Welcome to

CMSE 410/890 - Bioinformatics & Computational Biology

Arjun Krishnan

arjun@msu.edu | @compbiologist | thekrishnanlab.org

Week 1: Introductions

Course overview

- Introduction
- Course website
- Communication
- Course activities
- Tentative Schedule
- Course topics
- Wrap-up

Congratulations on surviving 2020!

I'm amazed by and grateful for the seemingly limitless courage, hard work, and sacrifice of healthcare workers, other essential workers, and the people who stood up for social justice.

2020 was an incredibly tough year for being a student, caregiver (esp. parent of young children), or daily-wage worker -- difficulties that were hugely compounded due to also belonging to a systematically minoritized/underrepresented/disadvantaged group.

In addition to these difficulties, my heart goes out to those who also have suffered the loss of loved ones.

Introductions

1. My name is **Arjun Krishnan**.

2. I prefer to be called **Arjun**.

3. My pronouns are **he/him/his**.

4. I'm an **Assistant Professor** in the **Dept. Computational Mathematics, Science, and Engineering** and the **Dept. Biochemistry and Molecular Biology**.

5. Three things to describe me: **1) Endlessly curious, 2) Excited about learning new things, and 3) Love hanging out with smart people like you.**

6. The most underrated part of teaching is learning. I learn a ton every time I teach a course. I design courses that help me learn. And you students point out my blind spots. It's such a joy!

Learn more about your fellow learners at http://flipgrid.com/compbio2021.

What you should get out of this course

How to become a practicing computational biologist in 60+h?

- Introduction to the inner-workings of methods in bioinformatics and computational biology:
  - Analytical techniques, algorithms, and statistical/machine-learning approaches developed to address key questions in biology and medicine.

- At the end you should be able to:
  - Critically read bioinformatics / computational-biology literature.
  - Apply the methods you have learned to problems both within & outside biology.

Modules & Topics

1. Sequence alignment & pattern finding	• Dynamic programming; Substitution matrices; Fast Local Alignment
2. Genome assembly & annotation	• de Bruijin graphs; Hidden Markov models
3. Comparative genomics; Phylogenomics	• Suffix trees; Tree construction
4. Quantitative genetics	• Regularized linear regression; Statistical inference, Multiple testing
5. Regulatory genomics	• Gibbs sampling; Expectation-Maximization
6. Functional genomics	• Two-sample tests; Hypergeometric test; Unsupervised/supervised learning
7. Single-cell genomics	• Missing value imputation; Dimensionality reduction; Trajectory inference
8. Molecular dynamics; Structure prediction	• Molecular simulation; Maximum entropy modeling
9. Biological networks	• Measuring associations; Network inference; Graph theory, Label propagation
10. Modeling cellular pathways	• Wiring diagrams; Dynamical simulation, State Space, Bifurcation
11. Whole-cell models; Digital evolution	• Constraint-based modeling; Artificial life; Whole-cell models

Modules & Topics

1. Sequence alignment & pattern finding
2. Genome assembly & annotation
3. Comparative genomics; Phylogenomics
4. Quantitative genetics
5. Regulatory genomics
6. Functional genomics
7. Single-cell genomics
8. Molecular dynamics; Structure prediction
9. Biological networks
10. Modeling cellular pathways
11. Whole-cell models; Digital evolution

- Dynamic programming; Substitution matrices; Fast Local Alignment
- de Bruijin graphs; Hidden Markov models
- Suffix trees; Tree constr
- Regularized linear regres
- Gibbs sampling; Expecta
- Two-sample tests; Hype
- Missing value imputatio
- Molecular simulation; Ma
- Measuring associations;
- Wiring diagrams; Dynam
- Constraint-based model

MW & first-half of F: Arjun presents

Each week:

- Lecture
- Paper discussion

Second-half of F: You present/lead as team of two.

Prerequisites & Expectations

- CMSE 201 and two semesters of introductory biology (LB 144 and 145 OR BS 161 and 162 OR BS 181H and 182H, or equivalent).
  - Statistics at the level of STT 231 is strongly recommended.

- Basically, it would be assumed that you:
  - know how to code in one of the mainstream languages like Python or R,
  - have an understanding of basic statistics and probability, and
  - have studied basic genetics, molecular biology, and cellular biology.

Course website

bit.ly/compbio2021

- Contact information
- Course outline and materials
- Schedule, location, calendar, and office hours
- Website and communication
- Course activities
- Grading information
- Attendance, conduct, honesty, and accommodations

- Lecture slides
- Learning materials
- Assignments

Communication

compbio2021.slack.com

- The primary mode of communication in this course (including major announcements) will be the course Slack account.

#announcements	#talks-meetings-conferences
#lectures-assignments	#fun
#project	#health-workout
#primers-articles	#random

My office: 2507H Engineering Building (2nd floor)

Course activities

- Assignments
- Class presence
- Project

Assignments

- For each topic, you will be given an assignment after the topic's first "Lecture" class on **Monday**:
  - A paper to read, summarize, and critique.
  - A software tool to explore.
  - An exercise to work on.
- Submit a report *before* the topic's "Paper discussion" class on **Friday**.
- Details on each assignment will be posted on Slack and on the class website.

Class presence

If you sincerely engage with the class and put in earnest effort, you will do well. Your can do so by:

- Completing any pre-class readings.
- Showing up to the online class on time and staying for the whole hour.
- Responding to prompts from Arjun that punctuate the lectures: Qs, polls, etc.
- Asking Qs about computational or biological concepts (because no one will have the perfect background).
- Working in groups during in-class discussion sessions.
- Correcting me when I am wrong.

Learning groups

A group of learners focused on a common goal and who respect & support each other in pursuit of that goal.

- Each group has a **dedicated slack channel**.
  - Post questions/comments; Help each other when needed; Coordinate meetings.
- **Meet** on Thursday or Friday (before class) to study/discuss that week's lectures & assignment.
- As a group:
  - Maintain a living document containing a **glossary** of terms & concepts [shared w/ me].
  - Bring at least **one question** to the class on Friday about topics/concepts [I will call out each group].
- Work on your assignments beforehand; Discuss w/ the group; **Submit individual assignments**.

Learning groups

Best practices:

- All class policies on **Presence, Conduct, & Honesty** apply to your shared space: https://github.com/krishnanlab/teaching/blob/master/2021-spring_compbio/policies.md.

- **Learn about your groupmates**: watch their FlipGrid profile & read their Research profile.
    - Learn their names & pronouns. [Please make your zoom/slack name to the name you go by.]

- **Prepare before your weekly meeting**:
    - Attempt the assignment on your own once before discussing it when you meet.

Friday paper presentations

- One group will present a paper for 15–20 min.
- Another group will lead a 5–10 min discussion on key Qs based on the paper / presentation.

Class Day	Date	Module	Presenting group	Discussion group
Day 05	F, Jan 29	Sequence Alignment	Group 1	Group 4
Day 08	F, Feb 05	Genome Assembly and Annotation	Group 2	Group 5
Day 11	F, Feb 12	Comparative Genomics; Phylogenetics	Group 3	Group 1
Day 14	F, Feb 19	Quantitative Genetics	Group 4	Group 2
Day 17	F, Feb 26	Regulatory Genomics	Group 5	Group 3
Day 26	F, Mar 19	Functional Genomics	Group 1	Group 4
Day 29	F, Mar 26	Single-cell Genomics	Group 2	Group 5
Day 32	F, Apr 02	Protein Structure Prediction and Dynamics	Group 3	Group 1
Day 35	F, Apr 09	Biological Networks	Group 4	Group 2
Day 38	F, Apr 16	Modeling Cellular Pathways	Group 5	Group 3

Friday paper presentations

- Each group will present twice & lead the discussion twice. Divide yourself into 2 subgroups to do the two presentations/discussions.
- If a group is presenting a particular week, they are not required to turn in the assignment for that week.

Class Day	Date	Module	Presenting group	Discussion group
Day 05	F, Jan 29	Sequence Alignment	Group 1	Group 4
Day 08	F, Feb 05	Genome Assembly and Annotation	Group 2	Group 5
Day 11	F, Feb 12	Comparative Genomics; Phylogenetics	Group 3	Group 1
Day 14	F, Feb 19	Quantitative Genetics	Group 4	Group 2
Day 17	F, Feb 26	Regulatory Genomics	Group 5	Group 3
Day 26	F, Mar 19	Functional Genomics	Group 1	Group 4
Day 29	F, Mar 26	Single-cell Genomics	Group 2	Group 5
Day 32	F, Apr 02	Protein Structure Prediction and Dynamics	Group 3	Group 1
Day 35	F, Apr 09	Biological Networks	Group 4	Group 2
Day 38	F, Apr 16	Modeling Cellular Pathways	Group 5	Group 3

Project

- A major goal of this course is to prepare your ability to perform original research in computational biology, and to present your ideas and research.
- Can be one of:
  - Design and implement a new computational method for a task in biology.
  - Improve an existing method.
  - Perform an evaluation of several existing methods.
  - Develop a fully-reproducible codebase for an existing analysis in a paper.

Project

1. Research profile | Fri, Jan 22
2. Project topic | Fri, Jan 29
3. Project pre-proposal | Fri, Feb 5
4. Project proposal | Fri, Feb 19
5. Proposal reviews | Fri, Feb 26
6. Mid-term project presentation recordings | Wed, Mar 10
7. Mid-course Project presentation reviews | Fri, Mar 12
8. Mid-course project report | Fri, Mar 19
9. Final project report | Wed, Apr 21
10. Final project poster presentations | Note: Diff. day/time | Thr, Apr 29, 12:45 – 2:45p

Groups @ MSU doing (a sigf. amount of) Bioinfo & Compbio

- Jianrong Wang
- Yuying Xie
- Alex Dickson
- Michael Feig
- George Mias
- Ana Vasquez
- Gustavo de los Campos
- Wen Huang
- Eran Andrechek
- Robin Buell
- Erik Goodman
- Kevin Liu
- Mark Reimers
- Sudin Bhattacharya
- Jiayu Zhou
- Shinhan Shiu
- MANY OTHERS

The **MSU_Bioinfo-Compbio-Investigators_2021.md** file in the class website contains a full(er) list.

You are welcome to work with me as well!

Questions/comments

Website

- It is set up nicely.
- The description and information in the course webpage is really extensive and detailed.
- The site is very well organized!
- It is very organized.
- Very detailed write up.

Questions/comments

Website

- It is very well laid out, however whenever I check D2L I slightly panic because I don't see CMSE411 show up. But I think I'll get used to that over the semester🙂
- I like how it is organized. Preferable to D2L.
- Will we also have a D2l page?
- Will there be assignments/important information posted on Slack that is unavailable on the course website and D2L?

Questions/comments

Website

- Will lectures and other material be posted early?
- How often will the website be updated?
- Will the lectures be updated weekly or by the day?
- Where do messages and notifications go?
- When are we assigned our "learning groups"?
- How to get to the shared google doc?
- Will the class syllabus, lectures, and other materials be available as pdfs?

Questions/comments

Slack

- Why do we have so many channels? Some of the channel names confused me.
- How often do students use it?
- Are the different channels (health-workout, fun, etc.) for things we are doing in this class? Or just general extracurricular communication?
- To what extent is student collaboration allowed in the slack?

Questions/comments

Slack

- I think it's a really good idea to have health and fun channels.
- It is a good place to ask questions.
- I think it's cool that I have a class that actually has an actual virtual environment for students to talk.
- Well organized. I like the health-workout channel idea :)
- I appreciate the Health/fun sections!
- Looking forward to utilizing the #health channel

Questions/comments

Slack

- I think it will help facilitate communication within the classroom
- Hopefully it's put to good use.

Questions/comments

Misc.

- Is it recommended to also learn R to prepare for this class, or is python knowledge sufficient to succeed?
- Where do we post our research profile?

- I'm a very excited about this class in Zoomland. It looks like a very challenging course!
- I'm looking forward to it!