Computational Requirements and Prerequisites

(a.k.a. Assignment #1 - Bioinformatics Environment & Linux Skills Prep)

Hello, all participants of the *DNA Sequencing and Data Analysis* course 2025! My name is Amit, and I will be your biocomputational tutor for this course.

The course will be interleaved with hands-on computational genomic experimentation, including the final assignment, and is oriented as a computational lab. We will start the computational part during the third lecture, on April 24th! ***Please remember to bring your laptop with you for every computational lecture!***

Bioinformatics software and data management are facilitated mostly using the Linux OS. To complete the home and final assignments you would need to make sure you have access to a UNIX based operating system, i.e., Linux, MacOS, FreeBSD, with a strong preference to a Linux based distribution. Windows OS users are encouraged to enable WSL functionality and review the following guides:

What is the Windows Subsystem for Linux?
How to install Linux on Windows with WSL
Set up a WSL development environment

As all of you are holding an undergraduate degree in computer science or equivalent, we will only briefly review introductory level Linux system and basic operations. If you feel your Linux skills are lacking, please make sure to go through a Linux basics tutorial. Here is a great workbook for you to refresh the ground level operations: *Introduction to Unix*

Furthermore, you'll need to install command line utilities. Please see that you understand how to download, compile and run packages from your chosen distribution package manager, and from software hosted services such as GitHub.

We recommend installing Anaconda as your environment package manager as most tools (probably all that we will be using in the scope of this course) are available via the Bioconda channel. Working with Conda for bioinformatics will allow you to easily install packages and resolve dependency conflicts. Please review the BIOCONDA channel for more information, and how to install it for your system (which is just performing a one-time set up using the 3 commands stated in the link). A general tutorial with great background information can be found here: GCB 2020 Tutorial

For your use - a practical guide for working with Conda: Mayrose Lab – Conda Tutorial
Please also notice the "Useful links" section at the end.

The first "assignment" is a Bioinformatics Environment & Linux Skills Prep exercise. This means reading this "Computational Requirements and Prerequisites" document, making sure you have access to a UNIX based operating system/environment, refreshing your memory concerning Linux system and basic operations, and installing Anaconda as your environment package manager. We want to make sure you read this form and understand the necessary steps we require of you to perform (**by April 24**th). Therefore, please fill in the following sheet as you complete each step:

https://docs.google.com/spreadsheets/d/1g-eCN8EtwWCS7IKQuaURwkyeES2KwCTII5eFA5QOH0U/edit?usp=sharing

This course assignment relies on the honor system, so I won't be checking to see if you completed each task. But this part is just as important and crucial as all the other assignments, which is why it gets an equal weight as the others in the final grade (10%). Please make sure you create the necessary environment, and we encourage you to acquire a comprehensive understanding of everything I mentioned here. Feel free to get the help of ChatGPT, Perplexity or any other AI helper. If you encounter any problems or feel you cannot complete this by the due date, let me know.

I would like to welcome you to this exciting field of DNA, sequencing, and bioinformatics which has revolutionized how we do medicine in the past decade. We wish you a fruitful, productive and stimulating term.

Good luck and see you after Passover break!

Amit Levon, amit.levon@post.runi.ac.il

Prof. Noam Shomron, nshomron@gmail.com