CPSC 567 / IB 505 :: BIOINFORMATICS AND SYSTEMS BIOLOGY

Course Information

Instructor: Gustavo Caetano-Anollés

Office: Room 332, National Soybean Research Center

Email: gca@illinois.edu; Phone: (217) 333-8172

Office hours: M and W 12 AM to 1 PM (by appointment)

Schedule: 4 h; Monday and Wednesday 10:00 – 12:00 AM CST.

Location: Real-time online via Zoom

Web Course Tools (*Illinois Compass*): http://compass.illinois.edu

Web page login and password: *net id* and password, respectively

Course web address: http://gca.cropsci.illinois.edu/courses.html

Objectives

Systems biology represents a novel view of the biological sciences in which the emergent properties of a biological system are explained by its components and their interrelationship. This view is possible thanks to advances in genomics, bioinformatics, and structural biology, and the generation of large amounts of complex data linked to genotype and biological function. *Bioinformatics and Systems Biology* introduces students to the crossroads of bioinformatics and systems approaches. The course combines lectures, selected hands-on exercises, and discussion of topics that are relevant to the emergent field.

Outline of Topics

1. JAN 19: Introduction to bioinformatics and systems biology
2. JAN 24: Phylogenomics
3. JAN 26: **Phylogenomics: presentations and discussion**
4. JAN 31: **Phylogenomics: presentations and discussion**
5. FEB 02: **Phylogenomics: presentations and discussion**
6. FEB 07: Basic tools for phylogenetic reconstruction
7. FEB 09: Phylogenetic reconstruction: experimental
8. FEB 14: Phylogenetic reconstruction: experimental
9. FEB 16: Structural bioinformatics: modeling and visualization
10. FEB 21: Structural bioinformatics: classification
11. FEB 23: **Structural bioinformatics: presentations and discussion**
12. FEB 28: **Structural bioinformatics: presentations and discussion**
13. MAR 02: **Structural bioinformatics: presentations and discussion**
14. MAR 07: Structural bioinformatics: experimental (HMWK 1)
15. MAR 09: *Project planning*
16. MAR 21: Networks in biology
17. MAR 23: **Networks: presentations and discussion**
18. MAR 28: **Networks: presentations and discussion**
19. MAR 30: **Networks: presentations and discussion**
20. APR 04: *Project discussion*
21. APR 06: Network visualization: experimental
22. APR 11: *Project: Presentation of project proposals*
23. APR 13: Systems biology: simulation and system-level analysis
24. APR 18: **Systems biology: presentations and discussion**
25. APR 20: **Systems biology: presentations and discussion**
26. APR 25: **Systems biology: presentations and discussion**
27. MAY 27: *Project think tank*
28. MAY 02: *Project presentations*
29. MAY 04: *Summary* (HMWK 2)

Some basic questions addressed in this course:

General
What is bioinformatics?
What is systems biology?
What are the challenges of biology in the post-genomic era?
Phylogenomics
How do genomes evolve?
Are processes of change gradual or punctuated?
Are there analogous processes of change at different scale in the genome?
Structural bioinformatics
How can we recognize patterns in structure?
Can we model and predict protein structure?
Can we define a protein universe?
What are current protein classification schemes?
Networks
What kinds of networks exist?
How can we visualize and describe network architecture?
How do networks evolve?
How can we analyze the dynamics of networks?
Systems biology
What are the properties required for system-level understanding?
Can we integrate different network architectures?
How can we model systems?
What are emergent properties of systems?
What is the role of modularity and redundancy in systems?

Grading and class activities

Grading will be based on seminar presentations, class participation, and homework assignments, totaling 100 points.

Seminar presentations (50 points): During the course there will be 12 classes (depending on the number of registered students) devoted to student presentations. These classes cover four central themes (see outline):

Phylogenomics
Structural bioinformatics
Networks
Systems biology

During these events, each student will have in two (2) occasions the opportunity to present a 25 minute-seminar presentation, engage actively in discussion, and receive 25 points for each of presentations. The subject of the presentations will be assigned based on preference during the first day of class, and will be associated usually with one or two individual research papers that puts forth a point of view. All research articles will be assigned in advance and will be posted in *Illinois Compass* for downloading. It is the responsibility of *all* students to read these articles before coming to class. Presentations will be followed by a general discussion of the subject, where each speaker will have to act as advocate and defend the point of view of their contributions. At the end of the class, students will summarize the theme. At that time, the students that were not speakers that day will privately rate the quality of speaker presentations and will submit their rating together with their next homework assignment (see below).

When preparing your presentation, seek other articles from the literature.

- In particular, if a background article is a review, you will want to read some papers it cites in order to present specific illustrations of its points.
- With the *Web of Science* you can find papers that cite an earlier paper to which you have a reference.
- Or, you can search by keywords in *PubMed*.
- Use the *Web of Science* or *PubMed* through the U of I Library's Online Research Resources (http://www.library.uiuc.edu/orr/), which gives links directly to PDF files for the papers. Ask the reference librarians for help, if you need it.

As you prepare your seminar, synthesize a coherent overview of your topic. Please feel free to consult with me, or your classmates, as you prepare. Think about, and then tell the students:

- What are the questions you will address?
- What answer will you offer for each question?
- What is the evidence to support each answer?

Your last slide should contain ***three (3) thought-provoking questions*** to jumpstart discussion in class.

When you present a seminar you should plan to deliver your topic in about 20 minutes. We will adopt this format and adhere to time limitations strictly. It is recommended that you use PowerPoint, PDF or a text file for your presentation. Usually, a 20-minute presentation involves up to 10-15 slides (or graphical items), but this depends on content and fluidity of delivery. You may connect your own laptop computer to the classroom computer. Note that if your laptop is a Mac, *you* need to bring an adapter to the connector for a PC, which is on the classroom computer. Your presentation should be made available to the rest of the class at least a day in advance. It should be sent directly to the instructor by e-mail so that the file can be placed in *Iliinois Compass* for downloading by the class.

The grading of seminar presentations is based on the merits of the effort invested, the presentation (materials, analysis, critical thinking, etc), and your participation. Your grade will be based on how well synthesized and presented was the subject. Fellow student will also rate your performance in relation to that of other speakers, and this rating will be taken into consideration.

Homework assignments linked to presentation topics (25 points): Two homework (hwk) assignments due the week immediately after the first two series of seminar presentations and close to the end of the course will provide you with the opportunity to earn 12.5 points each. The assignment involves the production of a review paper that summarizes content and discussion of the two corresponding seminar series. You will need to review the literature associated with the presentations, include items of discussion and conclusions from class, and produce a well-rounded paper. This task will be facilitated by how much effort was invested by individual speakers. At the end, you will rate the overall performance of speakers, not necessarily individually. The paper should be no more than five (5) pages single-spaced, in a font no less than 12 point, and should include a title, an overview of the presentations, a statement about the problem(s) and their significance, a brief indication of the evidence for each point in the overview, and a list of references that you have used. The homework deadlines are the following:

MAR 07 Homework 1: Phylogenomics and structural bioinformatics
MAY 04 Homework 2: Network and systems biology

All homework assignments must be electronically submitted using *Illinois Compass* by the respective deadlines. Electronic mail attachments or printed homework submissions will not be accepted. The homework file must be named with the *netid* of the student followed by the homework number (e.g., gca3). Only one homework assignment submitted up to 5 weekdays after the deadline will be accepted, but 20% of points will be discounted. These exceptional late submissions must be previously arranged with the instructor.

Bioinformatic project in molecular evolution, biological networks and protein structure (25 points): Group of students will work together on a project throughout the course. The groups will report on their project on the last day of class, and will provide a written report in the form of a research paper. Each group team member will earn up to 25 additional points.

Your group will need to draft a research project that is FEASIBLE within resources and time available. During the course, for example, information about resources listed below will be provided and will be available for your perusal in the project:

1. Given proteins with structural information (i.e., with corresponding PDB entries), CASTp provides structural information about pockets, DALI and GANGSTA+ allows comparison of 3D overlaps between proteins, and SCOP provides a structural classification scheme defining folds, fold superfamilies and families of protein architecture. Visualization tools such as CHIMERA allow analysis of molecules and complexes, and PDBsum summarizes structural information.

2. KEGG contains metabolic network information in the form of pathways of metabolites and enzymes, annotated with genomic and biochemical information. MANET traces evolutionary information related to the occurrence of fold architectures in genomes along pathways of enzymes in KEGG. This information is given as an ancestry value (ranging 0 to 1 from old to young), which is linked to SCOP classifiers and enzymatic function (given by EC numbers) in every subnetwork of the 11 mesonetworks of metabolism. PROCOGNATE provides information about cofactors.

3. Phylogenetic reconstructions that use for example the incidence of the nine most ancient fold architectures in metabolic subnetworks (derived from MANET) can be used to suggest ancestral-descendant relationship between subnetworks.

4. Finally, PAJEK provides tools for visualization of networks. Networks can be viewed using different spatial representations. Parameters that measure the size and connectivity of network components can be calculated. These parameters can be used to establish for example if a network has scale-free properties driven by power law behavior. Remember that the structure of networks can impact their function.

Your project needs to address some important questions. Examples related to point 1-4 above: Does the structure of subnetworks change in the course of evolution? Are there links between protein structure and function? Are ancient subnetworks more versatile? We recommend that you discuss possible projects with your group, rate top candidates, and then consult with the instructors on their feasibility.

Your project needs to include an experimental design that would somehow address that question. Include controls if these are available to you. Make explicit the individual step of your analysis and how you plan to achieve each of them. In this process you will evaluate how you are going to obtain the data and how you are going to analyze it. Consider what will be your expectations in light of possible outcomes.

You will then need to work with group members to accomplish goals and produce a collective report in which you state: 1) the basic question addressed, 2) the experimental design, 3) results, and 4) a discussion of results and their implication. You should also make explicit the contributions made by each of work group members to the project (very much as you would do in a standard scientific paper contribution). In presenting your results I would recommend you follow the organization of standard scientific papers. Provide relevant information in figures and table items. A single report crafted by the input of the entire group will suffice; i.e. you do not need to provide independent reports. Each group should send the report by e-mail to the instructors.

NOTE: Your research endeavor may produce valuable knowledge worthy of publication. In years 2011 and 2012 the work of two groups resulted in publications, both of which were presented in international conferences. Two groups in year 2020 published research that addressed COVID-19 epidemiology and genomics. Other manuscripts have been drafted and are still in the making.

Final grade: The final grade will be related to the total number of grade points you have earned during the course. Grades will be posted in *Illinois Compass* so that you can monitor your progress. Attendance to lectures, active participation, and culmination of laboratory activities is a plus. A total points-to-grade scale will be based on gaps in the total final score of the class students.

IMPORTANT: Students are responsible of ensuring that their work is correctly and successfully submitted electronically and should notify the instructor of any problems in this matter *at least 30 minutes* before the homework deadline. Students are encouraged to submit their homework assignments *at least 40 minutes* before the deadline.

Attendance

Class materials will be complemented with additional information provided during the class and laboratory hours by the instructor. **Students are expected to attend class and actively participate in learning activities.** Students that miss a class are expected to read the corresponding class materials and obtain the additional information provided in class from a student that attended the class. Speakers that miss their presentation will loose the grade points associated with the activity.

Academic integrity

The University of Illinois at Urbana-Champaign *Student Code* should also be considered as a part of this syllabus. Students should pay particular attention to Article 1, Part 4: Academic Integrity. Read the Code at the following URL: http://studentcode.illinois.edu/ .

Academic dishonesty may result in a failing grade. Every student is expected to review and abide by the Academic Integrity Policy: http://studentcode.illinois.edu/. Ignorance is not an excuse for any academic dishonesty. It is your responsibility to read this policy to avoid any misunderstanding. Do not hesitate to ask the instructor(s) if you are ever in doubt about what constitutes plagiarism, cheating, or any other breach of academic integrity.

Students with disabilities

To obtain disability-related academic adjustments and/or auxiliary aids, students with disabilities must contact the course instructor and the as soon as possible. To insure that disability-related concerns are properly addressed from the beginning, students with disabilities who require assistance to participate in this class should contact Disability Resources and Educational Services (DRES) and see the instructor as soon as possible. If you need accommodations for any sort of disability, please speak to me after class, or make an appointment to see me, or see me during my office hours. DRES provides students with academic accommodations, access, and support services. To contact DRES you may visit 1207 S. Oak St., Champaign, call 333-4603 (V/TDD), or e-mail a message to disability@uiuc.edu. http://www.disability.illinois.edu/.

Emergency response recommendations

Emergency response recommendations can be found at the following website: http://police.illinois.edu/emergency/. I encourage you to review this website and the campus building floor plans website within the first 10 days of class. http://police.illinois.edu/emergency/floorplans/ .

Family Educational Rights and Privacy Act (FERPA)

Any student who has suppressed their directory information pursuant to *Family Educational Rights and Privacy Act* (FERPA) should self-identify to the instructor to ensure protection of the privacy of their attendance in this course. See http://registrar.illinois.edu/ferpa for more information on FERPA.

Policies on computer resources and copyrights

All students must adhere to the rules and policies indicated by the software, websites and computer laboratories used for course related purposes. The policy on course notes and related printed and internet materials (e.g. published articles, website information) copyrights follows The General Rules Concerning University Organization and Procedure (U. of I. Board of Trustees, 1998) and can be found at http://www.vpaa.uillinois.edu/policies/patents.htm and any other rule mentioned in the materials.

Required reading materials

Caetano-Anollés G (2010) Evolutionary genomics and systems biology. John Wiley & Sons.
Caetano-Anollés G (2021) Untangling molecular biodiversity. World Scientific.

Other reading materials

Structural bioinformatics and systems biology

Bourne PE and Weissig H (2003) Structural bioinformatics. Wiley-Liss.
Orengo CA, Jones DT and Thornton JM (2003) Bioinformatics: genes, proteins and

computers. Bios Sci Pub.
Lesk AM (2001) Introduction to protein architecture. Oxford Univ Press
Baldi P, Brunak S and Brunak S (2001) Bioinformatics. MIT Press.
Kitano H (2001) Foundations of Systems Biology. MIT Press.

Applied Bioinformatics

Zvelebil M and Baum JO (2008) Understanding bioinformatics. Garland Science.
Lesk AM (2002) Introduction to bioinformatics. Oxford Univ Press
Baxevanis AD and Ouellette BFF (1998) Bioinformatics. A practical guide to the analysis of genes and proteins. John Wiley & Sons.
Waterman MS (1996) Introduction to computational biology. Maps, sequences and genomes. Chapman & Hall.
Bishop MJ (1999) Genetics Databases. Academic Press.
Setubal J and Meidanis J (1997) Introduction to computational molecular biology. PWS Publishing.
Weir BS (1996) Genetic data analysis II: methods for discrete population genetic data. Sinauer Associates.
Campbell AM and Heyer LJ (2002) Discovering genomics, proteomics, and bioinformatics. Pearson Education.
Mount DW and Mount D (2002) Bioinformatics: sequence and genome analysis. Cold Spring Harbor.
Krane DE and Raymer ML (2002) Fundamental concepts of bioinformatics. Pearson Education.
Claverie J and Notredame C (2003) Bioinformatics for Dummies. John Wiley & Sons.

Molecular Evolution

Higgs PG and Attwood T (2005) Bioinformatics and molecular evolution. Blackwell Publishing
Page RDM and Holmes EC (1998) Molecular evolution: a phylogenetic approach. Blackwell Science.