffit  staff    238 Jul 24 17:41 Documents
drwx------+ 157 ogriffit  staff   5338 Nov 17 15:54 Downloads
drwx------@  37 ogriffit  staff   1258 Nov 15 13:15 Dropbox
drwx------@  63 ogriffit  staff   2142 Oct 18 18:27 Library
drwx------+   4 ogriffit  staff    136 May 31  2013 Movies
drwx------+   5 ogriffit  staff    170 Aug 24  2013 Music
drwx------+   5 ogriffit  staff    170 Jun 30 13:39 Pictures
drwxr-xr-x+   4 ogriffit  staff    136 May 16  2013 Public
drwxr-xr-x    7 ogriffit  staff    238 Oct 22 14:32 Sync
drwxr-xr-x    3 ogriffit  staff    102 May 20 00:49 VirtualBox VMs
drwxr-xr-x   10 ogriffit  staff    340 Nov  2 17:44 bin
drwxr-xr-x   13 ogriffit  staff    442 Nov 11 01:39 git
drwxr-xr-x    6 ogriffit  staff    204 May 30  2013 igv
drwxr-xr-x    3 ogriffit  staff    102 Nov  2 17:45 lib
vpn-10-1-24-10:~ ogriffit$ mkdir cshl
vpn-10-1-24-10:~ ogriffit$ cd cshl/
vpn-10-1-24-10:cshl ogriffit$ ls -la
total 0
drwxr-xr-x   2 ogriffit  staff    68 Nov 17 16:20 .
drwxr-xr-x+ 50 ogriffit  staff  1700 Nov 17 16:20 ..
vpn-10-1-24-10:cshl ogriffit$
```

Obtaining your AWS 'key' file from the wiki




```
vpn-10-1-24-10:~ ogriffit$ mkdir cshl
vpn-10-1-24-10:~ ogriffit$ cd cshl/
vpn-10-1-24-10:cshl ogriffit$ ls -la
total 0
drwxr-xr-x   2 ogriffit  staff    68 Nov 17 16:20 .
drwxr-xr-x+ 50 ogriffit  staff  1700 Nov 17 16:20 ..
vpn-10-1-24-10:cshl ogriffit$ ls -la
total 8
drwxr-xr-x   3 ogriffit  staff   102 Nov 17 16:22 .
drwxr-xr-x+ 50 ogriffit  staff  1700 Nov 17 16:20 ..
-rw-r--r--@  1 ogriffit  staff  1692 Nov 17 16:22 CSHLRNA.pem
vpn-10-1-24-10:cshl ogriffit$
```

Viewing the 'key' file once downloaded

```
vpn-10-1-24-10:cshl ogriffit$ ls -la
total 8
drwxr-xr-x   3 ogriffit  staff   102 Nov 17 16:22 .
drwxr-xr-x+ 50 ogriffit  staff  1700 Nov 17 16:20 ..
-rw-r--r--@  1 ogriffit  staff  1692 Nov 17 16:22 CSHLRNA.pem
vpn-10-1-24-10:cshl ogriffit$ cat CSHLRNA.pem
-----BEGIN RSA PRIVATE KEY-----
MIIEowIBAAKCAQEAljC42AmfGYOH8lDgDHUK3gbdFFh8UQuphRF96U0FpOu9Jzre43U5S106vVhU
MvRNN4lmhhACkQb0aTTpZ458eDyXYUrf+1veIMxJzqN76e9/i4VN4BNJMaK9vr4jFUhM4ZH0pFKQ
9gn8T/hFicX+vNPTLVwHWqUJRNi8QLBzxW2Ba16gzAik6H9s05ahnr616jNCzj2DQ1wXTO7kpoS3
X7xmcn/GvpzLteVkb8OOKXnPsNbg8K+W4Ca4V2ZzDdLj7YaZJ98/IYRv5HjyBOnMIt9/Z34IuDFG
PLKgNLGT1+FLaczfbkMoUo47xHa435YKuC9L6UOPRMjyD2nNax0FJwIDAQABAoIBAD9P/Kv9qlcM
DoQEud4mfNVJ2WqZPpa/rs+MpyJxOJ7kTZG7DHigUu3FOFzXXm84c1ClGFYNa7eNUf0DtBzWgPwC
tuHpuW/xszrqQ3bbjgH41zCOmlyKZMGK1CgTaSCwoNA6QgH/WGKPpRBlHZNNpuwc85ncgLEifzzJ
jNpMSBGCSJ0cF9Th7Sf0orFgm7Zlowg/rNVyHg76rLWXJ3TSjcOI7VNHNcaNIX88LATr0uotXwWQ
UjbnyLQWQquSsFpTeoF9RjRNcrkQlJce+9stijs0Sf1qSxsCXuDUR7Kcaxs9s7S7Uo6osggHf+AR
e4P0qhNc5JGhAJJMxLZC3TW66+ECgYEA3OF2lRE/zZP+RbzhQ7tBjILUC0dXAlQPAxwAxVqTXDNA
FE/e63+q+5D7Mxp0BedU+6OGjJr5Sm4PcHfOnOqZMLThlf7P78H4UYUPBxjhmgkLYAsQdBcMkaqK
FUrlB009HZFIPIvlKC0JYrX3/Nup1BbsfhGiyN0zRDgDTrLHY3ECgYEArhHwOn8x+edb1NApTksc
VY8LLhodUYJLeotQnQDivN4WGwk0t96Os0a7CL75yJEEFFLaqwTYWxssZEHzl4/pTqeJUXbrH/sK
0eaVzZVZ9S1nSb0SqgNmFuJuS2ASrc+MGxF38FKnYhzGoVVG6KphfMQq2fByeC9c54O6yUCtdhcC
gYEAwJtqiDpuZJm0lKNL24AifKBovx0KZmFeZdu5YF7HORMHwa6UOM8vDgcyxTFAExXSKUxDO8C3
uYuXsv8GEBcSVkKko+N69xsxqgRZQlsT/vn8DVgYlOK0dyE7H64bYO6daHcTgpiKzynWkVkUUk5K
HUOQHa7LQFSDK+2cFLgY3kECgYA4aB21hwQ13A/4/V91PIe4/fp8fP/lKti0pDKCPgg7dagqKKiJ
sLgDRjdTlTcyFDQDNfogExjdJQgUkP6TrZHH6ChqWqPEoEAYVqVSkkQtXAItdfR3XVSWs9dT25UR
98CKAPYhzVbqRzLiktTAjn26xYKxTuZ5JbfwWdKMZsYPDwKBgHLZZD3abGgGVghW+yZ3/7ZDqHor
sUwSKvZ4B3VT6uKO2AmRDwjeHesbVSaAFyEzWWYMgaGfZGToAqJbTejov5E/GC11HlUVJyqx2he7
M5WCh3y6PKUknebocJWhn1AA80hvd+YBapX3Lq+AkfCvzKV2+IXNjRd3prARS2tUqDbm
-----END RSA PRIVATE KEY-----vpn-10-1-24-10:cshl ogriffit$
vpn-10-1-24-10:cshl ogriffit$ _
```

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 CBWRNA.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 in to AWS

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

```
chmod 600 CSHLRNA.pem
```

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

`##` is your assigned student number. Your student number is the number on the participant list. If your number is less than 10, please add 0 in front of it. `-i` selects a file from which the public key authentication is read. `ubuntu` is the name of a user on the system you are logging into (a default user of the Ubuntu operating system). `cshl##.dyndns.org` is the address of the linux system on Amazon that you are logging into.

Copying files from AWS 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 CSHLRNA.pem ubuntu@cshl##.dyndns.org: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://cshl##.dyndns.org/

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

We are on a Coffee Break & Networking Session