

```
#####
# 1) Connect to the aci cluster using the command line
#####
ssh -X username@aci-b.aci.ics.psu.edu
```

```
#####
# 2) Navigating files and directories
#####
# Print current working directory
pwd
# List the names of the files and directories in the current directory
ls
# List file names and display '/' for directory
ls -F
# To find out other options in 'ls'
ls --help
# List file names using a long listing format
ls -l
# Change directory to "scratch"
cd scratch
# List files
ls -l
# Go up one directory
cd ..
# Change directory to "scratch"
cd scratch
```

```
#####
# 3) Download datasets
#####
git clone https://github.com/PSUGenomix/BG_retreat_workshop_2017.git
```

```
#####
# 4) Working with files and directories
#####
# Make a directory called "test"
mkdir test
ls -F
# Change directory to "test"
cd test
cd ../
cd BG_retreat_workshop_2017
ls
# Copy file
cp sg11_reference_gene_annotation.gff copy.gff
ls
# Rename copy.gff to copy2.gff
mv copy.gff copy2.gff
ls
# Remove copy2.gff
rm copy2.gff
ls
```

```

# Create a directory called "test2"
mkdir test2
ls
# Delete "test2", which will produce an error
rm test2
# Recursively delete "test2"
rm -r test2
ls

#####
# 5) Pipes and Filters
#####
# Word count : Number of lines, words, characters
wc sg11_reference_gene_annotation.gff
# Print number of lines per file
wc -l sg11_reference_gene_annotation.gff
# Show contents of file
cat sg11_reference_gene_annotation.gff
# Look at first lines
head sg11_reference_gene_annotation.gff
# Look at last lines
tail sg11_reference_gene_annotation.gff
# Short contents : -n option is for numerical order
sort sg11_reference_gene_annotation.gff
# Sort contents and save it to sorted-lengths.txt
sort -n sg11_reference_gene_annotation.gff > sorted.gff
ls
# Look at first lines in "sorted-lengths.txt"
head sorted.gff
# Sort and look at the first 2 lines
sort -n sg11_reference_gene_annotation.gff | head
# Look at first column
cut -f1 sg11_reference_gene_annotation.gff
cut -f1 sg11_reference_gene_annotation.gff | head
# Look at first and third columns
cut -f1,3 sg11_reference_gene_annotation.gff | head
# Look at first column and remove duplicates
cut -f1 sg11_reference_gene_annotation.gff | uniq
cut -f1 sg11_reference_gene_annotation.gff | uniq | head
# Search for gene YAL047C
grep YAL047C sg11_reference_gene_annotation.gff
# Open vi text editor
vi draft2.txt

```