**Introduction to Unix - Instructor Notes**

**Resource Title:** Introduction to Unix

**Authors:** Serghei Mangul, UCLA; Kristen Butela, University of Pittsburgh; Michael Sierk, Saint Vincent College; Lihua Stefan, NCSU

## **Article Context**

### **Course:** Upper-level biology course (e.g. Microbiology, Genomics, Genetics, Bioinformatics, etc.)

### **Course Level:** Intermediate/Advanced

### **Class Type:** Lab

### **Audience:** Life Sciences Major

### **Class Size:** 1-50

### **Lesson Length:** One class period (3 hours). Can be extended to 2 or 3 lab sessions.

### **Key Words:** Unix, Linux, command line, sequence analysis

## **Scientific Teaching Context**

### **Learning Goal(s)**

* (NIBLSE Core Competency 5): Students will be able to find, retrieve, and organize various types of biological data.
* (NIBLSE Core Competency 7): Students will be able to use command-line bioinformatics tools and write simple computer scripts.

### **Learning Objective(s)**

* Students will be able to carry out simple tasks using the Unix command-line environment.

## **Introduction**

This laboratory module leads undergraduate biology students in learning how to operate in a command-line environment. It is derived from a workshop given by the Institute for Quantitative and Computational Biosciences at UCLA, primarily to graduate students and research scientists with the need to learn how to process sequence data in a Unix environment. The materials from that three-day workshop are available here: <https://github.com/QCB-Collaboratory/W1.UNIX.command.line/wiki>. These materials can be used as background for the Instructor and/or students, or as an extension of this exercise.

Biology increasingly requires analysis of “big data”, and many programs written for this purpose are only available in a Unix environment. If a biology student is going to be required to perform these types of analysis in graduate school (for example) operating in the command line environment will be much less of a shock if they have been exposed to it as an undergraduate. Of course undergraduates may also make use of the command line environment to carry out research for their own summer or senior research projects.

**Intended Audience**

The intended audience is junior- and senior-level undergraduate life science students. This activity has been adapted from a workshop used with graduate students and research faculty, but has only been used with a small number of undergraduates so far.

**Required Learning Time**

A single three-hour laboratory session is appropriate. The exercise could also be spread out over two laboratory sessions, and could be expanded to three or more sessions using the workshop materials at the link provided above if so desired. The second application exercise, on microbiome analysis using QIIME2, is more advanced and will take students at least an hour to work through.

**Prerequisite Student Knowledge**

Students will need some background in high-throughput sequencing methods and metagenomics or microbiome analysis in order to carry out the sequence analysis activities.

### **Prerequisite Teacher Knowledge**

The instructor will need some familiarity with the Unix environment, and an understanding of the sequencing methods and sequence analysis methods used in the application activities. The second application, in particular, will require significant background work by the instructor if he or she is not already familiar with this type of analysis. The instructor will need to either be able to install a Unix environment, help the students access a Unix environment on their own computers, or have access to IT support that can install a Unix environment for the students.

## **Lesson Plan**

***Classroom Management***

* Instructors will need to determine the best option for accessing a Unix environment depending on the resources available to them at their institution, the size of the class, etc. Various options are described below in the Supporting Materials section. They will need to provide the appropriate instructions for installation and/or access to their students.
* The instructor should work through the modules prior to classroom implementation to validate the instructions. This will also allow the instructor to be familiar with areas that students may have trouble with.
* Resources needed: internet access, a computer, a Unix environment.
* Depending on how the Unix environment will be installed, this may need to be completed prior to the class, or sufficient class time will have to be devoted to this.
* If performing the QIIME2 microbiome analysis for the application activity, the QIIME2 environment will need to be installed. This requires 5GB of disk space, and may take 30 minutes or more to install.

## **SUPPORTING MATERIALS**

**Running Unix on a Windows Machine**

There are multiple ways to run Unix/Linux on a Windows desktop or laptop computer. Three options, from simplest to most complicated, are:

1. Built in options.
   1. Enable Unix on a machine running Windows 10 or later. Windows 10 comes with the option to run a Linux subsystem (Windows Subsystem for Linux, or WSL). The instructions for doing this can be found here: <https://docs.microsoft.com/en-us/windows/wsl/install-win10>
   2. On Apple laptops and desktops, the Unix environment is already part of the operating system. Students just go to Applications -> Utilities -> Terminal to open a Unix command-line environment.
2. Log in to a Unix server.
   1. (Windows) Use the PuTTY program to connect to a Unix server. PuTTY allows users to log in to a remote Unix server and open a shell (command-line environment). In order to use this option you will need a server running Unix or Linux that is configured to allow students to log into it. PuTTY can be downloaded here: <https://putty.org>.
   2. (Apple) Students can log in to the server using the Terminal described in 1b above.
3. Install a Virtual Machine to run Unix on your local machine.
   1. (Windows) A more robust option than #1 is to install software on your Windows computer that enables you to run a Unix/Linux system on your computer. However this requires installing the software, which may require significant time to download and install, and will require more computer resources such as disk space and memory.
      1. Install virtualbox from <https://www.virtualbox.org>. Virtualbox is free to download and use.
   2. (Apple) Virtualbox can also be installed on Apple machines, if desired. This is the preferred way to install QIIME2 if using the microbiome analysis application activity.
   3. If performing the QIIME2 activity, the instructions for installing the QIIME2 package using Virtualbox are described here: <https://docs.qiime2.org/2018.11/install/virtual/virtualbox/>. It is recommended that this installation be performed first, then the Introduction to Unix slides & exercises be carried out in this environment, then the microbiome analysis activity be carried out. Note that this installation process requires approximately 30-60 minutes to complete over a normal internet connection and requires 5 GB of disk space.
4. Using a Jupyter notebook on Google Collaboratory. This is a cloud-based solution, which may be easier than installing Unix locally (and is platform agnostic). However, it requires internet access and a Google account.
   1. <https://colab.research.google.com/>
   2. Click “New Python3 Notebook”