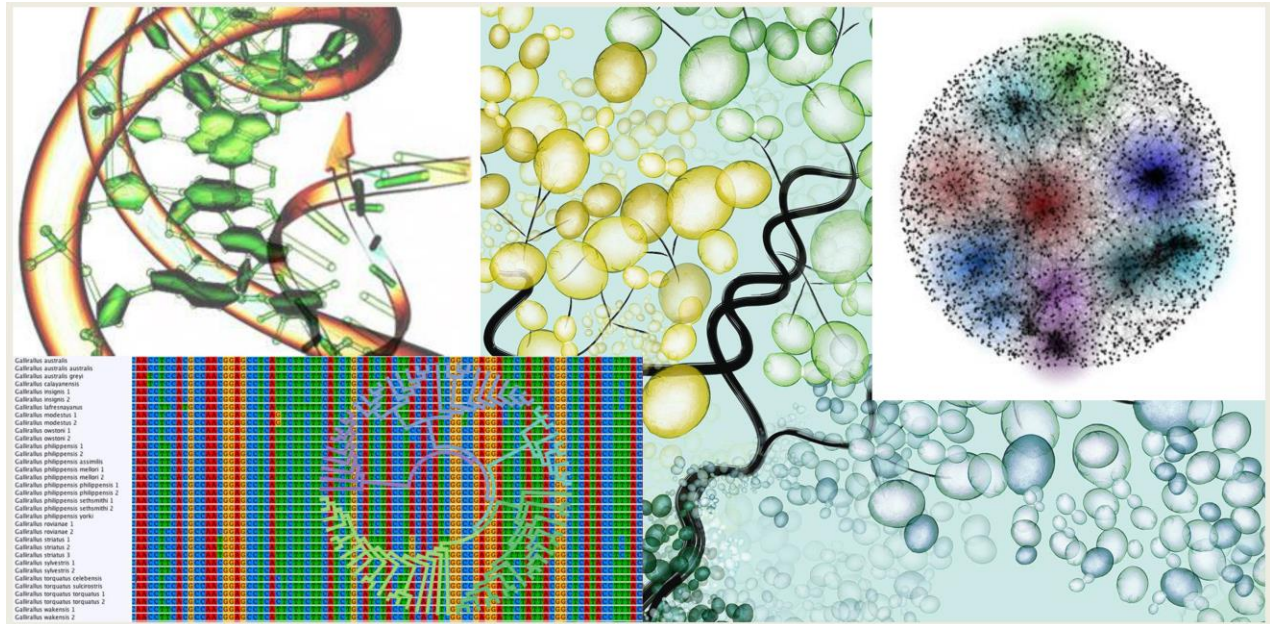# Introduction to Bioinformatics

BIOL-130   FALL 2022   Instructor: Dr. Gregory A Babbitt

contact Gregory A Babbitt, Office - Gosnell 08-1334, email - gabsbi@rit.edu

ZOOM OFFICE HOURS (by apt no passcode) https://rit.zoom.us/j/5727855006

SUMMARY

This course will introduce students to the background molecular biology, genetics and evolution needed to understand bioinformatics and also will introduce current databases and software tools employed in bioinformatics analyses. Students will work with DNA sequence databases, the PDB protein data bank, KEGG Drug Knowledgebase and some common software tools employed on them (UCSF Chimera/ChimeraX and MEGA). THIS IS A PURELY INTRODUCTORY COURSE. Computer programming experience or previous college level biology coursework is not expected nor required, however students with quite varied programming skills often take this course. This course is designed to provide basic domain knowledge in biology and bioinformatics to both 1st or 2nd year bioinformatics program students, as well as 3rd or 4th year students from other programs such as computer science or engineering. In class assignments, students will use scripting to conduct comparative genomics on the yeast genome and will use protein modeling software to conduct research on a drug-target protein interaction of their own choosing. A laptop computer is required. This course will be partly activity based.

OPTIONAL TEXTBOOK / MATERIALS

The textbook for the course is
Introduction to Bioinformatics by Authur Lesk (Oxford University Press)
ISBN 9780199651566

Throughout the course we will use the KEGG Knowledgebase, Protein Data Bank, GenBank at NCBI and several freely available software packages MEGA 10.0 and UCSF Chimera 1.14
KEGG   https://www.genome.jp/kegg/
MEGA   http://www.megasoftware.net/
UCSF Chimera   https://www.cgl.ucsf.edu/chimera/
UCSF ChimeraX   https://www.rbvi.ucsf.edu/chimerax/

TEACHING PHILOSOPHY

The mission of the modern university is to produce and disseminate new knowledge to the world. As an instrument towards this goal, I am an active researcher in the areas of molecular evolution and computational structural biology. However, I enjoy my role as a teacher here at RIT just as much as my research. I strive to do both of these tasks well. I ask each of you to take responsibility for the continued development of your own education. We are always here to help you, but ultimately the path of our learning is walked alone in each of our own minds. You must be actively committed towards this personal goal for yourself if you want to get a good grade in this class. Self-study is very important. Simply showing up without actively engaging your mind will not let you achieve success. Please communicate with me early in the semester if you feel you are having trouble understanding any concepts or material. All of us, PhD's included, have gaps in our understanding of various subjects. There is no such thing as a "dumb" question in my class. Success in academics at every level is often simply a matter of recognizing where missing information is holding us back. As an instructor, I will always try my best to achieve clarity by communicating to the least informed member of my audience, but I may miss the mark at times. During the lecture or lab, please don't feel that you cannot interrupt to ask me questions. **While I am with you, my time is always your time.**

Unlike some of the other natural sciences which are more mathematically-based, concepts in biology are often expressed in language through a specialized vocabulary. I highly recommend that you keep a vocabulary list while reading prior to class and attending lecture.

ATTENDANCE
Welcome to college, where YOU the student are ultimately paying for my attendance in class. Your attendance at lectures and/or studio sessions is hoped for, and probably required for a good grade but not monitored. There will always be information presented that is not otherwise available, i.e. it will not be in the books, and will be needed in projects, exams and/or quizzes. It is a good idea to be proactive if you are going to miss a class, and contact me to discuss material that may be missed.

GRADES
Grades will be based upon several in-class studio assignments, 3 midterm exams, a short final project, and an optional final exam.

HONESTY POLICY
PLEASE REVIEW THE RIT POLICY ON ACADEMIC DISHONESTY AND STUDENT MISCONDUCT AT THE WEBSITE BELOW

http://www.rit.edu/studentaffairs/studentconduct/rr_academicdishonesty.php

WE WILL ABIDE BY THESE POLICIES AND RECOMMENDED PENALTIES...NO EXCEPTIONS.

GRADE DISTRIBUTION – (60% 3 unit exams, 40% 7 assignments)
Note: grades are scaled as follows

A 93–100 | A- 90-93 | B+ 86–90 |B 83-86 | B- 80 – 83|C+ 76 – 80 | C 73 – 76 |C- 70 – 73 |D 60-70 | F <60

LECTURE AND STUDIO SCHEDULE ('studio' sessions will require students to bring a laptop computer, however all software is open source)

In-class studio assignments in blue (due one week after date in syllabus)

Unit 1 – bioinformatics (background and basics)

Week 1
A)   information and entropy (thermodynamics), central dogma of molecular biology
B)   Chemical bonding, DNA, RNA, and protein structure
C)   Gene and genome structure, protein synthesis and the genetic code

Week 2
A) genome differences in prokaryotes and eukaryotes, gene regulation and expression
B) NCBI, GenBank and Protein Data Bank databases
C) (studio 1) Intro to PDB, UCSF Chimera, and KEGG Knowledgebase

Week 3
A) no class
B) technologies behind the data – sequencing, sequencing assays, x-ray crystallography, NMR, cryo-EM
C) molecular evolution - genetic variation, mutation, selection, and population structure

Week 4
A) molecular evolution - genetic variation, mutation, selection, and population structure
B) (studio 2) Models of molecular evolution in MS Excel
C) open review

Week 5
A) in class **Exam I**

Unit 2 – bioinformatics (what to do with sequences and structures)

B) molecular evolution – alignment, homology and phylogenetic inference
C) (studio 3) Intro to GenBank and MEGA – multiple sequence alignment and quality assessment

Week 6
A) phylogenetics – inferring tree topology, likelihood, Markov model selection and bootstrapping
B) phylogenetics – inferring tree topology, likelihood, Markov model selection and bootstrapping
C) (studio 4) MEGA assignment – bootstrapped ML species tree (i.e. orthologs)

Week 7
A) (studio 4) MEGA assignment – bootstrapped ML gene tree (i.e. paralogs)
B) (studio 4) MEGA assignment – tests of natural selection acting on a gene
C) (studio 4) MEGA assignment – mutational tolerance calculations and disease prone genes

Week 8
A) protein enzymes, allostery, metabolism and cell signal transduction
B) metabolic and signaling pathways and mechanisms
C) (studio 5) Michaelis-Menton enzyme kinetics and KEGG drug database assignment

Week 9
A) structural pharmacology and drug discovery
B) structural comparison, molecular dynamics, biophysics
C) (studio 5 continued) Michaelis-Menton enzyme kinetics and KEGG drug database assignment

Week 10
A) human genetic variation and personalized and precision medicine
B) open review
C) project-based take home **Exam II**

Unit 3- computer programming in bioinformatics

Week 11
A) comparing syntax and control structures in python, perl and R
B) writing/running scripts from command line, reading bioinformatics file types
C) (studio 6) scripting to find sequence pattern matches using RegEx (perl and python)

Week 12
A) molecular modeling with UCSF Chimera and ChimeraX
B) molecular dynamics with UCSF Chimera and other programs
C) (studio 7) python scripting to control UCSF Chimera/ChimeraX

Week 13
A) gene regulatory evolution, protein evolution and protein space
B) new technologies – sequencing, structure determination, imaging, machine learning
C) new technologies – sequencing, structure determination, imaging, machine learning

Week 14
A) coding-based take home **Exam III**
B) (open studio) help on late assignments and Exam III
C) (open studio) help on late assignments and Exam III

Week 15 and Finals week
A) TBA (optional in class **FINAL EXAM** to replace 1 midterm exam grade)