Home

Introduction to Bioinformatics Programming and Analysis

Welcome to the home page for Introduction to Bioinformatics Programming and Analysis, developed by Ann Loraine in the Department of Bioinformatics and Genomics at UNC Charlotte. For more information about Ann Loraine and her research, visit the Loraine Lab home page.

The main goal of this class is to help you develop skills you would need in order to do the job of a bioinformatics programmer/analyst working in academia or industry.

Topics

UNIX and shell scripting

- UNIX tools for data exploration and running analyses
- Scripting bioinformatics pipelines
- Simple systems administration tasks everyone should know
- How to set up a simple Web site on your own UNIX server

Programming (python)

- Interactive programming in python
- How to write (and run) simple scripts
- Testing programs and program outputs
- Version control for managing code and analyses

Analyzing data (R and BioConductor)

- Using R/Bioconductor
- Microarray data analysis
- RNA-Seq data analysis

Techniques for reproducible research

- Survey of literate programming tools
- Unit testing (in python)

Objectives

The main goal of this class is to introduce you to core computer skills bioinformatics programmer/analysts use on a daily basis. By the end of the class, you'll have developed a strong sense of what these skills are and how bioinformatics professionals use them in research. Practically speaking, this means that when you read a job ad for a bioinformatics programmer, you'll be able to recognize the technical terms and explain what they mean. Of course, the best way to learn about a technology is to use it, and so this class will give you a chance to practice these skills through guided assignments designed to be accessible to computer-savvy non-programmers. Throughout the course, you'll be working in a Unix enviornment, a total immersion experience designed to give you the number one skill needed by all bioinformatics researchers: the ability to work productively in a Unix-based, command-line environment. By the end of the class, you'll have gained mastery over the basics of working in Unix. We'll accomplish this goal by giving you access to Unix virtual machines that allow you to experiment, make mistakes, and start over as needed during the course of the semester.

Topics we'll cover include:

- techniques for ensuring reproducible and accurate research, including:
    - version control and how scientists and software engineers use it
    - identifying and testing assumptions about data

- validating code through testing
- understanding basic data types in bioinformatics (genes, reads, alignments, etc) and file formats we use to represent them, including
  - fasta (sequence data)
  - fastq (sequence and quality scores)
  - BAM/SAM
  - BED (for annotations)
- pipelining data processing and analysis steps with commonly used command-line tools
- working productively in a UNIX environment
  - simple systems administration
  - working at the command line
- procedural programming (scripting) in python
- useful features and libraries available in python including
  - object-oriented regular expressions
  - BioPython
  - PyUnit (for unit testing)

Textbook (for UNIX section)

Learning the bash Shell, 2nd Edition - O'Reilly Press

Slides and attachments

For slides or other materials, click **Tools > Attachments** to get a copy.

Topics

See child pages below for a listing of classes, topics, and exercises.