# Computation for 21st Century Biologists – BIOL 4590/5590; CMSS 5590/6590

## Department of Life Sciences

## Fall 2019

## A. COURSE INFORMATION

**Course number/section: BIOL 4590/5590; CMSS 5590/6590**
**Class meeting time: F 1-3:30**
**Class location: CCH-207, WEBEX https://tamucc.webex.com/meet/cbird**
**Course Website: https://github.com/comp-bio-fall-2019**
**https://bb9.tamucc.edu/**

## B. INSTRUCTOR INFORMATION

**Instructor: Dr. Christopher E. Bird**
**Office location: TH 234**
**Office hours: TWTh 3-5 **Coding Café In Library Coming Soon****
**Telephone: 361-825-6024 (office), 361-443-5676 (cell)**
**e-mail: chris.bird@tamucc.edu**
**Appointments: arrange via email or text**

## C. COURSE DESCRIPTION

This is a 3 credit course for upper-level undergraduates and graduate students that introduces the powerful open-source computing tools that are used in biological research for the creation, organization, manipulation, processing, analysis, and archiving of "big data". This course is designed to prepare and enable students to use computational tools for bioinformatic applications in advanced courses and independent research projects. The primary topics covered are: data formats and repositories, command line Linux computing and scripting, regular expressions, super-computing, computer programming with PYTHON and R, data visualization with R, version control and dissemination of scripts and programs with GIT, typesetting with LATEX, and organizing data with SQL relational databases.

## D. PREREQUISITES AND COREQUISITES

While not a formal requirement, it is assumed that students have a firm command of basic algebra.

## E. REQUIRED TEXTBOOK(S), READINGS AND SUPPLIES

Computing skills for biologists: a toolbook. Allesina & Wilmes 2019.

A computer designed for content creation (Linux, OSX, Windows, not chrome, not iOS, not Android). Course will be delivered in a computer lab for those that need a computer.

## F. STUDENT LEARNING OUTCOMES AND ASSESSMENT

Assessment is a process used by instructors to help improve learning. Assessment is essential for effective learning because it provides feedback to both students and instructors. A critical step in this process is making clear the course's student learning outcomes that describe what students are expected to learn to be successful in the course. The student learning outcomes for this course are listed below. By collecting data and sharing it with students on how well they are accomplishing these learning outcomes, students can more efficiently and effectively focus their learning efforts. This information can also help instructors identify challenging areas for students and adjust their teaching approach to facilitate learning.

Upon the successful completion of this course, students should be able to:

1. Recognize, describe, and organize data into standard biological data structures
2. Locate scientific data repositories and extract data
3. Operate UNIX/LINUX computers from command line
4. Construct and modify computer programming/scripting logic structures for processing biological data
5. Use version control software (git)
6. Describe and use regular expressions to query data
7. Typeset with LaTeX or MarkDown
8. Use the most popular open-source tools for biological data manipulation
    a. Shell scripting (bash)
    b. Scientific computing (python)
    c. Statistical computing (R)
    d. Tool repositories

## G. INSTRUCTIONAL METHODS AND ACTIVITIES

Computation for 21st Century Biologists will convene on Fridays at 1pm for 2.5 hours. Class periods will involve interactive lectures that require each student to have a computer designed for content creation (Linux, OSX, Windows, not chrome, not iOS, not Android). Homework exercises will embellish upon concepts addressed in lecture. **Participation** involves attending lectures and performance on unannounced quizzes. Weekly **Assignments** will be given to reinforce concepts covered in lectures and encourage students to start using computational tools. **Exams** will be used to evaluate comprehension of the materials covered in lectures and assignments. For *undergraduates only*, a comprehensive **Final Exam** will be used to assess the learning objectives detailed above.

Rather than having a final exam, *graduate students* are expected to complete a **Final Project** involving the automation of the manipulation and/or analysis of data, The code should be archived on GitHub. A report written in Latex or Markdown will be due during the final exam period. The report should be concise in stating what the problem is, describing the strategy used for the solution, and describing how the code works (be sure to include a flow-chart or outline describing what code does). Each student will give a 10-minute presentation during the Final period on their project.

Project examples: automatically process data from experimental apparatus; image analysis; automated reporting of experimental results; downloading and organizing data from online repositories; etc…

## H. MAJOR COURSE REQUIREMENTS AND GRADING

Student learning outcomes will be assessed using in class exercises, semester-long assignments, and exams. Your final grade will be based on the percentage you earn out of the total possible points, extra points may be built into exams or other assignments. It is also possible to lose points by turning in assignments late. Statistical manipulations to adjust grades, *if* used (at the Instructor's discretion), will be performed for each exam individually and all assignments in aggregate. A standard grading scale will be used:

A = 90 - 100 %
B = 80 - 89.9 %
C = 70 - 79.9 %
D = 60 - 69.9 %
F = 0 - 59.9 %

**Undergraduates:**

| ACTIVITY | % of FINAL GRADE |
|---|---|
| Participation | 15 |
| Assignments | 40 |
| Exam 1 | 12.5 |
| Exam 2 | 12.5 |
| Final Exam | 20 |

**Graduates:**

| ACTIVITY | % of FINAL GRADE |
|---|---|
| Participation | 10 |
| Assignments | 20 |
| Exam 1 | 10 |
| Exam 2 | 10 |
| Final Project | 50* |

* See section I for breakdown of credit for Final Project

## I. COURSE CONTENT/SCHEDULE

| Date | Lecture Topic | HW Due |
|---|---|---|
| | **Theme I: Welcome to the Matrix** | |
| **Wk 1** | 1. Course overview<br>2. Biological Data Repositories, Structures, Formats<br>3. Computer set up | |
| **Wk 2** | Linux Boot Camp I (Ch 1)<br>1. UNIX philosophy<br>2. Navigating/creating/manipulating directories & files<br>3. How to get help: man pages<br>4. Basic commands,<br>5. Commands useful for manipulating data files in text streams | Assignment 0 |
| **Wk 3** | Linux Boot Camp II (Ch 1)<br>1. Manipulating text streams with grep, sed, cut, tr, paste, cat<br>2. Advanced regular expressions (grep) | Assignment 1 |
| **Wk 4** | Linux Boot Camp III (Ch 1)<br>1. Intro to Computer Programming & Scripting with bash<br>2. Shebang!, if-then-else; looping with for, while, and GNU parallel<br>3. Functions: diy tools | Assignment 2 |
| **Wk 5** | Version Control & Supercomputing (Ch 2)<br>1. Linux repositories and tools for biologists<br>2. Version control with git<br>3. Super computing | Assignment 3 |
| | **Theme II: Programming the Matrix** | |
| **Wk 6** | Python Boot Camp I (Ch 3)<br>1. Intro to Python<br>2. Data structures<br>3. Functions<br>4. Decision logic and loops<br>5. Reading and writing files | **Exam 1**<br>Install Anaconda & Jupyter |
| **Wk 7** | Python Boot Camp II (Ch 4)<br>1. Writing code<br>2. Modules & Program Structure<br>3. Errors and exceptions<br>4. Debugging<br>5. Testing & Profiling | Assignment 4 |
| **Wk 8** | Scientific Computing w/ Python (Ch 6)<br>1. NumPy and SciPy<br>2. Pandas | Assignment 5 |

| | | |
|---|---|---|
| | 3. Biopython<br>4. Other modules | |
| **Wk 9** | Scientific Typesetting w/ Latex (Ch 7)<br>1. Latex document structure<br>2. Typsetting<br>3. Latex packages for biologists | Assignment 6 |
| | **Theme III: Becoming THE ONE** | |
| **Wk 10** | R Boot Camp I (Ch 8)<br>1. R Philosophy<br>2. Similarities and differences from bash and python<br>3. Installing R & R Studio<br>4. R data structures<br>5. Reading and writing data<br>6. Scripting<br>7. Logic structures | **Exam 2**<br>Install R & R Studio |
| **Wk 11** | R Boot Camp II (Ch 8)<br>1. Functions<br>2. Libraries<br>3. Random numbers<br>4. Vectorize<br>5. Debugging<br>6. Stats<br>7. Plots<br>8. Packages | Assignment 7 |
| **Wk 12** | R Boot Camp III (Ch 9)<br>1. tidyverse<br>2. Manipulating data<br>3. Computing statistics<br>4. Wrangling data<br>5. Visualization of data | Assignment 8 |
| **Wk 13** | Relational Databases (Ch 10)<br>1. Intro to relational data bases<br>2. SQLite<br>3. Designing databases<br>4. Working w/ databases<br>5. Scripting<br>6. Interfacing with databases from R and Python | Assignment 9 |
| **Wk 14** | Putting It All Together (Ch 11) | Assignment 10 |
| **Final** | **Final Exam: "Welcome to the real world."** | |

| Date | Final Project Schedule |
|---|---|
| | **Theme I: Welcome to the Matrix** |
| **Wk 1** | |
| **Wk 2** | |
| **Wk 3** | Submit Project Idea (2.5%) |
| **Wk 4** | |
| **Wk 5** | Submit Project Plan/Outline (2.5%) |
| | **Theme II: Programming the Matrix** |
| **Wk 6** | Link to GitHub repository for project w/ readme (2.5%) |
| **Wk 7** | Commit at least 1 working function to GitHub (2.5%) |
| **Wk 8** | Commit at least 2 working functions with data I/O to GitHub (2.5%) |
| **Wk 9** | |
| | **Theme III: Becoming THE ONE** |
| **Wk 10** | |
| **Wk 11** | Latex or Markdown draft/ progress report (see syllabus section G for description of report); include a description of tasks left to achieve. Code and data committed to GitHub (2.5%) |
| **Wk 12** | |
| **Wk 13** | |
| **Wk 14** | Final Report in Latex or Markdown; Working code and data committed to GitHub. (25%) |
| **Final** | Final Presentations (10%) |

## J. COURSE POLICIES

**Attendance/Tardiness:** Attendance is expected. If you are late, don't make a disturbance.

**Late Work and Make-up Exams:** 10% of total possible score is deducted per day late. Inform professor as soon as you find out that you will miss and exam. Make arrangements with professor for make up.

**Cell Phone, Tablet, and Laptop Use:** Required

**Food in Class:** food is not allowed in computer labs

**Participation:** Required

## K. COLLEGE AND UNIVERSITIY POLICIES

- **Academic Integrity (University)**
  University students are expected to conduct themselves in accordance with the highest standards of academic honesty. Academic misconduct for which a student is subject to penalty includes all forms of cheating, such as illicit possession of examinations or examination materials, falsification, forgery, complicity or plagiarism. (Plagiarism is the presentation of the work of another as one's own work.) In this class, academic misconduct or complicity in an act of academic misconduct on an assignment or test will result in a failing grade.

- **Classroom/Professional Behavior**
  Texas A&M University-Corpus Christi, as an academic community, requires that each individual respect the needs of others to study and learn in a peaceful atmosphere. Under Article III of the Student Code of Conduct, classroom behavior that interferes with either (a) the instructor's ability to conduct the class or (b) the ability of other students to profit from the instructional program may be considered a breach of the peace and is subject to disciplinary sanction outlined in article VII of the Student Code of Conduct. Students engaging in unacceptable behavior may be instructed to leave the classroom. This prohibition applies to all instructional forums, including classrooms, electronic classrooms, labs, discussion groups, field trips, etc.

- **Statement of Civility**
  Texas A&M University-Corpus Christi has a diverse student population that represents the population of the state. Our goal is to provide you with a high quality educational experience that is free from repression. You are responsible for following the rules of the University, city, state and federal government. We expect that you will behave in a manner that is dignified, respectful and courteous to all people, regardless of sex, ethnic/racial origin, religious background, sexual orientation or disability. Behaviors that infringe on the rights of another individual will not be tolerated.

- **Deadline for Dropping a Course with a Grade of W (University)**
  I hope that you never find it necessary to drop this or any other class. However, events can sometimes occur that make dropping a course necessary or wise. ***Please consult with your academic advisor, the Financial Aid Office, and me, before you decide to drop this course.*** Should dropping the course be the best course of action, you must initiate the process to drop the course by going to the Student Services Center and filling out a course drop form. Just stopping attendance and participation WILL NOT automatically result in your being dropped from the class. Please consult the Academic Calendar (http://www.tamucc.edu/academics/calendar/) for the last day to drop a course.

- **Grade Appeals (College of Science and Engineering)**
  As stated in University Procedure 13.02.99.C2.01, Student Grade Appeal Procedures, a student who believes that he or she has not been held to appropriate academic standards as outlined in the class syllabus, equitable evaluation procedures, or appropriate grading, may appeal the final grade given in the course. The burden of proof is upon the student to demonstrate the appropriateness of the appeal. A student with a complaint about a grade is encouraged to first discuss the matter with the instructor. For complete details, including the responsibilities of the parties involved in the process and the number of days allowed for completing the steps in the process, see University Procedure 13.02.99.C2.01, Student Grade Appeal Procedures. These documents are accessible through the University Rules website at http://www.tamucc.edu/provost/university_rules/index.html, and the College of Science and Engineering Grade Appeals webpage at http://sci.tamucc.edu/students/GradeAppeal.html. For assistance and/or guidance in the grade appeal process, students may contact the chair or director of the appropriate department or school, the Office of the College of Science and Engineering Dean, or the Office of the Provost.

- **Disability Services**
  The Americans with Disabilities Act (ADA) is a federal anti-discrimination statute that provides comprehensive civil rights protection for persons with disabilities. Among other things, this legislation requires that all students with disabilities be guaranteed a learning environment that provides for reasonable accommodation of their disabilities. If you believe you have a disability requiring an accommodation, please call (361) 825-5816 or visit Disability Services in Corpus Christi Hall 116.

  If you are a returning veteran and are experiencing cognitive and/or physical access issues in the classroom or on campus, please contact the Disability Services office for assistance at (361) 825-5816.

  http://disabilityservices.tamucc.edu/

- **Statement of Academic Continuity**
  In the event of an unforeseen adverse event, such as a major hurricane and classes

could not be held on the campus of Texas A&M University–Corpus Christi; this course would continue through the use of Blackboard and/or email. In addition, the syllabus and class activities may be modified to allow continuation of the course. Ideally, University facilities (i.e., emails, web sites, and Blackboard) will be operational within two days of the closing of the physical campus. However, students need to make certain that the course instructor has a primary and a secondary means of contacting each student.

## A. OTHER INFORMATION

- **Academic Advising**
  The College of Science & Engineering requires that students meet with an Academic Advisor as soon as they are ready to declare a major. The Academic Advisor will set up a degree plan, which must be signed by the student, a faculty mentor, and the department chair. Meetings are by appointment only; advisors do not take walk-ins. Please call or stop by the Advising Center to check availability and schedule an appointment. The College's Academic Advising Center is located in Center for Instruction 350 or can be reached at (361) 825-3928.

## GENERAL DISCLAIMER

I reserve the right to modify the information, schedule, assignments, deadlines, and course policies in this syllabus if and when necessary. I will announce such changes in a timely manner during regularly scheduled lecture periods.

**B.**