# Activity 1: In silico digestion

In this activity, you will computationally “digest” genomic DNA to predict numbers and lengths of fragments in a GBS experiment.

1. Log on to crick using the Terminal (Mac) or an SSH client such as PuTTY (Windows):   
   crick.cgrb.oregonstate.edu  
   \*Note: Don’t forget to designate port 732!
2. Use the “print working directory” command to show the full path to your home directory  
   pwd

1. Make a subdirectory for this workshop in your home directory:  
   mkdir GBS\_ws
2. Class data, scripts and other files can be found here:  
   /nfs1/Teaching/data/viningk/GBS  
   Use the “change directory” cd command to go to this directory, and then list the contents:  
   cd /nfs1/Teaching/data/viningk/GBS  
   ls  
   Notice that there are two subdirectories:   
   scripts/  
   PhytozomeV10/
3. Go into the scripts subdirectory and list contents. Copy the “GBScovCalc.pl” script to the GBS\_ws subdirectory you created in your home directory:  
   cp GBScovCalc.pl /your/home/directory/GBS\_ws
4. Now, find your way into the PhytozomeV10 subdirectory and list its contents. These fasta files were pulled from the Phytozome plant genome database (<http://phytozome.jgi.doe.gov/pz/portal.html>).
5. Use the pwd command to view the full path to the PhytozomeV10 subdirectory and copy this path into your notes. (Hint: Click/drag to select, right-click to copy). Then, copy the rice genome fasta file, Osativa\_204\_v7.0.fa, and append it to the path in your notes, so that it looks like this:  
   /nfs1/Teaching/data/viningk/GBS/PhytozomeV10/Osativa\_204\_v7.0.fa
6. Go back to your GBS\_ws directory and run the GBScovCalc.pl script on the rice genome file. You can do this by making a symbolic link to the file, or by pasting in the full path to that file from your notes as shown below. Redirect the output to a file called ricedigest.txt (all on one line):  
   GBScovCalc.pl /nfs1/Teaching/data/viningk/GBS/PhytozomeV10/Osativa\_204\_v7.0.fa > ricedigest.txt
7. Look at the ricedigest.txt file using the less command. The -S flag keeps lines from wrapping.  
   less -S ricedigest.txt
8. How many ApeKI cut sites are there in this genome assembly? What is the expected coverage per ApeKI site for each individual in this GBS experiment?
9. Try running the script on another of the plant genomes in the PhytozomeV10 subdirectory to compare their numbers to that of the rice genome. What is different about the poplar (Ptrichocarpa) and soybean (Gmax) genome assemblies relative to the rice and potato genome assemblies that might make the results less accurate for Ptrichocarpa and Gmax? (Hint: Use grep “>” to look at all of the fasta headers in a file).  
   grep ">" Ptrichocarpa\_210\_v3.0.fa