

Hello everyone,

Thank you very much for registering the Bioinformatics Training Workshop at PAG 2020, which is only two weeks away.

The room for the workshop is **Royal Palm Salon 5/6 at Town and Country San Diego.**  
The Google map location link is:  
<https://www.google.com/maps/place/Town+and+Country+San+Diego/@32.7618584,-117.1685843,16.54z/data=!4m8!3m7!1s0x0:0xcba5ae24f7a445ad!5m2!4m1!1i2!8m2!3d32.7620096!4d-117.1684998>:

**Town and Country San Diego** **Address:** 500 Hotel Cir N, San Diego, CA 92108

**Town and Country San Diego** **Phone:** (619) 291-7131

Thanks to the great efforts of the instructors, we have got all materials ready and uploaded to my GitHub repository: <https://github.com/haibol2016/BioinformaticsTrainingWorkshop>.

All the instructors are expecting all attendants to familiarize themselves with the materials due to the tight time. Please download the file in the repository by click the green “clone or download” button on the top right, then the “Downlod ZIP” button. After downloading, please unzip the compressed file in “.zip” format.

For **UNIX session**, please read the two PPTX files in the 1.UNIX/**presentations folder**.

For **R session**, please install **R and RStudio** by following instruction on this webpage: <https://www.datacamp.com/community/tutorials/installing-R-windows-mac-ubuntu>. Then install all the required R packages by running the following code snippet. We will not have enough time to install all these packages on site.

```
# Install all required packages for this R session, and later RNA-seq and ChIP-seq data analysis practice.
```

```
packages_list <- c("aplpack", "bootstrap", "broom", "ca", "car", "cluster", "colorspace", "ComplexHeatmap", "corrplot", "d3heatmap", "DAAG", "datasets", "dendextend", "devtools", "doBy", "downloader", "dplyr", "factoextra", "FactoMineR", "fpc", "gclus", "ggforce", "ggm", "ggthemes", "gplots", "grDevices", "grid", "gvlma", "hexbin", "Hmisc", "lattice", "leaps", "MASS", "mclust", "methods", "multcomp", "NbClust", "nFactors", "openxlsx", "pastecs", "pheatmap", "polycor", "psych", "pvclust", "Rcmdr", "RColorBrewer", "readxl", "relaimpo", "rgl", "rmarkdown", "scatterplot3d", "sm", "tibble", "tidyverse", "vcd", "vioplot", "XLConnect", "DESeq2", "edgeR", "pca3d", "sva", "limma", "genefilter", "WriteXLS", "DiffBind", "ChIPQC", "ChIPseeker", "GenomicFeatures", "magick")
```

```
if (!require("BiocManager")) {install.packages("BiocManager", dependencies = TRUE)}
```

```
for (pkg in packages_list)
```

```
{
```

```
  if (!pkg %in% installed.packages())
```

```
  {
```

```
    BiocManager::install(pkg, dependencies = TRUE)
```

```
  }
```

```
}
```

# After finishing running the above code, to make sure all packages has been installed, please run the following code snippet. You should see nothing printed out.

```
for (pkg in packages_list)
```

```
{
```

```
  if (!pkg %in% installed.packages())
```

```
  {
```

```
    print(pkg)
```

```
  }
```

```
}
```

### **Further Resources for learning basics of UNIX (you can pick one of them you like most)**

- UNIX Tutorial for Beginners: <http://www.ee.surrey.ac.uk/Teaching/Unix/>
- Learning UNIX through exercises: <https://www.doc.ic.ac.uk/~wjk/UnixIntro/>
- Learn Shell Programming - Free Interactive Shell Programming Tutorial: <https://www.learnshell.org/>
- Free Linux shell access is available here: <https://bellard.org/jslinux/> or <https://linuxzoo.net/>

### **Further Resources for learning R basics (you can pick one you like most)**

- Learn R, in R: <https://swirlstats.com/>
- Quick R: <https://www.statmethods.net/r-tutorial/index.html>
- Introduction to R (DataCamp free course, providing interactive programming): <https://www.datacamp.com/courses/free-introduction-to-r>
- Free RStudio cloud for practice: <https://rstudio.cloud/>

For **Introduction to NGS and common tools, RNA-seq, and ATAC-seq data analysis**, please read through the PPTX files. For ChIP-seq, please read the **ChIP-seq data analysis.md** file on the GitHub folder: **BioinformaticsTrainingWorkshop/4.ChIP-seq/**.

I will upload the last file for RNA-seq hands-on practice as soon as possible.  
Thank you very much again and look forward to seeing at the Workshop.

Best,

Haibo Liu

### Software installation:

0. **Install Integrative Genome Viewer (IGV)**. By following direction :  
<https://software.broadinstitute.org/software/igv/download>

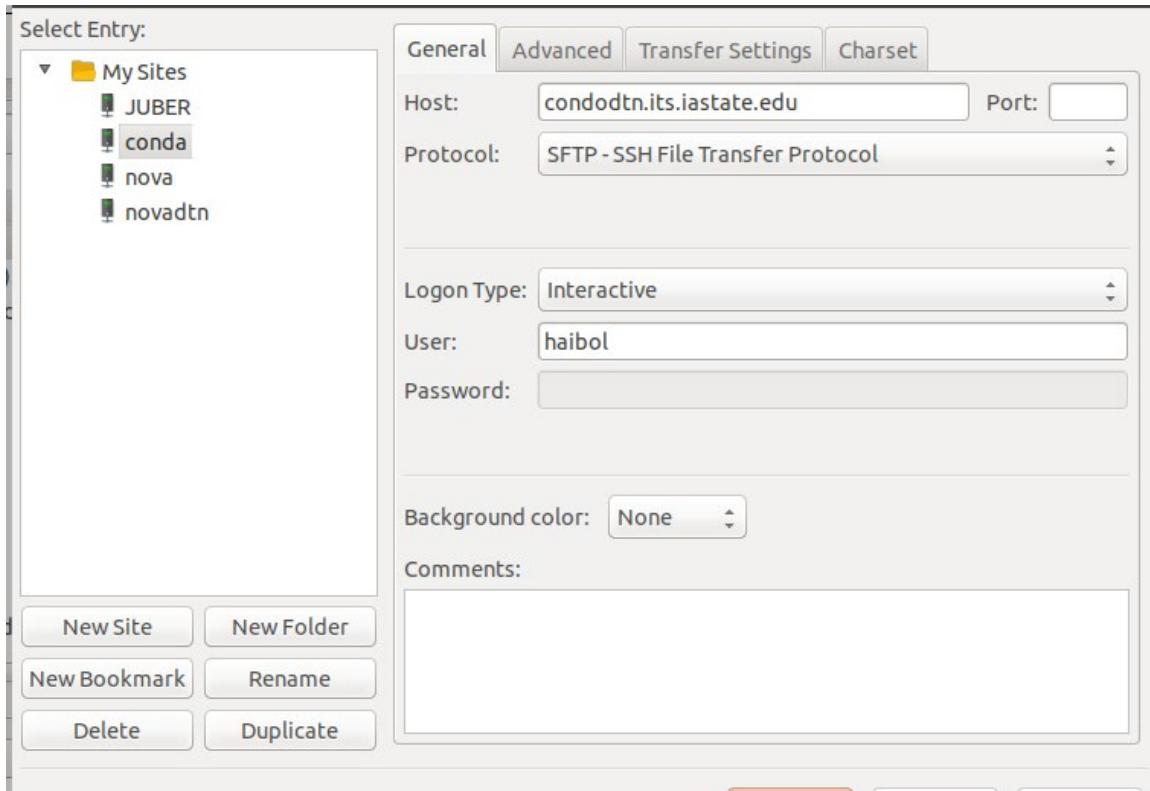
The pig reference genome Sus Scrofa11.1 files for IGV will be available from the condo HPC cluster for downloading.

1. **FileZilla** Downloading and installation **for Mac users**:

Download FileZilla Client from <https://filezilla-project.org/> by clicking the button and install the software by following the instruction: [https://wiki.filezilla-project.org/Client\\_Installation](https://wiki.filezilla-project.org/Client_Installation)



Set up the FileZilla



## 2. Secure Shell Client (SSH) download and installation for Windows users

Please download and install and install SSH by following the direction here:

<https://www.ohlone.edu/download-and-install-ssh-secure-shell-secure-file-transfer-client-windows-web-center>

## 3. Install ISU VPN for Login from off-campus

Access to condo2017.its is limited to machines in the iastate.edu domain. To access this machine from an off-campus computer, please install the ISU VPN using the following link: [Installation of ISU VPN](#)

## 4. Condo HPC Access and Login

Please follow instruction on this page to access the HPC of Iowa state University:

<https://www.hpc.iastate.edu/guides/condo-2017/access-and-login>. Google Authentication is needed to access the HPC. Depending on the type of SmartPhone you have, please refer to :

Android Phone: <https://www.hpc.iastate.edu/guides/condo-2017/access-and-login/google-auth-on-android>

Blackberry Phone: <https://www.hpc.iastate.edu/guides/condo-2017/access-and-login/google-auth-on-blackberry>

Iphone-ipad-ipod: <https://www.hpc.iastate.edu/guides/condo-2017/access-and-login/google-auth-on-iphone-ipad-ipod>

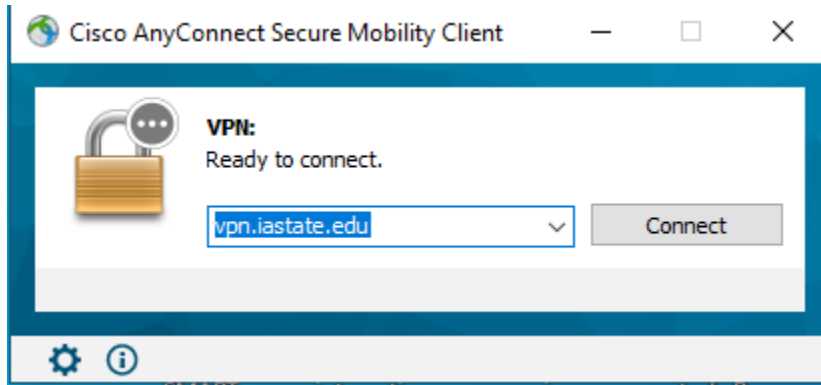
Windows Phone: <https://www.hpc.iastate.edu/guides/condo-2017/access-and-login/google-auth-on-windows-phone>

## 5. About the Condo Cluster

We will use the Condo Free Tier for our hand-on practices on RNA-seq and ChIP-seq data analysis. The computing resource is listed here: <https://www.hpc.iastate.edu/guides/condo-2017>.

## 6. Connect to the condo HPC.

- A. Start the Cisco AnyConnect Secure Mobility Client, and click “Connect”.



- B. perform Cisco AnyConnect Login by input your ISU NetID and password associated with it. Click “Sign In”.

**IOWA STATE UNIVERSITY**

**Sign In**

**Username**

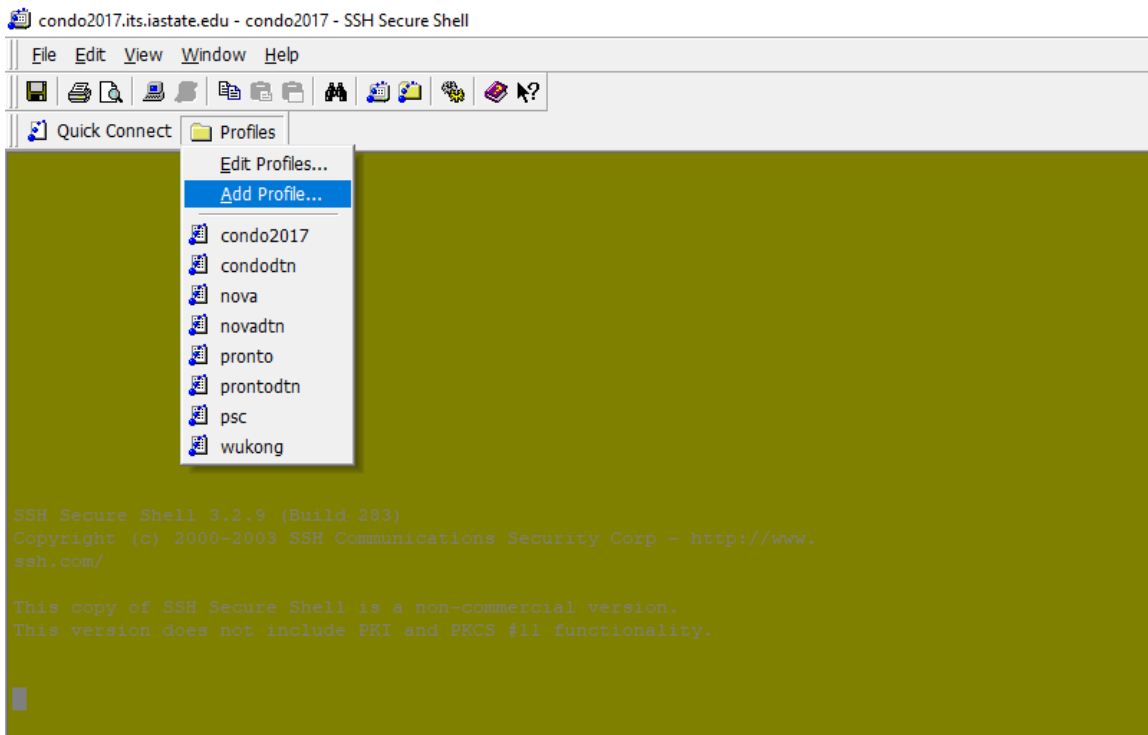
**Password**

**Sign In**

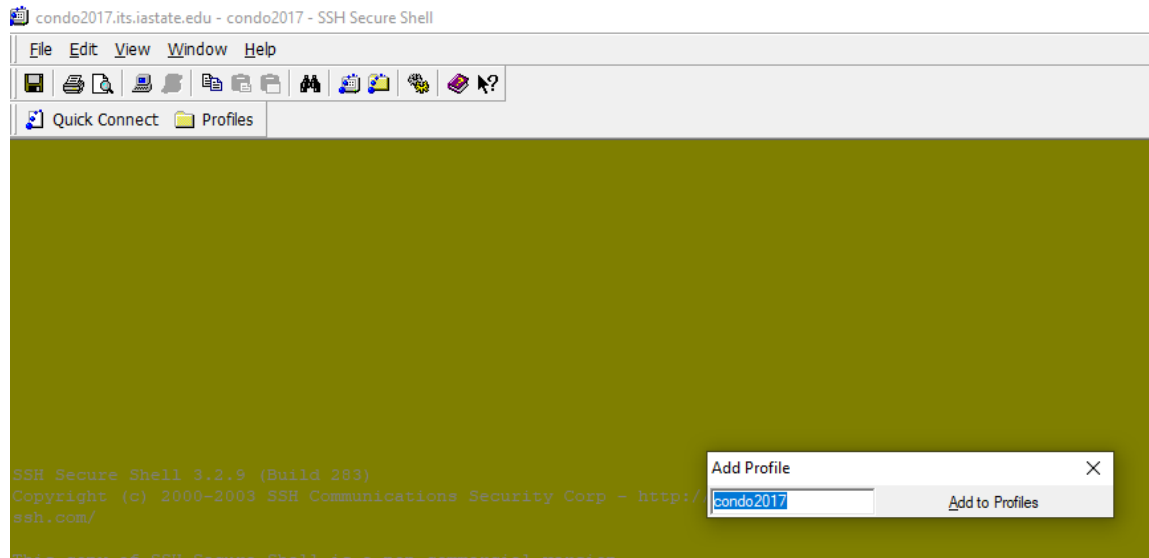
[Need help signing in?](#)

**C. For Windows users, please setup SSH connection to `condo2017.its.iastate.edu` by following the instruction below:**

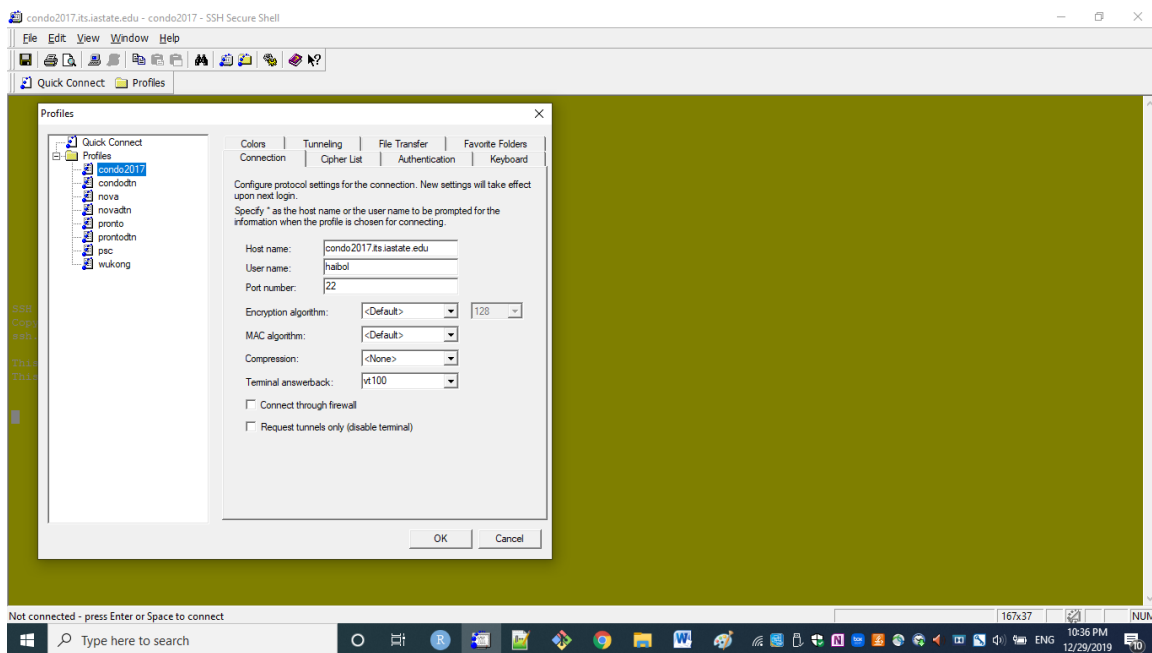
1. Click "Profile", then "Add profile"



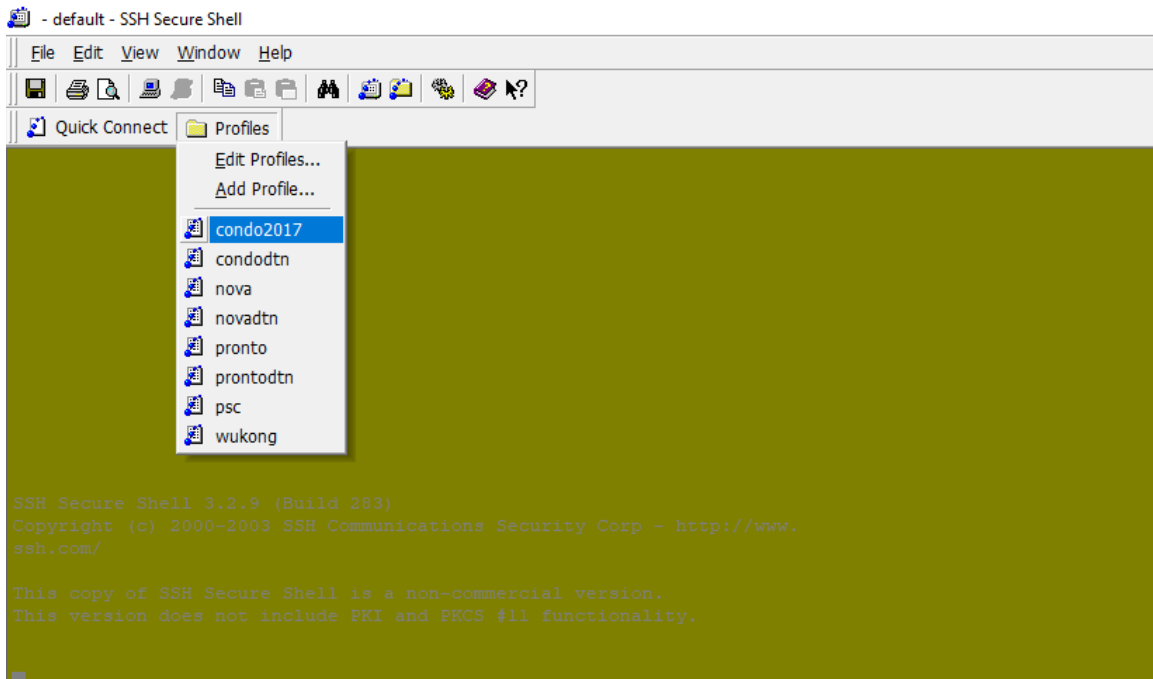
2. Fill in the prompt box with "condo2017" and click "Add to Profiles"



3. Click "Profile", then "Edit Profiles". Fill in the Host name: condo2017.its.iastate.edu and User name: your ISU NetID; Port number: 22  
Click "OK".

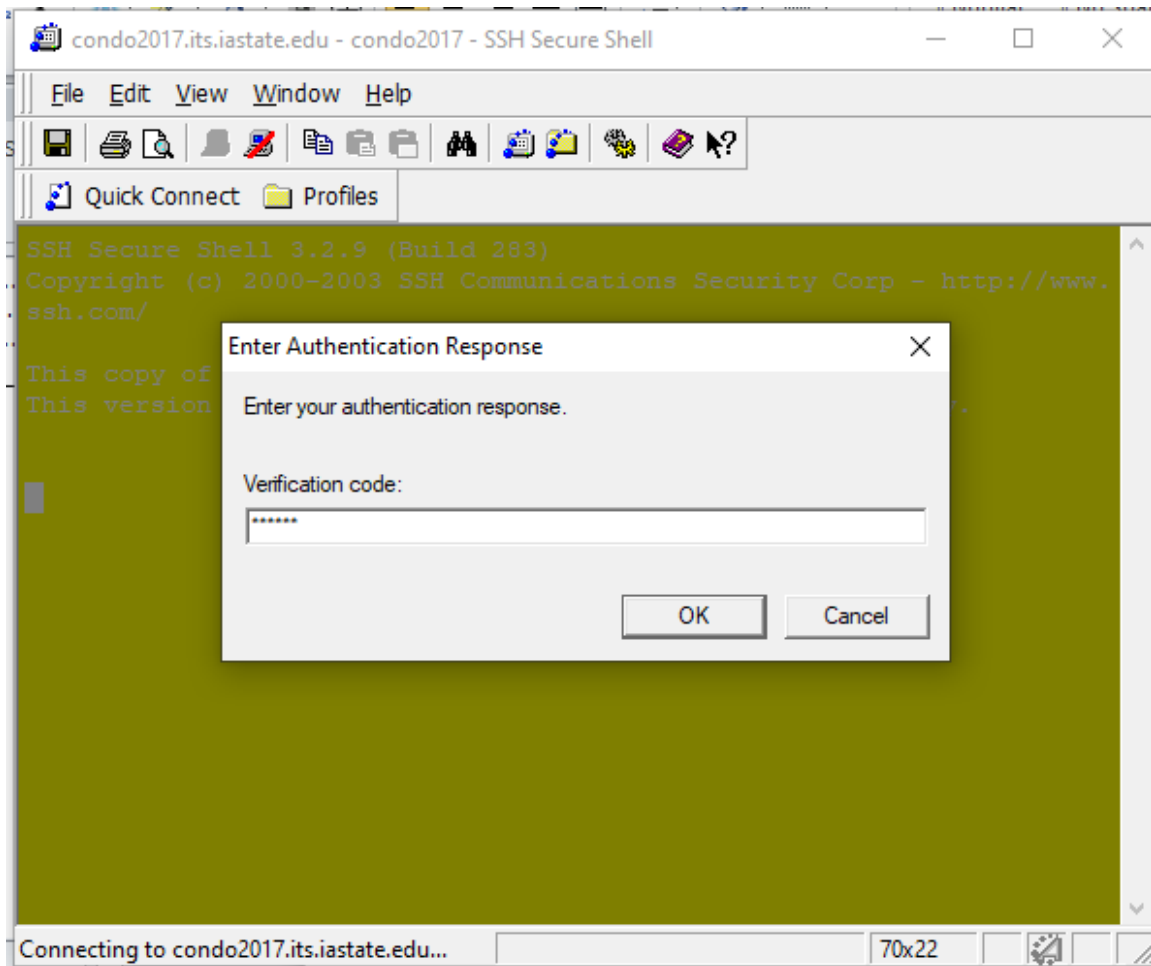


4. Click "profile", then "condo2017"



5. Fill in the verification code shown on your Google Authentication App, a number of 6 digits.

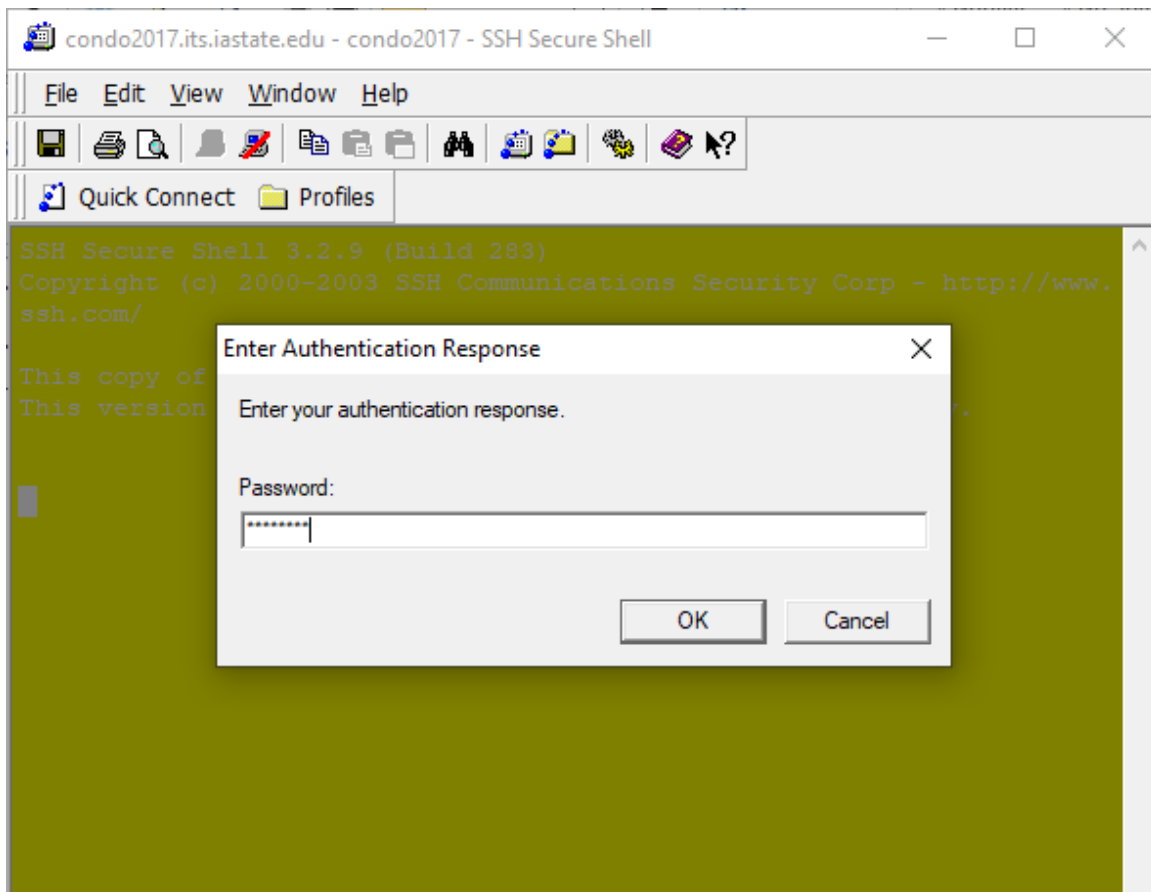




6. Click "OK" and Fill in your password which is associated with your ISU NetID.

If you have problems in installing those R packages in your own computers, **You can use the R and R packages I have installed on the condo HPC** during the workshop, though you have to eventually get R and its packages installed in your own computer in the future if you need R for your work.

To use that, please log in your account on <ISU NetID>@condo2017.its.iastate.edu. And use "vi" to open the shell configuration file ".bash\_profile" as follows:



7. Now you should be connected to the Condo HPC cluster.

**D. For Mac users, use the built-in terminal and type in the following command by replacing your actual ISU NetID with <ISU NetID>.**

```
ssh -X <ISU NetID>@condo2017.its.iastate.edu
```

Login with your ISU NetID and your ISU email password.

Storage Space for attendants:

Large temporary space `/freetmp` is mounted on free nodes and on `condodtn` (it's not mounted on `condo` login node). Users should **use `/freetmp/cktuggle-free`**. They can **ssh to `condodtn` from `condo` login node**.

## use R and R packages in the condo cluster

If you have problems in installing those R packages in your own computers, you can use the R and R packages I have installed on the condo HPC during the workshop, though you have to eventually get R and its packages installed in your own computer in the future if you need R for your work.

To use that, please log in your account on @condo2017.its.iastate.edu. And use “vi” to open the shell configuration file “.bash\_profile” as follows:

Hide

```
vi .bash_profile
```

After open the file, hit the “I” key on your keyboard to get into “INSERT” mode and move the cursor to the end of the file and hit the “Enter” key. Then copy the following lines ( CTRL + C for Windows, and Command + C for Mac) and paste (SHIFT + insert for Windows, and Command + V for Mac) to the line where the cursor locates.

Hide

```
# User specific environment and startup programs

export PATH=$PATH:/bin:/usr/bin:/usr/local/bin:$HOME/bin/~/./local/bin

### alias to avoid mess up files
alias rm='rm -i'
alias cp='cp -i'
alias mv='mv -i'
alias ll='ls -l'
set -o noclobber

# R library path
```

```
R_LIBS=${R_LIBS}:/home/haibol/R/x86_64-pc-linux-gnu-library/3.6
```

```
### set colorful prompt
```

```
export PS1="\[\e[31;40m\]\[\[\e[m]\]\[\e[32;40m\]\u\[\e[m]\]\[\e[31;40m\]@  
\[\e[m]\]\[\e[32;40m\]\h\[\e[m]\]\[\e[31;40m\]:\[\e[m]\]\[\e[32;40m\]\w\[\e  
[m]\]\[\e[31;40m\]\]\[\e[m]\]\[\e[31;40m\]\$\[\e[m]\] "
```

Then hit the “ESC”, the “:” key and type “wq”. The modified file will have been saved.

To make the modification effective, run the following command:

Hide

```
source .bash_profile
```

No you can load the R and run R interactively, by running the following command.

Hide

```
# log in to a node for interactive job running
```

```
salloc -N 1 -t 05:00:00
```

```
# load R3.6.0 module
```

```
module load r/3.6.0-py2-fupx2uq
```

```
# start R
```

```
R
```

## Use the storage space in /freetmp/cktuggle-free

Large temporary space /freetmp is mounted on free nodes and on condodtn (it’s not mounted on condo login node). Users should use /freetmp/cktuggle-free. To access that

space and use it as working space. Please follow the instruction belows after logging in the head node (login node):

Hide

```
# you should be at your home directory
pwd

# replace <your ISU NetID> with your real NetID
mkdir -p /freetmp/cktuggle-free/<your ISU NetID>_work

# Create a symbolic link to your workspace
ln -s /freetmp/cktuggle-free/<your ISU NetID>_work ./

# now you can start real work in this directory
cd <your ISU NetID>_work
```