

Week 3 Using UNIX and Scripting UNIX

Introduction

This week, you'll get even more experience with using UNIX to perform bioinformatics data processing and analysis tasks. You'll also get introduced to UNIX scripting, which means: writing programs that automate tasks you would normally perform by typing and entering commands one by one.

The assignment for this week will feature several basic tasks you'll need to do again and again as a bioinformatics programmer or analyst, including:

- obtaining and installing command-line utilities
- obtaining bioinformatics data sets from the public domain ([PlantGDB downloads](#) and [IGBQuickLoad](#))
- processing data sets in a pipeline of linked steps
- checking results using a visualization tool ([Integrated Genome Browser](#)) and UNIX filter commands (e.g., cut)

Once you've performed the annotation process manually, you'll write a shell script that automates each step. For this, you'll learn

- how to create and use variables in a script
- how to check whether a file exists
- how to use bash string operators to extract components of a file name
- how to run shell commands from inside a script (and capture their output)
- control structures, including
 - how to iterate (for loops)
 - conditions (if/else statements)

Read and/or Watch

- **Learning the bash shell**
 - [Chapter 4](#) Basic Shell programming
 - [Chapter 5](#) Flow Control

References Dr. Loraine recommends

- **Beginners guide to the BASH shell** http://tldp.org/LDP/Bash-Beginners-Guide/html/sect_01_05.html
- **Test if a file exists with the BASH shell** <http://www.electrictoolbox.com/test-file-exists-bash-shell/>

Assignments

- [Annotate a genome with ESTs](#) due 5 pm, Thursday Sept 13



If you are using the base image, you only have 10 Gb of disk space on your VM. Contact iPlant support (support@iplantcollaborative.org) to request an EBS (elastic block storage) volume. 50 Gb should be enough. See:

- **Mounting an EBS Volume** <https://pods.iplantcollaborative.org/wiki/x/PKxm>
- **Attaching an EBS Volume to an Instance** <https://pods.iplantcollaborative.org/wiki/x/OKxm>