

## 03-511/711 Computational Molecular Biology and Genomics Fall 2021

### Instructors:

Professor Dannie Durand, ([durand@cmu.edu](mailto:durand@cmu.edu))

Dr. Maureen Stolzer ([mstolzer@andrew.cmu.edu](mailto:mstolzer@andrew.cmu.edu))

Yuting Xiao ([yutingx1@andrew.cmu.edu](mailto:yutingx1@andrew.cmu.edu))

Course web site: <https://www.cs.cmu.edu/~durand/03-711/>

Teaching and learning this semester will (hopefully) be more normal, but it will still be an adventure. Class is planned to be in-person, but we should all be prepared for changes due to Covid. For this reason, the plan is tentative and subject to change.

If you have any questions about the policies explained here, or any other aspect of the course, please do not hesitate to ask. I am here to help you learn this material and succeed. My course policies are designed to explain my expectations of students and foster an open learning environment, while ensuring a fair and honest system.

### Course description

An advanced introduction to computational molecular biology, using an applied algorithms approach. The course will survey established algorithmic methods, including pairwise sequence alignment and dynamic programming, multiple sequence alignment, sequence evolution models, derivation of amino acid substitution matrices, hidden Markov models for molecular motifs, database search heuristics, and local alignment statistics. We will explore emerging computational problems in genomics through assignments in 03-711.

Here is a tentative schedule of topics covered. The course web site provides a detailed view of course materials: <https://www.cs.cmu.edu/~durand/03-711/>.

1. Introduction to course
2. Pair-wise sequence alignment
3. Sequence evolution models
4. Amino acid substitution matrices
5. Modeling motifs, PSSMs, Hidden Markov Models
6. Database searching, Blast, alignment statistics
7. Multiple Sequence alignment

### Objectives

Upon completion of this course, students

1. will have a working knowledge of algorithms for global, semi-global, and local alignment and be able apply those algorithms to concrete examples;
2. understand how parameter selection in a scoring function influences the results obtained using an alignment algorithm;
3. have an understanding of Markov models of sequence evolution;

4. be familiar with the problems of sequence motif discovery, representation and recognition in a probabilistic framework;
5. understand the application of Hidden Markov Models in this framework, including the Viterbi, Forward, and Backward algorithms;
6. understand the Baum Welch algorithm and the Gibbs Sampler in terms of sequence analysis;
7. understand the purpose of amino acid substitution matrices that are parameterized by evolutionary distance;
8. have an in-depth knowledge of the BLAST database search heuristic and its parameters; be able to select appropriate parameter values for a given query sequence and retrieval goal; and be able to interpret the statistical output of BLAST.

### **Prerequisites**

Research in bioinformatics requires an interdisciplinary effort between molecular biologists and computer scientists, and students from both disciplines are encouraged to participate in the course. A variety of backgrounds will bring different strengths to the course.

To succeed in this course, you will need a solid introductory background in biology (molecular, cell and genetics), data structures and algorithms, and discrete math (probability, graphs, trees). There are *no programming assignments* in this course and the ability to program is not evidence that you have the right foundation.

To take this course you must have taken the following CMU courses or have the permission of the instructor.

- Modern Biology (03-121),
- Programming for Scientists (02-601) or Principles of Imperative Computation (15-122),
- Concepts of Mathematics (21-127). Note that 21-127 is a co-requisite of 15-122.

### **Course organization**

Lecture topics, reading assignments, lecture notes, assignments, work sheets and ***due dates*** will be posted the course website: <https://www.cs.cmu.edu/~durand/03-711/>

Canvas will be used for you to submit assignments and for us to return graded assignments to you. Canvas will not be the primary vehicle for communicating with you about the progression of the course. *Do not send me messages on Canvas, they will not be read.*

Piazza will be used as a class bulletin board and communication system. Any questions you have about course organization or questions about the material should be posted to Piazza. Matters of a private nature should be sent to the instructors by email or via face-to-face communication (in-person or virtual). If you have a question about the course or content, at least two other students will have the same question and will be delighted that you asked. The question deserves to be on Piazza.

Office hours will be held via zoom. We will have a poll to determine the best times for weekly office hours.

## **Covid-19**

In order to attend class meetings in person, all students are expected to abide by all behaviors indicated in [A Tartan's Responsibility](#), including any timely updates based on the current conditions.

In terms of specific classroom expectations, whenever the requirement to wear a facial covering is in effect on campus, students are expected to wear a facial covering throughout class. *Note: the requirement to wear a facial covering is in effect for the start of the Fall 2021 semester.* If you do not wear a facial covering to class, I will ask you to put one on. (If you don't have one with you, I will direct you to a distribution location on campus, see <https://www.cmu.edu/coronavirus/health-and-wellness/facial-covering.html>). If you do not comply, you will be referred to the Office of Community Standards and Integrity for follow up, which could include student conduct action. Finally, please note that sanitizing wipes should be available in our classroom for those who wish to use them.

## **Reading**

You do not need to purchase a textbook for this course. Reading materials drawn from several textbooks, as well as the primary literature, are available through course web site. In addition, a comprehensive set of lecture notes (~150 pages) in pdf format will be provided to you electronically.

For supplementary reading, the following textbooks are recommended:

- Biological Sequence Analysis: Probabilistic Models of Proteins and Nucleic Acids, R. Durbin et al ([electronic version](#)).
- Introduction to Computational Molecular Biology, J.C. Setubal and J. Meidanis
- Statistical Methods in Bioinformatics: An Introduction Ewens and Grant ([electronic version](#)).

*The Cartoon Guide to Genetics*, by Larry Gonick and Mark Wheelis, is recommended for those needing additional biological background. The paperback edition is available on Amazon for about \$15, new; used copies start below \$10.

## **Course work**

Course work for all sections includes 7 to 9 homework assignments, one mid-term exam and a final exam. Homework assignments focus on problem solving and *do not require programming*. Students taking the 12-unit version of the course (711) will complete 6 to 8 additional assignments that extend the course material to topics in genomics and involve problem solving in more complex or open-ended scenarios.

The final exam is cumulative, although there will be greater emphasis on material covered in the second half of the course.

## **Grading policy:**

- Participation: 5%
- Homework: 40%
- Midterm Exam: 20%
- Final: 35%

## Participation

Participation makes class more valuable for everyone. During class, I expect students to be attentive and actively engaged in the learning process. Asking questions will help you, me, and the rest of the class. Chances are, if something is confusing to you, it's confusing to at least one other person in class. I will ask questions of the class and ask for participation in walk-through examples. This helps me assess whether a concept has been effectively conveyed or areas that need more discussion. I understand that public speaking can be more difficult for some than others. Therefore, I will also consider your attendance at office hours and questions submitted electronically as part of your participation.

The participation requirement can be satisfied through interaction in-class or via Piazza. In any given week, participation could include

- Posting a substantive question on the course bulletin board on Piazza. ("When is the homework due?" is not a substantive question.)
- Providing a substantive answer to a question posted by someone else on the course bulletin board. ("Friday" is not a substantive answer.)
- Attending class and asking or answering a question or participating in the discussion.

You do not have to participate in the same way each week.

## Problem Sets

Canvas will be used to distribute assignments, for handing in completed assignments, and for returning graded assignments with comments. To facilitate this process, assignments will be made available in both pdf and latex formats. You may submit your assignments in any of the following ways:

- Download the assignment in pdf format and print it out. Write your solutions in the space provided. Scan your handwritten solution and upload it to Canvas.
- Download the assignment in pdf format. Use the commenting features in adobe or a similar tool to enter your solution directly on the pdf. Upload the completed assignment, annotated with your solution.
- Download the assignment in latex format. Enter your solution in latex format, compile the assignment, and turn in the resulting pdf.

For option 1 and 2, make sure that your answers are clear to read.

Due dates Homework will typically be assigned on a weekly basis, with the exception of exam weeks, and will be due 7 to 10 days later. Please turn your assignments in on time. Late homework inconveniences your instructors and other students in the course. Late homework also puts you at a disadvantage because problem sets are designed to help you stay up to date on the material.

Late homework will not be accepted if doing so would disrupt the administration of the course. In particular, late assignments will not be accepted once graded homework has been returned to the class and solutions have posted. If late homework is accepted, the penalty is 10% per day. Extensions will be granted only in extraordinary circumstances (e.g., an extended illness). *There will be no "makeup" homework assignments.* Please let me know as soon as possible, preferably through email, if you will require an extension.

To offset these draconian rules and give you some flexibility in these uncertain times, I will drop your two lowest homework assignment grades when calculating your final score, *with the proviso* that all assignments have been submitted by the last day of classes. Your mid-term grade will be based on in-class examinations and all homework assignments administered before mid-semester. No homework scores will be dropped at midterms.

Collaboration on Homework Unless stated otherwise, you may use textbooks, your own class notes from previous courses, and material you find on the internet, in preparing homework assignments. Include a list of the reference materials you used when you hand in your assignment. You may not consult solution sets from previous years. Nor may you quote from any source without attribution.

*Problem Set 0* (PS0) is a self-administered placement quiz, to help you (and me) determine if you have the foundation required for the course. On PS0, you may not

- collaborate with other students,
- refer to previous homework assignments and/or solution sets from the core courses from prior years, or
- discuss this assignment with anyone other than Dr. Durand or Dr. Stolzer.

Your score on Problem Set 0 counts towards your final grade.

*All other problem sets:* Discussion and collaboration on homework problems between students is allowed, but you may not hand in jointly prepared assignments. Students may not copy any portion of a homework assignment from another student, nor may they jointly prepare all or part of an assignment. Students who collaborate on homework assignments must write the names of their collaborators on the front of their homework assignments, with a short explanation of what was discussed. This information is to help me understand each individual student's comprehension of the material.

An example of acceptable collaboration would be the discussion of strategies for a particular task, followed by each student implementing the strategy independently. Examples of unacceptable collaboration are:

- jointly doing an analysis and then handing in multiple copies of the results;
- following a method suggested by someone else without being able to explain the method;
- getting help from a classmate or source without acknowledgement; or
- receiving answers from students who have taken the course in previous years.

### **Academic Integrity**

Integrity and honesty are integral aspects of good science.

**All students are responsible for familiarizing themselves with the Carnegie Mellon [policies on Academic Integrity, Cheating and Plagiarism](#) and adhering to those policies.** In order to deter and detect plagiarism, online tools and other resources may be used in this class.

You should be aware that plagiarism, cheating and other violations of the academic integrity policy are considered to be serious [infractions](#) at Carnegie Mellon. Any act of cheating or plagiarism will be treated in accordance with Carnegie Mellon's Policy on Academic Integrity. Depending upon the individual

violation, students could face penalties ranging from failing the assignment to a penalty as severe as failure in the course for a single violation.

### **Lectures**

Most lectures will be presented at the chalkboard. Lecture summaries will be posted on the syllabus page. These are not comprehensive notes and are not guaranteed to cover all the material discussed in class. **You are responsible for taking notes in class and learning the material from your notes and reading assignments.** Occasionally, lectures will include PowerPoint presentations, software demonstrations, and guest lectures.

### **Electronic Devices**

Unexpected noises and movement can automatically divert and capture your classmates' attention, which means you are affecting everyone's learning experience if your device makes noise or is visually distracting during class. Personally, I find laptops and other electronic devices in class distracting. When I am distracted, I am a poor lecturer. Further, [evidence](#) from learning research suggests that laptops in the classroom do not enhance the academic performance of the student using the laptop - taking notes by hand helps you retain information better!

For this reason, I ask you to turn off your mobile devices and close your laptops during class. Exceptions will be made for students with disabilities.

*Classroom Recording Allowed for Personal Educational Use:* Classroom activities may be recorded by the lecturer or a student for the personal, educational use of that student or for all students presently enrolled in the class only, and may not be further copied, distributed, published, or otherwise used for any other purpose without the express written consent of Dr. Durand. Recordings of class sessions are covered under the **Family Educational Rights and Privacy Act (FERPA)** and must *not* be shared with anyone outside your course-section. All students are advised that classroom activities may be taped.

### **Accommodations for Students with Disabilities**

If you have a disability and have an accommodations letter from the Disability Resources office, I encourage you to discuss your accommodations and needs with me as early in the semester as possible. I will work with you to ensure that accommodations are provided as appropriate. If you suspect that you may have a disability and would benefit from accommodations but are not yet registered with the Office of Disability Resources, I encourage you to contact them at [access@andrew.cmu.edu](mailto:access@andrew.cmu.edu).

### **Attendance**

Attending class is one of the best ways to gain a better understanding of the material. Being attentive for lectures, taking notes, and engaging in examples are important to your success in this course. In order to learn, you must be present both physically and mentally!

*Tardiness:* Being late to class disrupts the lecturer and the learning environment of your fellow classmates. Please respect your peers and show up to class on time. If there are extenuating circumstances, please contact me.

## **Study Tips**

These are some tips and tricks on how to do well in class. (These are general tips that can help you succeed in learning!)

1. Come to class prepared with paper and writing implement for note taking.
2. Lecture notes and/or slides are posted to the syllabus page on the course website. Read over these notes and your class notes before the next lecture. This will only take a short amount of time, but can yield great benefits in transferring material to your memory and highlighting areas of confusion, which can be clarified in class.
  - Highlight important concepts and/or make marginal notes.
  - Re-do in-class examples, without looking at your answer. When going through examples in class, it may seem easy. Doing the example yourself can highlight areas that need clarification.
  - Write down your questions and bring them to the next lecture. If you're uncomfortable asking a question in class, give it to me on a piece of paper at the start of class or email me the night before.
3. Do the homework. Doing the problems is like training for your brain.
  - Start early.
  - Ask me for clarification on any questions that are unclear.
  - Attempt to solve the problems before collaborating with other students or coming to office hours.
  - Do the problems again, after the solutions have been posted. Be sure you understand how to solve the problem.
4. Come to office hours and use the time productively.
5. Read the assigned readings. As with lecture notes, you should highlight, make marginal notes, do the examples, and write down questions you may have.

## **Student Health & Well-being**

Take care of yourself. Do your best to maintain a healthy lifestyle this semester by eating well, exercising, avoiding drugs and alcohol, getting enough sleep and taking some time to relax. ***This is especially important given the uncertainty and isolation many of us face during the pandemic.***

All of us benefit from support during times of stress or struggle. You are not alone. There are many helpful resources available on campus and an important part of the college experience is learning how to ask for help. Take the time to learn about all that's available and take advantage of it. Ask for support sooner rather than later – this always helps.

If you or anyone you know experiences any academic stress, difficult life events, or feelings like anxiety or depression, we strongly encourage you to seek support. Consider reaching out to a friend, faculty or family member you trust for assistance in obtaining support that can help. Counseling and Psychological Services (CaPS) is here for you: call 412-268-2922 and visit their website <http://www.cmu.edu/counseling/>. Over 25% of students reach out to CaPS at some point during their time at CMU.

If you or someone you know is feeling suicidal or in danger of self-harm, call someone immediately, day or night:

CaPS: 412-268-2922

Re:solve Crisis Network: 888-796-8226

If the situation is life threatening, call the police

On campus: CMU Police: 412-268-2323

Off campus: 911

*Food Insecurity:* If you are worried about affording food or feeling insecure about food, there are resources on campus who can help. Email the CMU Food Pantry Coordinator to schedule an appointment:

Pantry Coordinator

[cmu-pantry@andrew.cmu.edu](mailto:cmu-pantry@andrew.cmu.edu)

412-268-8704 (SLICE office)

### **Academic Resources**

- The [Student Academic Success Center](#) offers many services and resources to students in need of academic help, including: academic counseling and individual tutoring. They can also help you improve general learning strategies, such as note-taking, time-management, exam preparation, and stress management. Services are free to all Carnegie Mellon students.
- [Communication Support](#) provides free communication instruction and support to help CMU students, faculty, and staff convey their ideas logically, clearly, and effectively. Services include one-on-one tutoring and campus workshops.
- [Language and Cross-cultural Support](#) offers services and resources to students from other cultural backgrounds to help with developing language skills and adjusting to the culture of the American classroom.

### **Diversity and Inclusion**

*We must treat every individual with respect. We are diverse in many ways, and this diversity is fundamental to building and maintaining an equitable and inclusive campus community. Diversity can refer to multiple ways that we identify ourselves, including but not limited to race, color, national origin, language, sex, disability, age, sexual orientation, gender identity, religion, creed, ancestry, belief, veteran status, or genetic information. Each of these diverse identities, along with many others not mentioned here, shape the perspectives our students, faculty, and staff bring to our campus. We, at CMU, will work to promote diversity, equity, and inclusion not only because diversity fuels excellence and innovation, but because we want to pursue justice. We acknowledge our imperfections while we also fully commit to the work, inside and outside of our classrooms, of building and sustaining a campus community that increasingly embraces these core values.*

*Each of us is responsible for creating a safer, more inclusive environment.*

Unfortunately, incidents of bias or discrimination do occur, whether intentional or unintentional. They contribute to creating an unwelcoming environment for individuals and groups at the university. Therefore, the university encourages anyone who experiences or observes unfair or hostile treatment



on the basis of identity to speak out for justice and support, within the moment of the incident or after the incident has passed. Anyone can share these experiences using the following resources:

- Center for Student Diversity and Inclusion: [csti@andrew.cmu.edu](mailto:csti@andrew.cmu.edu), (412) 268-2150
- [Report-It](https://reportit.net) online anonymous reporting platform: [reportit.net](https://reportit.net) username: *tartans* password: *plaid*

All reports will be documented and deliberated to determine if there should be any following actions. Regardless of incident type, the university will use all shared experiences to transform our campus climate to be more equitable and just.