

**INTRODUCTION TO COMPUTATIONAL BIOLOGY: BIOL 5800/6800
FALL 2019**

INSTRUCTOR: Dr. Scott R. Santos, Department of Biological Sciences
(844-7410, santos@auburn.edu)

OFFICE: 101 Life Sciences. **Office hours:** 8 AM-10:00 AM Fridays

LECTURES: Lectures are from **4-5:15 PM on MW** in **Science Center Classroom (SCC) 118**.

REQUIRED TEXTS:

Barrett, D.J. - "Linux Pocket Guide" by O'Reilly Book (Amazon for \$7.99; Book-A-Million for \$9.95)

Haddock, S.H.D. and Dunn, C.W. - "Practical Computing for Biologists" (Paperback) (\$60.00 from Amazon)

ALTERNATIVE TEXT:

Wünschiars, R. - "Computational Biology: A Practical Introduction to BioData Processing and Analysis with Linux, MySQL, and R" (Paperback) (\$80.00 from Amazon)

STUDENTS ARE REQUIRED TO HAVE THEIR OWN WINDOWS OR APPLE LAPTOP FOR USE IN CLASS. ALL FILES ON THE LAPTOP SHOULD BE BACKED UP PRIOR TO THE FIRST DAY OF CLASS AND THERE SHOULD BE (AT LEAST) 20GB OF FREE HARD DRIVE SPACE ON THE MACHINE.

PREREQUISITES: Background in molecular genetics and statistics is strongly required. It is expected that everyone has already taken some type of statistics course as well as courses in molecular biology, etc. Lectures will assume that everyone is comfortable with these subjects. If you are not, I would suggest dropping the course and taking it at a later date.

COURSE DESCRIPTION: Over the last decade, biological data sets have been rapidly growing in size and complexity. This course focuses on how to use computers to streamline the analysis of biological data, with an emphasis on "working smart" rather than hard. Developing a solid background in how computers can facilitate biological research will not only help with your thesis projects, but will also make you more "marketable" for post-docs and faculty positions in the future.

COURSE-SPECIFIC GOALS AND OBJECTIVES

Throughout the semester, students are expected to:

1. Learn basic LINUX/UNIX command line utilities
2. Generate quick one- and multi-line line scripts for data analysis
3. Develop basic programming skills in shell and Perl
4. Perform basic statistical analysis of biological data using R
5. Produce publication quality graphics using R and Inkscape

DEPARTMENT OF BIOLOGICAL SCIENCES STUDENT LEARNING OBJECTIVES (SLOs) COVERED BY THIS COURSE ARE:

SLO 2: Students will be able to effectively locate, evaluate, and summarize published information in the life sciences. They will be able to identify appropriate sources for specific information needs and use appropriate search tools and search strategies to access needed

information.

SLO 4: Students will be able to communicate effectively to the appropriate audience in the written genres common to biological sciences (e.g., lab reports, research articles). Components of written communication effectiveness include logical organization of data and ideas, appropriate language use, and correct use of scientific citations.

SLO 5: Students will be able to apply the scientific method to formulate testable hypotheses, gather data that address the hypotheses, and analyze the data (statistically, graphically) to assess the degree to which their scientific work tests their hypotheses and draw appropriate conclusions from the data.

SLO 6: Students will develop metacognitive skills and be able to distinguish between broad categories of metacognition as applied to their major. In particular, they will distinguish between foundational (i.e., knowledge recall) and higher order (i.e., creative, analysis, synthesis) metacognitive skills.

ADDITIONAL STUDY AIDS FOR THIS COURSE:

Additional reading materials, in the form of PDFs and HTML documents, will be used to supplement readings from the above books. These will be distributed either via email attachments or posted links from a computer server. These materials will be made available on the Friday following the class that they are assigned in.

IMPORTANT INFORMATION OF SPECIAL NOTE:

• **Read the assigned materials prior to coming to lecture.** The lectures are meant to clarify and discuss concepts, not to serve as your first exposure to them. Each lecture is presented with the assumption that you have read the material and are at least vaguely familiar with it.

Since the lectures will tend to build upon the previously covered materials, it is highly recommended that you review your notes on a weekly basis and attend lectures on a regular basis. Since this class only meets once a week, **ATTENDANCE IS MANDATORY. Failure to attend a class without prior notification will result in the loss of a letter grade (see below).**

If needed, I encourage you to make use of my office hours for additional instruction/discussion.

INFLUENZA H1N1 CONTINGENCY PLAN:

If normal class and/or lab activities are disrupted due to a high number of students experiencing illness or an emergency or crisis situation (such as a widespread H1N1 flu outbreak), the syllabus and other course plans and assignments may be modified to allow completion of the course. If this occurs, there is the potential that modified course assignments will replace the original materials.

GENERAL POLICY and PROCEDURES: You should retain this schedule of lecture topics and relevant instructions for reference throughout the semester. You are responsible for learning the material that will be covered, for preparing for lectures by reading assignments beforehand, and for being present at all lectures without further notice or additional reminders.

Missing classes should be avoided at all cost. Valid reasons for absences as outlined in the Tiger Cub are: 1) **severe** illness, 2) documented personal or family emergencies, 3) official University excuses. Illness will necessitate a note from the doctor or infirmary (as will a family emergency). Official University

excuses will likewise require documentation (see Tiger Cub, p. 86). Preferably, advise me beforehand regarding absences.

CHEATING: VIOLATION OF THE UNIVERSITY ACADEMIC HONESTY CODE WILL BE DEALT WITH AS OUTLINED IN THE SGA CODE OF LAWS, TITLE XII (TIGER CUB, pp. 125-126). All forms of academic dishonesty will be reported to the Academic Honesty Committee. This may result in failing grade, suspension, and/or expulsion from the University. These are serious situations, and any discovered attempt at academic dishonesty will be treated as extremely grave. (Note this includes turning in an excuse for an absence that cannot be verified as true).

SPECIAL ACCOMMODATIONS: Students who need special accommodations are encouraged to see me after class or in my office so we can discuss your situation, **confidentially**. Please bring your memo from The Program for Students with Disabilities (PSD) to me as soon as possible; we can discuss it during your appointment. Exam accommodations should be arranged at least one week in advance. If at any time during the semester you feel that the accommodations we have put in place are not working, please consult with me and /or the professional staff in the PSD office. If you do not have a memo from the PSD office that alerts me about your accommodations, it is recommended that you make an appointment to see them in **1232 Haley Center (844-2096)**. Without this memo no accommodations can be granted.

Special Request: Cell phones and pagers should be turned off for the duration of the lecture. Students will be asked to leave the classroom for the remainder of the lecture in the event one of these devices is activated during the lecture.

GRADING: There will be no structured midterms or final exam. Instead, “pre” and “post” tests will be use to assess the knowledge that students have gained by taking the course. Along with this, “mini” projects will be given on a weekly basis in order for students to practice the skills that they have ben exposed to in class.

**LECTURE SCHEDULE BY SUBJECT MATTER
(TENTATIVE AND SUBJECT TO CHANGE)**

Introduction and VirtualBox

Installation of Linux (PCs) into VirtualBox

The “shell” and basic UNIX commandline

The structure of the UNIX file system

Simple programs and how to automate commands
via scripts (shell, PERL, etc)

Compiling programs from scratch (using EMBOSS
as an example)

Analysis of molecular data

Statistics with R

Graphics with R

Figure and map creation using GMT

Image files: what’s the difference between them?

Creation of graphs and figures for publications
using GIMP and Inkscape