Advanced Bioinformatics Lab (BIMM 185)
Tu &Th, 5:00 – 7:50pm, WLH

Instructor:	Luis Arturo Medrano-Soto (Arturo)	**IA:**	Hanbin Lu
Office:	4256 Muir Biology Bldg.	**E-mail:**	hal213@ucsd.edu
Phone:	(858) 534-2457		
E-mail:	l1medranosoto@ucsd.edu		

Course Summary

This course emphasizes the hands-on application of bioinformatics methods to solve biological problems. Students will gain experience in the development and application of software tools, the interaction with public databases, as well as in combining methodologies to answer relevant biological questions.

This is a fourth-year course. Thus, students are expected to understand the fundamentals of molecular genetics, cell biology, statistics, and programming. It is advisable for students taking this course to be familiar with the UNIX terminal. The UNIX shell will be introduced in the first lab session, but prior experience will be helpful because we will use the terminal to run most bioinformatic tools. Since a substantial part of preparing biological data for bioinformatic analyses involves text processing, students will need to become familiar with a high-level programming language that efficiently handles complex regular expressions (e.g., Perl or Python).

Given the nature of this lab, a high level of participation, engagement, and independent research is expected. Therefore it is required that the student:

- Attends classes.
- Completes assignments and submits them on time.
- Demonstrates critical reflection when discussing topics in class.

Learning Objectives

By the end of this course, students will:
- Understand the relevance of bioinformatics and its impact on the life sciences.
- Understand, access, and process data from publicly available biological databases.
- Be able to understand, explain, and apply currently available bioinformatic methods.
- Develop critical analysis and research skills to understand and improve on existing methodologies in bioinformatics.
- Be able to design and implement novel bioinformatic strategies to answer complex biological questions.

Note: Private appointments to discuss accommodations or special circumstances can be arranged by email.

Course Grading

For the first seven weeks, students will be assigned weekly projects and some questions to answer. Students must prepare weekly reports, which are due the following Tuesday at the start of the class. Lab reports will be worth 60% of the final grade. *If all students fill out a course evaluation at the end of the quarter, the lowest scoring weekly lab report will be dropped*. For the last 3 weeks of the quarter, students will work on an independent final project and write a report, which will be worth 25% of the grade. These are the evaluation components:

- Quizzes and participation in class: 15%
- Final project: 25%
- Weekly lab reports: 60%
- <u>Note</u>: no reports will be accepted after the established deadline.

Attendance is mandatory, but in the case of special circumstances exceptions can be made (must be approved by the instructor in advance).

Getting Help

Questions are essential in science and you must never be afraid to ask them. The preferred medium for asking course-related questions is via the TritonEd Discussions. If you are confused, chances are that another student has the same doubt, so by asking questions there and having them answered publicly, everybody will benefit.

Team Work vs Individual Work

In the classroom, while working on coding and analysis, you are encouraged to interact with your classmates, form teams, do research online, and ask the instructor and IA for help. However, when writing your lab reports, work independently, and make sure that the lab report is in your own words and reflects your own understanding.

You may use scientific literature but must properly cite it in the text. DO NOT directly copy/paste code to/from other students in the class or websites. DO NOT post questions about assignments on sites like SEQanswers or Stack Overflow (you can research existing threads on these websites, just don't start new ones for the coursework).

Academic Honesty and Plagiarism

There is a zero tolerance policy regarding academic dishonesty. See the Academic Integrity Agreement (which you must sign to receive a passing grade in the class) for more details on the topic. Violators of these policies may be subject to UCSD rules for academic integrity.

Plagiarism is the unacknowledged presentation of the work of another person as one's own. To present someone else's work as one's own is dishonest and academically worthless. Plagiarism is unethical and will be treated as a serious offense. If a student is uncertain whether a course of action might constitute plagiarism or cheating, they should consult the instructor in advance.

Course Schedule (tentative)

Week 1 (April 4/6):
Setting up your environment, the UNIX shell, working with regular expressions. Generating flat files vs storing your data in a relational database (e.g., MySQL). Accessing, downloading, interacting locally and remotely with biomolecular databases (e.g., GenBank, UniProt, PDB, etc.). BLAST as a tool for querying sequence databases.

Week 2 (April 11/13):
Genomic projects, genome organization, gene prediction, proteins, operons, regulons. Sequence composition (e.g., codon usage, G+C content), relationship between sequence composition, translatability and gene expression.

Week 3 (April 18/20):
Homology and its different flavors. Working definitions of orthology and paralogy. Horizontal gene transfer. Identifying sets of proteins shared between full genomes and identifying proteins that share specific characteristics (e.g., molecular function, biological process, etc.)

Week 4 (April 25/27):
Comparative genomics methods to infer functional relationships between genes, physical protein-protein interactions, etc. Functional classification of proteins and resources available to extract annotations (e.g., KEGG, TCDB, RegulonDB, Gene Ontology, etc.).

Week 5 (May 2/4):
Genome annotation (putting it all together). Genome browsers and integrated analyses (e.g., UCSC, PATRIC, MicrobesOnline). Identification of conserved motifs in DNA and proteins sequences (e.g., active sites, regulatory binding sites, etc.). Inference of transport systems, metabolic pathways, and more in a genome.

Week 6 (May 9/11):
Multiple sequence alignments. Building phylogenies, evolutionary models, rooted vs unrooted trees, species tree vs gene tree. When sequence divergence makes multiple alignments unreliable. Statistical significance of a tree (bootstrapping).

Week 7 (May 16/18):
Gene expression analysis. Normalization. Differential expression. Clustering, classification, correlation and inverse correlation analysis. Co-expression vs co-regulation.

Week 8-10 (May 23 to June 12):
Independent projects.