

## MSBI 32400 – LAB 1

LARRY HELSETH PHD &  
JASON EDELSTEIN

January 9, 2019

## Lab Motivation

2

- Bioinformatics uses Linux
  - ▣ Mac OS uses BSD Linux under the hood but shields users from command line
  - ▣ Windows users can use Cygwin
- Develop comfort with command line
  - ▣ Working on remote computers
  - ▣ Transferring files
  - ▣ Installing software
  - ▣ Controlling complex software with command line flags

MSBI 32400 - Lab 1 1/09/2019

## Linux

3

- CaSe SenSiTiVe!
- Permissions
- Home directory
- “root” user is in charge
- Sometimes, “sudo” granted
  - Use with caution

MSBI 32400 - Lab 1 1/09/2019

## Linux comes in different flavors

4

- Ubuntu, Debian, CentOS (= Red Hat)
- Bio-Linux
  - ▣ Field D, et al., Nature Biotech., 24, 801 - 803 (2006),  
*Open software for biologists: from famine to feast*
  - ▣ <http://environmentalomics.org/bio-linux-download/>

MSBI 32400 - Lab 1 1/09/2019

## Bioinformatics Software

5

- Always changing so document which version you used
- Usually have to install from source (since so many Linux flavors) using “make” and “make install”
- Need root privileges to install for all users
  - ▢ Follow instructions to build local copy
  - ▢ Use virtual environment (Python)
  - ▢ Get to know your Linux admin team!
- Red Hat Software Collection Library offers newer builds (Python, Java, databases, etc.) without upgrading whole server

MSBI 32400 - Lab 1 1/09/2019

## Sources of Bioinformatics Software

6

- GitHub.com
  - ▢ See “Git for scientists”, chapter 5, Bioinformatics Data Skills (on-line at [http://proquestcombo.safaribooksonline.com.proxy.uchicago.edu/book/bioinformatics/9781449367480/5dot-git-for-scientists/ch05\\_html](http://proquestcombo.safaribooksonline.com.proxy.uchicago.edu/book/bioinformatics/9781449367480/5dot-git-for-scientists/ch05_html))
- Sourceforge.net
- AWS (prebuilt images)
- Check publications, Biostars.org, SeqAnswers.com, Bioinformatics.org, blogs, etc., for software recommendations

MSBI 32400 - Lab 1 1/09/2019

## Bioinformatics Blogs

7

- Start with Titus Brown!
  - ▢ <http://ivory.idyll.org/blog/>
- <https://nsaunders.wordpress.com>
- <https://bcbio.wordpress.com>
- <https://ewanbirney.wordpress.com>
- <http://crazyhottommy.blogspot.com>
- <http://www.ensembl.info>
- <http://www.gettinggeneticsdone.com>
- <https://blastedbio.blogspot.com>
- <https://macarthurlab.org/blog/>
- <http://www.cureffi.org>

MSBI 32400 - Lab 1 1/09/2019

## Connecting to Linux

8

Use SSH (secure shell) protocol for command line  
Windows:

- Putty (install from putty.org)

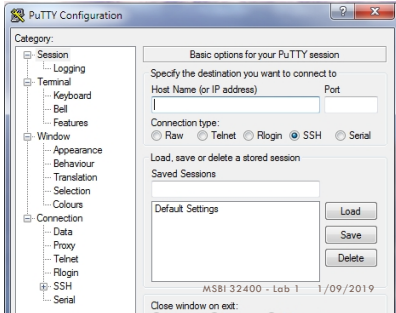
Mac OS X

- Terminal (installed in Applications/Utilities)

MSBI 32400 - Lab 1 1/09/2019

## Windows: Using Putty

- Install from [www.putty.org](http://www.putty.org)



## Mac OS: Using Terminal

- Open from Applications/Utilities

MSBI 32400 - Lab 1 1/09/2019

## Moving Files

- Browser upload limited to 4 GB
- Secure file transfer protocol (SFTP)
  - ▣ Advanced use: scp or rsync
- Recommend installing Filezilla an open source client (<https://filezilla-project.org>)
- Command line options:
  - ▣ wget, curl, ftp



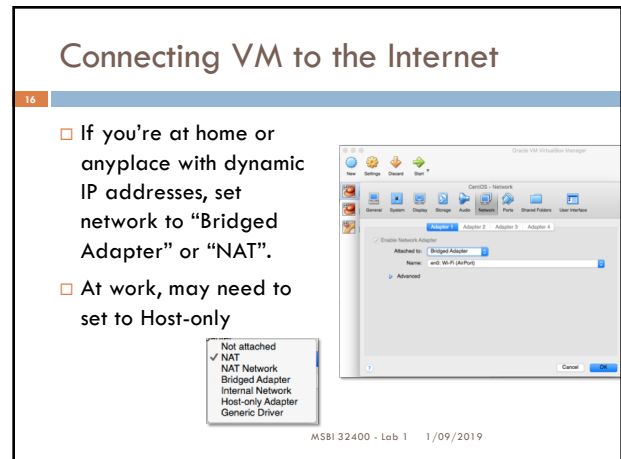
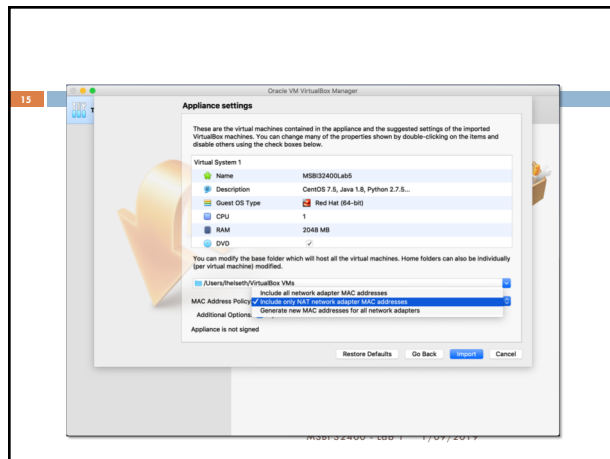
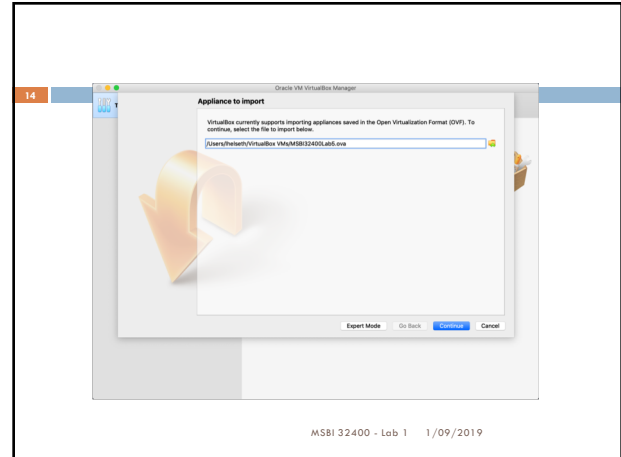
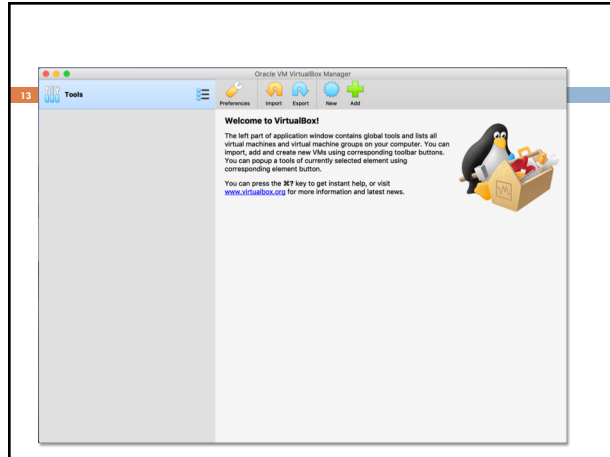
MSBI 32400 - Lab 1 1/09/2019

## VirtualBox



- Start the VirtualBox application you installed on your computer
- Select the Import arrow
- Browse to find where you saved the lab image you downloaded (MSBI32400Lab5.ova) & import
  - Recommend specifying the Virtual Machine folder as exempt from Time Machine or other backup
- Check Description in settings for username & PW
- Press Start

MSBI 32400 - Lab 1 1/09/2019



## Working with Virtual Machine

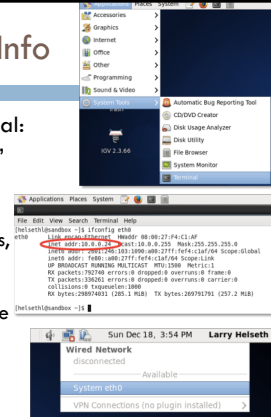
- To return ownership of keyboard and mouse to your host operating system, VirtualBox reserves a special key on your keyboard for itself: the “host key”. By default, this is the right Control key on your Windows keyboard; on a Mac host, the default host key is the left Command key.
- We recommend Pausing machines before quitting VirtualBox, and **saving** the current state when prompted.



MSBI 32400 - Lab 1 1/09/2019

## Getting Connection Info

- Sign in, then launch a Terminal:
- Type “ifconfig eth0<enter>”
  - Write down the IP address for later
  - If you don't see an IP address, check the network icon at top right. If red X by icon, click & select **System eth0** to activate
- Type in terminal: ping -c3 www.google.com



MSBI 32400 - Lab 1 1/09/2019

## Accessing your VM from Host OS

- On your VirtualBox, choose Devices/Network then uncheck Connect Network Adapter
- Click on Network Settings and choose Host-Only
  - If not found, setup per: <http://askubuntu.com/questions/293816/>
- Reselect Connect Network Adapter
- Open a terminal on the VM and type ifconfig

MSBI 32400 - Lab 1 1/09/2019

## Connecting to your Linux system

- Press your “Host key” and toggle back to your laptop (“Host Operating System”)
- Open a terminal (either Putty or Terminal) and type: ssh student@<your IP address>
- Enter the same password as you used on the VM. You should be in the “student” home directory. Type “pwd<enter>” to see /home/student. Type “ls<enter>” to see the files in your home directory.
- Type “exit<enter>” to sign off

MSBI 32400 - Lab 1 1/09/2019

## Linux Cheat Sheets

21



\$19.00 + S&H  
OR  
Just learn them!

<https://store.xkcd.com/products/linux-cheat-shirt>

MSBI 32400 - Lab 1 1/09/2019

## Commands you'll use a lot

22

<input type="checkbox"/> ls (ls -la, ls -ltr, ls -lh)	<input type="checkbox"/> head
<input type="checkbox"/> pwd	<input type="checkbox"/> tail
<input type="checkbox"/> cd	<input type="checkbox"/> man
<input type="checkbox"/> cp	<input type="checkbox"/> sftp
<input type="checkbox"/> rm	<input type="checkbox"/> echo
<input type="checkbox"/> cat	<input type="checkbox"/> Redirect commands like '>', '>>', '<'
<input type="checkbox"/> more	<input type="checkbox"/> Pipe command ' '
<input type="checkbox"/> find	<input type="checkbox"/> Ctrl-C, Ctrl-Z
<input type="checkbox"/> clear	

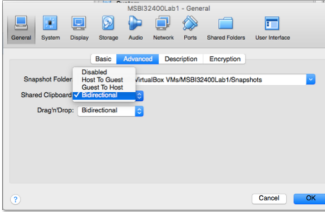
Learn syntax: <command> -h OR man <command> OR Google!

MSBI 32400 - Lab 1 1/09/2019

## Enable bi-directional copy & paste

23

- ☐ In Linux terminal:
  - ☐ Copy is Ctrl-Shift-C
  - ☐ Paste is Ctrl-Shift-V
- ☐ Change VirtualBox General/Advanced tab
  - ☐ May have to **reinstall Guest Extensions**

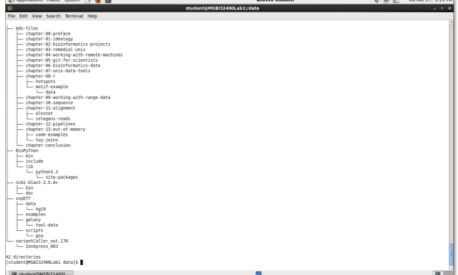


MSBI 32400 - Lab 1 1/09/2019

## Preloaded with tools & sample data

24

- ☐ Installed Buffalo files in /data:



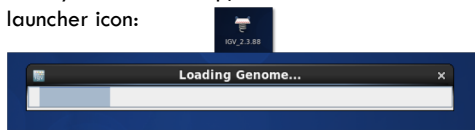
MSBI 32400 - Lab 1 1/09/2019

## Let's look at some sequence data

25

Change your VM back to NAT (Slide 13-Connecting VM to the Internet) to access the Internet.

- We'll be using the Integrative Genome Viewer (IGV) from the Broad Institute<sup>§</sup>
- From your VM Desktop, double click on the IGV launcher icon:



<sup>§</sup> <http://software.broadinstitute.org/software/igv/>

MSBI 32400 - Lab 1 1/09/2019

## Using IGV

26

- Default setting is coordinates on top, Variant file, BAM file (genome reads) and RefSeq at the bottom.
- Navigate the genome by choosing chromosomes, typing coordinates, dragging, or pinching & zooming.

<http://software.broadinstitute.org/software/igv/Navigate>

MSBI 32400 - Lab 1 1/09/2019

## Viewing a saved session

27

- Preloaded using the Buffalo sample BAM file & VCF from chapter 6 files by using File, Open Session, choose the igv\_session.xml in student directory and click OK.
- Buffalo prepared these from public 1000 Genome data on volunteer NA12891 (CEU). The BAM file only has data on one gene, USH2A.

MSBI 32400 - Lab 1 1/09/2019

## Viewing reads & SNPs

28

- Click and drag to zoom around "215,900bp"
- Point to exons at bottom and double click on exon 59 to zoom until you see sequence
- Compare codons:  
[https://en.wikipedia.org/wiki/DNA\\_codon\\_table](https://en.wikipedia.org/wiki/DNA_codon_table)

MSBI 32400 - Lab 1 1/09/2019

## OPTIONAL

29

- Terminal only VM from UMass course:  
[http://bioinfo.umassmed.edu/bootstrappers/bootstrappers-courses/pastCourses/unix\\_2015-09/index.html](http://bioinfo.umassmed.edu/bootstrappers/bootstrappers-courses/pastCourses/unix_2015-09/index.html)
  - ▣ See CentOS-6.5.ova in Files/Lab1 on Canvas
  - ▣ Use: student/student123
    - No bioinformatics software installed but can use to practice working on the command line
    - Use Machine/ACPI Shutdown when done

MSBI 32400 - Lab 1 1/09/2019

## Homework

30

- Upload to Canvas OR E-mail Jason Edelstein ([jason@uchicago.edu](mailto:jason@uchicago.edu)) with "Lab #1" in the subject line and the following information before next class:
  - ▣ A screen shot of you signing in
  - ▣ A list of files you found in your home directory
  - ▣ A list of folders you found in the /data directory
  - ▣ A screen shot from IGV showing you found the SNP
    - Either File/Save Image from IGV or print screen

MSBI 32400 - Lab 1 1/09/2019