BIO321G: Big Data In Biology (47960)

Class Tuesdays, Thursdays 11-12:30, GDC 7.514

Website `https://wikis.utexas.edu/display/BDIBstream/Big+Data+In+Biology+Stream`

Personalities

Instructors

Dennis Wylie

Tues 12:30–2
GDC 7.516
denniscwylie@gmail.com

Dennis Wylie focuses on application of statistical and machine learning approaches to the analysis of biological data, especially next-generation sequencing data.

Benni Goetz

Weds 3–4:30
GDC 7.414D
benni@utexas.edu

Benni Goetz is a bioinformatics consultant at CCBB. He writes custom scripts for analysis of next-generation sequencing data. Recently, transcriptome assembly and annotation, bacterial genome annotation, and next-generation sequence databases have been primary focuses.

Dhivya Arasappan

On maternity leave.

Dhivya Arasappan is a CCBB research scientist with 6+ years experience analyzing next gen sequencing data. She is currently enjoying time at home with her new baby, but is looking forward to returning in April to work with the Big Data in Biology stream.

James Derry

jderry@austin.utexas.edu

James Derry is a Senior Systems Administrator for the College of Natural Sciences. James has been teaching Python for Biologist and a microcontroller course for the past three years.

TA

Groves Dixon

Thurs 12:30–2
PAT 427
grovesdixon@gmail.com

Groves is a fourth year graduate student in Mikhail Matz's laboratory (`http://www.bio.utexas.edu/research/matz_lab/matzlab/People.html`) and peer mentor with the CCBB (`http://ccbb.utexas.edu/trainersmentors.html`). He takes evolutionary and functional genomic approaches to explain how coral genomes interact with the changing world around them.

Mentors

You will be part of a small group of roughly five students assigned to each of the mentors. Your assigned mentor is your first point of contact for questions, help, etc. But you can ask other mentors for help too, and you are encouraged to take advantage of the instructors and TA.

Pranav Bhamidipati

pbhamidipati@utexas.edu

In his research in Prof. Hofmann's lab, Pranav Bhamidipati looks at the evolution of developmental gene networks in vertebrates. He like to geek out about the human microbiome, and hopes to engineer probiotics in the future.

Vinita Gottipati

vinitag.95@gmail.com

Vinita Gottipati is a junior biochemistry major in the Bridging Disciplines Program, and is working to get a CSE certificate. She enjoys statistics as well as topics in healthcare, with specific interests in human genetics and public health.

Miranda Grabowski

mirandargrabowski@utexas.edu

Miranda Grabowski is a last-semester senior getting a Bachelor of Science and Arts in Biology with a minor in English. She is interested in computational analysis of next-generation sequencing data.

Sachit Saksena

sachitdsaksena@utexas.edu

Sachit Saksena is a sophomore getting a Bachelor of Science and Arts in Biochemistry and an Elements of Computing certificate. In his research, he has gained experience creating next generation sequencing pipelines for small bacterial genomes using packages like SPAdes, Velvet Genome Assembly and Pacbio SMRT Toolkit. He also loves statistics! He is interested in exploring computational modeling of neural systems in the future.

Principal Investigators

Prof. Hans Hofmann

Mon 1–2
PAT 319
hans@utexas.edu

Dr. Hans Hofmann is a Professor of Integrative Biology (`http://cichlid.biosci.utexas.edu/`) and the Director of the Center for Computational Biology and Bioinformatics (`http://ccbb.utexas.edu/`). Research in his lab focuses on the neural and molecular basis of social behavior and its evolution.

Prof. Vishy Iyer

Office Hours By Appointment
MBB 3.212A
vishy@utexas.edu

Dr. Vishy Iyer is a Professor in the Department of Molecular Biosciences and Institute for Cellular and Molecular Biology, co-Director of the Center for Systems and Synthetic Biology and interim Director of the Genome Sequencing and Analysis Facility. His research interests include the genomics of gene regulation, epigenetics, non-coding RNA, cancer, genetic variation and human disease.

Course Outline

Introduction to Biological Computing
- Broad overview of molecular biology
- Introduction to computing for data analysis

Introduction to Next-Generation Sequencing
- Overview of sequencing technology
- Computational analysis of sequencing data

Discovering Genetic Expression Patterns
- Overview of statistical methods
- Differential expression analysis

Group Projects

Course Flags

Quantitative Reasoning

This course carries the Quantitative Reasoning flag. Quantitative Reasoning courses are designed to equip you with skills that are necessary for understanding the types of quantitative arguments you will regularly encounter in your adult and professional life. You should therefore expect a substantial portion of your grade to come from your use of quantitative skills to analyze real-world problems.

Independent Inquiry

This course carries the Independent Inquiry flag. Independent Inquiry courses are designed to engage you in the process of inquiry over the course of a semester, providing you with the opportunity for independent investigation of a question, problem, or project related to your major. You should therefore expect a substantial portion of your grade to come from the independent investigation and presentation of your own work.

Writing

This course carries the Writing Flag. Writing Flag courses are designed to give students experience with writing in an academic discipline. In this class, you can expect to write regularly during the semester, complete substantial writing projects, and receive feedback from your instructor to help you improve your writing. You will also have the opportunity to revise one or more assignments, and you may be asked to read and discuss your peers work. You should therefore expect a substantial portion of your grade to come from your written work. Writing Flag classes meet the Core Communications objectives of Critical Thinking, Communication, Teamwork, and Personal Responsibility, established by the Texas Higher Education Coordinating Board.

Assignments & Grades

The first roughly two-thirds of the class will focus on developing your knowledge and skill with biology, programming, and statistics as applied to next-generation sequencing technology. The last third of the class will focus on a group final project, analyzing real biological data using the techniques you've learned.

The focus of this course will be to get you started with research problems in new computational aspects of biology. This is not a weed-out class. This is not a prerequisite for later classes. You are expected to learn the material and work hard on your final project, but this is intended to be a collaborative rather than competitive class.

3 problem sets: 210 pts (35%)

Problem sets will be tailored to the section.

3 quizzes: 120 pts (20%)

- One for each of the first three topics.
- Check understanding of basic ideas, not solving difficult problems.

3 papers: 180 pts (30%)

1. Short report on chosen journal article.
2. Grant aims page (grant as in asking for money from an institution).
3. Writeup of project intent and goals, any preliminary results.

Final Presentation: 60 pts (10%)

Present as a group on preliminary results of group project.

Keck seminar participation: 30 pts (5%)

Attend Keck seminar presentations on entrepreneurship.

The Keck Foundation is a major source of funding for our FRI stream. As part of the funding package, the Foundation wants students to receive extra lectures on entrepreneurship and industry. The Keck lectures for all Keck FRI streams will be given by Ana Ward, COO of AptamiR Therapeutics. All lectures will be 6–8pm on their respective evenings.

- Jan 27: What is out there?
- Feb 17: What do you need to get there?
- Mar 9: What is life like at a company?
- Mar 30: How do know if you're succeeding?

Materials & References

Laptops

A laptop will be very useful for this class. If you do not own a laptop, please come talk to Dennis or Benni immediately, and we will work accomodate you. UT Libraries have several computer labs on campus that you can use to connect to our dedicated server for this class.

CCBB Short Courses

The Center for Computational Biology and Bioinformatics offers several 3-hour short courses over the course of the semester. Courses cover a variety of topics in biological computing, ranging from an introduction to the Linux command line, to more advanced topics like ChIP-seq and RAD-seq. For more information see: `http://www.ccbb.utexas.edu/shortcourses.html`

References

These aren't required, and we won't be following them, but they may be useful references to get started. But part of this class is exploring, talking to your fellow students, and finding your own sources!

Python `http://www.greenteapress.com/thinkpython/`

Bioinformatics `https://wikis.utexas.edu/display/bioiteam/Home`

Statistics `http://www.greenteapress.com/thinkstats/`

Canvas

We will use the Canvas system for sending and receiving emails and posting important announcements. Please check the Canvas website regularly for course information. You will submit assignments using Canvas. Canvas will serve as the official repository of the BIO321G scores that will comprise your total BIO321G final grade. It is your responsibility to ensure that your scores for all of the various BIO 321G assignments are correctly displayed in Canvas. Canvas is accessible at: `http://canvas.utexas.edu/`

University Policy On Scholastic Dishonesty

Students who violate University rules on scholastic dishonesty are subject to disciplinary penalties, including the possibility of failure in the course and/or dismissal from the University. At minimum, the incident will be reported to the Deans office and SJS, where a permanent record of the violating incident will be placed in the students file, and a score of zero will be given for the assignment or exam in question. Since such dishonesty harms the individual, all students, and the integrity of the University, policies on scholastic dishonesty will be strictly enforced. You must read and be familiar with the expected standards of conduct for students (`catalog.utexas.edu/general-information/appendices/appendix-c/student-discipline-and-conduct/#subchapter11400.prohibitedconduct`), as well as University policies on scholastic dishonesty (`deanofstudents.utexas.edu/sjs/acint_student.php`).

Academic Accommodations For Students With Disabilities and Other Needs

The University provides, upon request, appropriate academic accommodations for qualified students with disabilities. Any student with a documented disability who requires academic accommodations should contact the Division of Diversity and Community Engagement, Services for Students with Disabilities at 512-471-6259 (voice) or 512- 410-6644 (Videophone) (or see `http://www.utexas.edu/`

diversity/ddce/ssd) as soon as possible to request an official letter outlining authorized accommodations. Accommodations for religious reasons will be made as per University policy, if you notify me at least 14 days prior to your pending absence.

Policy On Religious Observance

The University of Texas at Austin policy is that the student must notify each instructor at least fourteen days prior to the classes scheduled on dates he or she will be absent to observe a religious holy day. For religious holy days during the beginning of the semester, the notice should be given on the first day of the semester. The student may not be penalized for these excused absences but the instructor may appropriately respond if the student fails to notify *prior* to the observance or complete the missed assignment or examination within a reasonable time after the excused absence.

Emergency/Safety Procedures

Familiarize yourself with procedures for emergency evacuation, available at the Office of Campus Safety and Security, 512-471-5767 (www.utexas.edu/safety). You can sign up to receive emergency text alerts at www.utexas.edu/cellphonealert. Occupants of buildings on the UT Austin campus are required to evacuate buildings when a fire alarm is activated. Alarm activation or announcement requires exiting and assembling outside. Familiarize yourself with all exit doors of each classroom and building you may occupy. Remember that the nearest exit door may not be the one you used when entering the building. Students requiring assistance in evacuation must inform their instructor in writing during the first week of class. In the event of an evacuation, follow the instruction of faculty or class instructors. Do not re-enter a building unless given instructions by the following: Austin Fire Department, UT Austin Police Department, or Fire Prevention Services office. Information on emergency evacuation routes and emergency procedures can be found at www.utexas.edu/emergency. The UT Austin Behavior Concerns Advice Line (BCAL) is: 512-232-5050.

Course Schedule

Week	Date	Item
1 - intro comp bio	Jan 19	
	Jan 21	
2 - intro comp bio	Jan 26	
	Jan 27	entrepeneurship module 1 (6-8 PM)
	Jan 28	Hans Hofmann lecture
	Jan 28	problem set 1 assigned
3 - intro comp bio	Feb 02	
	Feb 04	select journal article for report
	Feb 04	problem set 1 due
4 - sequence alignment	Feb 09	
	Feb 11	quiz 1
5 - bioinformatics tools	Feb 16	
	Feb 17	entrepeneurship module 2 (6-8 PM)
	Feb 18	journal article report due
	Feb 18	problem set 2 assigned
6 - intro ngs	Feb 23	begin grant aims page assignment
	Feb 25	Vishy Iyer lecture
	Feb 25	problem set 2 due
7 - pandas, descriptive stats	Mar 01	
	Mar 03	
8 - probability	Mar 08	
	Mar 09	entrepeneurship module 3 (6-8 PM)
	Mar 10	grant aims page due
9 - *SPRING BREAK*		
10 - statistical inference	Mar 22	
	Mar 24	quiz 2
	Mar 24	problem set 3 assigned
11 - differential expression	Mar 29	
	Mar 30	entrepeneurship module 4 (6-8 PM)
	Mar 31	
12 - differential expression	Apr 05	
	Apr 07	problem set 3 due
	Apr 07	define group projects
13 - group projects	Apr 12	
	Apr 14	quiz 3
14 - group projects	Apr 19	
	Apr 21	
15 - group projects	Apr 26	
	Apr 28	individual project report due
16 - group projects	May 03	group presentations
	May 05	group presentations