GENERAL BIOLOGICAL INFORMATICS

Class Meeting Times: Mon & Thu 12:30 AM – 1:50 PM

Instructors
Dr. Justin M. Bernstein jbernstein@amnh.org

Office Location
TBD

Brief Summary of General Biological Informatics
Why study biological systems in the wild? How do we use modern technology to accomplish this task? In this course, we will learn about many organismal systems (groups of fish, mammals, birds, reptiles, amphibians) and use a combination of computers and data science methods to reinforce our understanding of key themes in evolutionary biology. We will review topics such as natural selection, genetic drift, systematics and phylogenetics, and many more. In addition to these reviews, I will teach participants of this course how to use computer coding to process and analyze biological datasets (bioinformatics). Students will be taught how computers are structured and how we can use coding practices to run tasks on our own computers to perform tasks that would otherwise be difficult manually. After this course, all students will not only leave with a greater theoretical knowledge of evolutionary biology and different organismal systems, but also have a different perspective on how we can study biological systems in different frameworks. Students will leave with tangible skills that are broader than the scope of this course, such as presentation and communication skills, database searching, and computer coding in different computer languages.

Course Objectives.
This course will teach students concepts in evolutionary biology and bioinformatic skills to reinforce these theories. Course sessions will involve a combination of PowerPoint lectures, coding specific analyses that serve to teach specific topics in biology, and hands-on exercises with groups such as reading or working with objects that represent evolutionary phenomena. The main objectives of this course are to:

(i) Study biological and evolutionary themes and topics through computer coding.

(ii) Learn about empirical research using published studies in the primary literature.

(iii) Obtain tangible skills for your career paths that are broadly applicable to all career paths.

Section Objectives.
Weekly sessions will serve to:

(i) Have students understand a new key topic in biology.

(ii) Learn a new computer skill and a new analysis that emphasizes the weekly topic.

(ii) Provide a breakdown of a group-chosen publication from primary literature.

(iii) Complete a 'mini quiz' about the computer coding done during the week.

Textbook. There is no textbook for this course, but several resources are available to get additional information on coding outside of class sessions:
Introduction to R: https://www.geeksforgeeks.org/r-programming-language-introduction/
Swirlstats in R: https://swirlstats.com/
Bash coding: https://www.freecodecamp.org/
NCBI GenBank: https://www.ncbi.nlm.nih.gov/genbank/

Grading. This course will be graded through a combination of participation and attendance (20%) and the completion of course projects (80%). Note, course projects will involve weekly coding sessions that are primarily group-based or done together as a class. I will administer either worksheets to assess your understanding of course material or students will submit code to me, which will be analyzed by myself and given a grade. The focus of this course is to obtain a better understanding of how we study biological systems, **NOT** to write the perfect code. If code fails, this does ***not*** mean a failure to complete an assignment. We are here to learn, together.

1) 20% attendance and participation
2) 80% course assignments
- a) 60% bi-monthly coding assignments, submitted to instructor
- b) 10% weekly worksheets reviewing key topics (5–10 questions)
- c) 10% brief, written summaries of primary literature through class discussions (group-based)

Lecture Attendance (20%): Students are strongly encouraged to attend lecture. While coding can be done at home and on the computer, classes will, in part, be spent explaining the theory behind biological concepts and the computer programs we use to understand them. While students can read about these concepts at home, the analyses we will run during sessions are niche software that can be difficult to run on your system. Coding should never be done on your own or any computer system without the knowledge of what code does, as this can risk the functioning of your own computer(s).

Bi-monthly coding assignments (60%) – Evolutionary topics can involve memorization of many definitions, making the broader themes difficult to understand. This course will highlight broader topics and, while it will involve definitions to learn, we will focus on using bioinformatics to enhance student understanding. Every other week, we will perform coding modules that will end in assignments to be handed in on a subsequent week. These assignments will involve individuals or groups writing code that will solve a puzzle or problem given by the professor (that represents a biological topic). Some assignments will also be 'chosen' by students, in that you will identify a system you find interesting and ask a question about a datatype that can be answered using bioinformatics. These student-chosen assignments will come later in the course once students have had time to learn what datatypes exist and how we can analyze them.

Weekly worksheets reviewing key topics (10%) – In addition to coding assignments, students will be given a weekly worksheet towards the end of the week. These will review biological and evolutionary concepts, or definitions in computer science/bioinformatics, to make sure students are understanding the weekly topics and are prepared for the subsequent week's work.

Written primary literature review and class discussions (10%) – While students will learn new topics, definitions, and methods in class, it is important to know how to search and understand primary literature. Students will find research papers of their choosing that pertains to the week's topic and write a ***brief*** (1–2 paragraphs) summary of this paper, and present it to the class. They are expected to not only present to the class, but to ***lead*** the class and learn how to keep a group engaged in discussion. These will be graded as a group, and groups will do this twice in the semester, with groups changing between the first and second presentation so to ensure students are learning with new classmates each time.

Assignment Submission Policy. Assignments will be turned in on given dates for each assignment. For coding assignments, 2 weeks will be given to complete the project, which is enough time for students to ask questions during class time or office hours. Points will be deducted per late day for these assignments. Weekly review worksheets will be done during class time and handed in at the end of class. Group-based discussions summaries will be given 1 week to complete after the discussion, but groups and presentation days will be decided on the first week of class to provide ample preparation time. I understand that students may require extension times or there may be unexpected illnesses or emergences that affect the submission of assignments. In these cases, please contact the Office of the Dean of Students (URL). To ensure student privacy and confidentiality, if you require resources and extensions due to personal reasons, please contact the Office of the Dean of Students, and let me know you have done so. You need not provide me any information other than that you contacted the Office of the Dean of Students. This office is located on (Location), and can be contacted at (Phone Number).

Students with Documented Disabilities

Students that may need academic accommodations based on disability impacts must reach out to the Disability Services office of (Institution) (URL). This will allow me to obtain the pertinent level of information from the university to provide these accommodations for you. This office is located at (Location), and can be reached at (Email) or (Phone Number).

Course Conduct and Community Learning

The overarching goal of this course is to work together to enhance our understanding of biological concepts and learn new skills. The key here is that we will be learning together, even through individual exercises. I encourage students to be themselves and be comfortable in class, and I have an expectation that all will maintain courtesy and respect for your fellow classmates and all other individuals. All students are required to follow the principles and guidelines of (Institution)'s Student Conduct (URL). By communicating with one another, respecting one another, and learning with one another, this course will progress through the semester in fun and collaborative ways!

Course Schedule:

Date	Day	Topic	Type
8/19	Tuesday	General Biological Informatics: Introduction	Group
8/21	Thursday	The Machine Behind Your Computer: Understanding Computers	Individual
8/26	Tuesday	Computer Hierarchy and Data Structure – PART 1	Individual
8/28	Thursday	Computer Hierarchy and Data Structure – PART 2	Individual

9/2	Tuesday	Learning the Basics of Bash – Coding Tutorials in the Terminal – PART 1 (file and directory manipulation)	Individual
9/4	Thursday	Paper Discussion	Group
9/9	Tuesday	Learning the Basics of Bash – Coding Tutorials in the Terminal – PART 2 (text editing, REGEX) loops)	Group
9/11	Thursday	Bash Practice: Coding Exercises and Practical	Individual
9/16	Tuesday	Introduction to the R Coding Language – Part 1 (RStudio, user interface, environments, data structure)	Individual
9/19	Thursday	Paper Discussion	Group
9/23	Tuesday	Introduction to the R Coding Language – Part 2 (loops and biological data)	Individual
9/25	Thursday	Understanding the intersection of biology and data science: bioinformatics – Part 1	
9/30	Tuesday	Understanding the intersection of biology and data science: bioinformatics – Part 2	

10/2	Thursday	Paper Discussion	Group
10/7	Tuesday	Introductions to Data Science: Bash Coding & DNA/RNA – PART 1	Individual
10/9	Thursday	Introductions to Data Science: Bash Coding & DNA/RNA – PART 2	Individual
10/14	Tuesday	Systematics & Phylogenetics: Maximum Likelihood	Group
10/16	Thursday	Systematics & Phylogenetics: Bayesian Inference	Group
10/21	Tuesday	Handling Molecular Datasets: Coding in R to Understand Natural Selection and Genetic Drift – PART 1 (theory and data structure)	Individual
10/23	Thursday	Handling Molecular Datasets: Coding in R to Understand Natural Selection and Genetic Drift – PART 2 (data and coalescent simulations)	Individual
10/28	Tuesday	Introduction to Machine Learning Algorithms	Individual
10/30	Thursday	Machine Learning Algorithms and Phenotypic Data: Identifying Intraspecific and Interspecific Variation	Group

11/4	Tuesday	Environmental Data and Bioclimatic Layers: How we use climatic data in bioinformatics	Group
11/6	Thursday	Ecological Niche Models – PART 1 (Downloading and Curating Species Occurrence Data)	Individual
11/11	Tuesday	Ecological Niche Models – PART 2 (Algorithms and Types)	Individual
11/13	Thursday	Ecological Niche Models – PART 3 (Assessing Model Performance and Output)	Individual
11/18	Tuesday	Paper Discussion	Group
11/20	Thursday	In-class open time: work on Final Coding Project	Individual
11/25	Tuesday	In-class open time: work on Final Coding Project	Individual
11/27	Thursday	In-class open time: work on Final Coding Project	Individual
12/2	Tuesday	In-class open time: work on Final Coding Project	Individual
12/4	Thursday	In-class open time: work on Final Coding Project	Individual

12/9	Tuesday	Present on Final Coding Project	Individual
12/11	Thursday	Present on Final Coding Project	Individual