Bioinformatics Algorithms
CSCE 4810.001 / CSCE 5810.001
BIOL 4810.001 / BIOL 5810.001
MATH 4810.001

Course Information & Syllabus

Instructor: Dr. Xuan Guo
Lecture Time: Mondays & Wednesdays, 10:40 AM – 12:00 PM, Gate 049
Student Hours: Mondays, 2:00 – 4:00 PM, or by appointment
Office Location: NTDP F290
Phone: 940-565-2817
Email: xuan.guo@unt.edu
Class Website: Canvas

Recommended Textbooks:

- Bioinformatics Algorithms, Phillip Compeau and Pavel A. Pevzner, Active Learning Publishers; 3rd edition (2018); ISBN-10: 0990374637; ISBN-13: 978-0990374633; URL: https://www.bioinformaticsalgorithms.org/
- Exploring Bioinformatics: A Project-Based Approach, Caroline St. Clair & Jonathan E. Visick, Jones & Bartlett Learning; 2nd edition, ISBN-10: 1284034240, ISBN-13: 978-1284034240

Course Description: Introduction to computational problems inspired by life sciences and overview of available tools. Methods to compute sequence alignments, regulatory motifs, phylogenetic trees and restriction maps.

This course introduces computational problems arising in molecular biology and the algorithms underlying modern bioinformatics tools. Students will learn to formulate biological questions as computational problems; design, implement, and analyze algorithms; and evaluate their performance on biological data. Topics include sequence alignment and similarity search, genome assembly and pangenome graphs, motif discovery and probabilistic sequence models, combinatorial optimization, machine learning and dimensionality reduction, single-cell and high-dimensional data analysis, phylogenetic inference, protein sequence analysis, and computational proteomics. Emphasis is placed on algorithmic reasoning, computational efficiency, validation, reproducibility, and critical evaluation of current research.

Course Prerequisites or Other Restrictions:
Required: Intermediate background in programming and data structures.

Recommended: Basic knowledge of molecular biology, discrete mathematics, probability and statistics, and linear algebra. Supplemental molecular biology and quantitative review resources will be provided through Canvas. Students who lack some of this background are expected to review the relevant materials as needed.

This course has digital components. To fully participate in this class, students will need internet access to reference content on the Canvas Learning Management System. If circumstances change, you will be informed of other technical needs to access course content. Information on how to be successful in a digital learning environment can be found at Learn Anywhere (https://online.unt.edu/learn).

Learning Objectives

By the end of this course, students will be able to:

- Formulate biological questions as computational problems by specifying inputs, outputs, assumptions, constraints, and appropriate evaluation criteria.
- Explain the biological context and data types underlying major problems in computational molecular biology.
- Analyze the correctness, time complexity, space complexity, assumptions, and limitations of bioinformatics algorithms.
- Design and implement dynamic-programming and sequence-analysis algorithms for alignment, similarity search, and related problems.
- Apply graph algorithms and combinatorial optimization methods to genome assembly, pangenome analysis, and other biological problems.
- Apply randomized and probabilistic methods, including motif-finding algorithms, Markov models, and hidden Markov models, to biological sequence data.
- Apply and critically evaluate machine-learning, dimensionality-reduction, and clustering methods for biological data, including single-cell and other high-dimensional datasets.
- Explain and apply computational methods for phylogenetic inference, protein sequence analysis, and mass-spectrometry-based proteomics.
- Evaluate computational methods using appropriate benchmarks, validation strategies, error measures, and statistical controls, including false discovery rate where applicable.
- Critically analyze current bioinformatics research and communicate algorithmic methods, experimental evidence, limitations, and conclusions through presentations and a final project.

Grading:

Undergraduate version		Graduate version	
Assignments	60%	Assignments	50%
Research Article Presentation 1	20%	Research Article Presentation 1	15%
Research Article Presentation 2	20%	Research Article Presentation 2	15%
		Project	20%
Attendance & Participation	1% bonus	Attendance & Participation	1% bonus
In-class practices	4% bonus	In-class practices	4% bonus

Grading Scale: A = 90-100, B = 80-89.99, C = 70-79.99, D = 60-69.99, F < 60

Tentative Schedule

The following is a list of topics we will attempt to cover during this course:

Week #	Description	Assignments
1	Biological Data, Problems, and Computational Thinking	
2	Algorithmic Foundations: Complexity, Search, and AI-Assisted Coding	HW#1

3-4	Sequence Alignment and Biological Similarity	
5-6	Graph Algorithms in Genomics: Assembly, de Bruijn Graphs, and Pangenomes	HW#2
6-7	Probabilistic Sequence Models: Motifs, Markov Models, HMMs, and Gene Prediction	HW#3
7-8	Integer Linear Programming	HW#4
8-9	Machine Learning for Bioinformatics: Prediction, Validation, and Biological Generalization	Research Article Presentation 1
10-11	High-Dimensional Biology: PCA, Clustering and Single-Cell Analysis	HW#5
12-13	Protein Sequence Algorithms, Evolutionary Analysis, and Introductory Phylogenetics	Research Article Presentation 2
13-14	Proteomics Algorithms: Spectrum Matching, FDR, DIA, and Learned Spectral Models	
15	Thanksgiving Break	
16	Project presentations & Project Reports	

The course calendar above is subject to change. Students will be notified by Eagle Alert if there is a campus closing that will impact a class and describe that the calendar is subject to change, citing the Campus Closures Policy (https://policy.unt.edu/policy/15-006).

Homework: Homework assignments will include both theoretical and programming exercises. There will be approximately five assignments, each with a duration of 1–2 weeks.

Research Article Presentations: Students may work individually or form a self-selected group of no more than two students. Each student or group will select and present two bioinformatics research papers from the list provided by the instructor; other papers may be used with prior approval. One paper will be presented during each presentation round. Presentations will be evaluated as a group effort, and both group members must contribute meaningfully to researching, preparing, and delivering each presentation. Each presentation will consist of a 15-minute overview covering the paper's background, motivation, research problems, methods, results, and conclusions, followed by a 5-minute Q&A of the computational methods, technical details, limitations, and possible improvements. Groups and papers must be selected by September 18. Presentations are scheduled for October 12 and November 2.

In-class Exercises: There will be approximately 10 in-class exercises, each lasting around 15-30 minutes. Students will work individually on a problem and submit their solutions during class. The schedule for these activities will be communicated via Canvas. The format of these questions will be similar to the homework assignments.

Submission: All assignments and projects must be submitted electronically via Canvas. Work submitted within 24 hours after the deadline will be eligible for 50% credit. No homework will be accepted after this extended deadline.

Attendance: Attendance will not be formally recorded but is expected. Students are responsible for all material covered in class. Active participation can earn you up to 1% bonus toward your

final grade. To receive full participation credit, you are expected to attend each class and engage by asking questions, answering questions, and participating in class activities.

Graduate Term Project: The goal of the term project is to apply computational methods to contemporary problems in biology. The project topic is open-ended; you may select a topic that is of interest to you or relevance in your profession. Some project ideas will be given by the instructor, too. Term projects can be done in groups up to 3. Each member must contribute equally to the project to avoid a lower score including 0.
The specific milestones of the term project are as follows. More information for each milestone will be made available on Canvas.

Milestone	**Due by (tentative)**	**Grade Weight**
Submit background/interests and choose teams	9/4	5%
Submit Project Title and 1 Paragraph Summary	10/9	10%
Submit Specific Aims and Research Strategy	10/23	15%
Submit Preliminary Results	11/6	20%
Oral Project Presentation	11/30	25%
Final Report	12/7	25%

Academic Integrity Policy:
According to UNT Policy 06.003, Student Academic Integrity, academic dishonesty occurs when students engage in behaviors including, but not limited to cheating, fabrication, facilitating academic dishonesty, forgery, plagiarism, and sabotage. A finding of academic dishonesty may result in a range of academic penalties or sanctions ranging from admonition to expulsion from the University.

Homework assignments are intended to reflect individual effort in the course. Students may discuss assignments in a general way, i.e., discussing the nature of the assignment or providing clarifications. However, sharing code (even a few lines), pseudo-code, program outputs, analysis results are strictly prohibited. Borrowing any code from other sources (internet, book, friend, AI-assisted tools, etc.) are not allowed unless stated otherwise by the instructor. The instructor may ask you to explain your logic for any source code that you upload to Canvas.

You are encouraged to refer to outside material such as journals, web pages, and books. Do not feel guilty about using outside material; just make sure that you cite your references. It is not acceptable to directly copy material from another source. Failure to properly cite your references may result in a charge of plagiarism. Give proper credit where credit is due! The instructor may cross-check submitted work via computational tools such as Turnitin to find highly similar submissions.

A Note on AI-assisted Technologies
AI tools may be used to support learning when authorized by the instructor. Authorization may differ by assignment. Each assignment will specify one of the following:

- **No AI use:** AI tools may not be used to complete any part of the assignment.
- **Learning support only:** AI may be used to explain concepts, generate practice questions, or help diagnose errors in student-written work. It may not generate code, analyses, answers, or prose submitted for credit.

- **AI use permitted with disclosure:** AI may assist with specified parts of the assignment, provided its use is fully documented.

If an assignment does not explicitly authorize AI use, students should assume that AI-generated code, analyses, answers, and prose are prohibited.
When AI is used, the submission must include an **AI-Use Appendix** containing:

1. The tool and model used, if known.
2. The relevant prompts and AI responses, submitted as an exported transcript, screenshots, or copied text.
3. Identification of any AI-generated material incorporated into the submission.
4. A brief reflection explaining:
    - what was trying to accomplish;
    - what suggestions were accepted, rejected, or modified;
    - what errors or limitations were identified; and
    - how the resulting work was independently verified.

Unrelated or private portions of a conversation may be redacted, but the record must contain every interaction that materially influenced the submitted work. Students must not enter confidential data, unpublished research data, another person's work, or personally identifying information into public AI systems. UNT similarly warns against sharing sensitive information with unapproved AI services.

Americans with Disabilities Act: We cooperate with the Office of Disability Accommodation to make reasonable accommodations for qualified students (cf. Americans with Disabilities Act and Section 504, Rehabilitation Act) with disabilities. If you have not registered with ODA, we encourage you to do so. If you have a disability for the accommodation, please discuss your needs with the instructor or submit a written Accommodation Request on or before the fourth-class day. The Office of Disability Access (ODA) can work with you on academic accommodation if needed. The University of North Texas makes reasonable academic accommodation for students with disabilities. Students seeking reasonable accommodation must first register with the Office of Disability Access (ODA) to verify their eligibility. If a disability is verified, the ODA will provide you with a reasonable accommodation letter to be delivered to faculty to begin a private discussion regarding your specific needs in a course. You may request reasonable accommodation at any time; however, ODA notices of reasonable accommodation should be provided as early as possible in the semester to avoid any delay in implementation. You must obtain a new letter of reasonable accommodation each semester and must meet with each faculty member prior to implementation in each class. For additional information, refer to the Office of Disability Access website (https://studentaffairs.unt.edu/office-disability-access). You may also contact ODA by phone at (940) 565-4323 or visit their website to begin the registration process for additional support.

Rules of Engagement: Rules of engagement refer to the way students are expected to interact with each other and with their instructors. Here are some general guidelines:

- While the freedom to express yourself is a fundamental human right, any communication that utilizes cruel and derogatory language on the basis of race, color, national origin, religion, sex, sexual orientation, gender identity, gender expression, age, disability, genetic information, veteran status, or any other characteristic protected under applicable federal or state law will not be tolerated.
- Treat your instructor and classmates with respect in any communication online or face-to-face, even when their opinion differs from your own.

- Ask for and use the correct name and pronouns for your instructor and classmates.
- Speak from personal experiences. Use "I" statements to share thoughts and feelings. Try not to speak on behalf of groups or other individual's experiences.
- Use your critical thinking skills to challenge other people's ideas, instead of attacking individuals.
- Avoid using all caps while communicating digitally. This may be interpreted as "YELLING!"
- Be cautious when using humor or sarcasm in emails or discussion posts as tone can be difficult to interpret digitally.
- Avoid using "text-talk" unless explicitly permitted by your instructor.
- Proofread and fact-check your sources.
- Keep in mind that online posts can be permanent, so think first before you type.

See these Engagement Guidelines (https://clear.unt.edu/online-communication-tips) for more information.

Acceptable Student Behavior: Student behavior that interferes with an instructor's ability to conduct a class or other students' opportunity to learn is unacceptable and disruptive and will not be tolerated in any instructional forum at UNT. Students engaging in unacceptable behavior will be directed to leave the classroom and the instructor may refer the student to the Dean of Students to consider whether the student's conduct violated the Code of Student Conduct. The University's expectations for student conduct apply to all instructional forums, including University and electronic classroom, labs, discussion groups, field trips, etc. Visit UNT's Code of Student Conduct (https://deanofstudents.unt.edu/conduct) to learn more.

Unauthorized Aids During Examinations and Assessments: Students may use only those materials and resources that are expressly authorized by the instructor for a particular examination or assessment. Touching, holding, or accessing any unauthorized aid during the examination period constitutes a violation of the Student Code of Conduct, regardless of whether the aid is used.
Any item that may be considered an unauthorized aid must be kept out of view and in a closed container (e.g. zipped backpack, buttoned purse, etc.) during the entirety of the examination or assessment.
Unless specifically authorized in writing by the instructor, prohibited aids include, but are not limited to:
• Mobile phones and smartphones
• Smartwatches and fitness trackers with communication, recording, or internet capabilities
• Tablets, laptops, and other computing devices
• Earbuds, headphones, and other audio devices
• Smart glasses, cameras, recording devices, and wearable technology
• Artificial intelligence (AI) tools, chatbots, and other generative AI technologies
• Notes, textbooks, study guides, formula sheets, or reference materials
• Any other device, material, or resource not expressly permitted by the instructor
Students are responsible for clarifying any questions about permitted materials before the examination begins.

Syllabus Change Policy: The instructor reserves the right to modify course policies, the course calendar, assignment or project point values, and due dates.