**BSE322A Bioinformatics and Computational Biology**

**3-0-0-1[10]**

**Course objective:** The objective of the course is to introduce students to the rapidly evolving field of bioinformatics. The term "bioinformatics" often means different things to different scientists, and our goal is not to cover all those things. Rather, we will aim to cover in the lectures the most fundamental topics, such as sequence alignment and pattern finding, and then explore some of the frontier areas through the "team debates". After completing this course, the students will gain an understanding of the computational challenges (and their solutions) in the analysis of large biological data sets; they will understand how some of the commonly used bioinformatics tools work, how to use these tools effectively, and how to read and evaluate research articles in the field.

**Lecture schedule:**
Mon: 5:10 - 6:00 pm, Wed: 5:10 - 6:00 pm, Thu: 2:00 - 2:50 pm

Venue: L-4

Instructor: Nitin Gupta (guptan@iitk.ac.in)

Office hour: Mon 6-7 pm

TAs: Arka Ghosh (arkag@iitk.ac.in) and Surbhit Wagle (wagles@iitk.ac.in)

Course Website: https://piazza.com/iitk.ac.in/secondsemester2018/bse322a/

Google calendar for this course:
https://www.google.com/calendar/embed?src=q1mnido7aqiicrbdvav5285i2o%40group.calendar.google.com&ctz=Asia/Calcutta

**Prerequisites**

Although there are no strict prerequisites, the course requires some familiarity with molecular biology (what is a gene? a protein?) and some comfort level with probability and statistics (what is standard deviation? a probability distribution?).

**Course Content (and approx. number of lectures for each topic)**

- Introduction to bioinformatics, biological databases and their growth, concept of homology, pairwise sequence alignment, dotmatrix plot (4)
- Dynamic programming, global (Needleman-Wunsch) and local (Smith-Waterman) alignments (7)
- Scoring matrices (PAM and BLOSUM families), gap penalty, statistical significance of alignment, BLAST (7)
- Multiple sequence alignment, Sequence Logos (6)
- Pattern finding in protein and DNA sequencing, Gibbs Sampler, Hidden Markov Models, Profiles construction and searching (9)
- Phylogenetics, gene prediction (7)

**Grading**

Following the principle of continuous evaluation, the grading will be based on these components:

- 5 periodic quizzes, 8% each. Best 4 of 5 will be used for grading. 32% total.
- Assignments: 14%
- Mid-sem exam: 15%
- End-sem exam: 15%
- Team Debates: 15%
- Class participation and attendance: 9%

In the team debates, students will form teams. The instructor will select topics (debatable claims) related to current developments in the field. Each team will be assigned in advance to prepare a case “for” or “against” one claim. You must read and cite recent research papers to support your arguments.

Reports: The teams will prepare and submit a report on their topics (about 3 pages, including a brief introduction of the topic, and then the points of the team, with references to research papers). You must not copy sentences from the papers; any information you are taking from other sources should be first understood by you and then written in your own words (with citation to the original source). Do not try to copy and change one or two words here and there; copying will result in heavy penalty in marks for the whole team (some teams were penalized last year). You will get more marks for a brief but original report than a long but plagiarized one.

Classroom debates: After report submission, we will hold classroom debates where two teams, one "for" and one "against" a claim, will present their arguments and challenge the points of the opposite team. The debates will be moderated by the instructor to ensure that all team members participate. The concepts covered in these debates will be included in the syllabus for subsequent quizzes and exams.

Break-up of the 15% marks for the debates:

- 6% for the report (all team members make the report together)
- 3% for effectively communicating their points in the debate (each team member will be allowed to speak and make a point in about 1-2 minutes in the beginning of the debate)
- 3% for effectively defending the claims during the debate
- 3% for effectively challenging the opposition's claims

Clarifications on debates: the goal of the debates is not to evaluate your communication skills. Rather, it is to see whether you have read recent research papers and thought carefully about the topic. I understand that some students have weak communication skills and have difficulty with English language. I will moderate the discussions so that these students can also present their points comfortably. It is not the case that one team will win and other will lose in the debate. Both competing teams can get high marks. The evaluation will be based on the points presented by each team (and that will depend mostly on how much reading and thinking the team has done).

The marks for class participation and attendance will be decided by the instructor based on the attendance (which will be taken only on randomly selected days) and the contribution to discussions in the classroom and on the course website.

Each quiz will cover material covered after the previous quiz. Each exam will cover material from the beginning of the course.

The copies will be rechecked only if a correct solution is marked incorrect. No requests for more partial marks will be entertained. If further mistakes are found in an answer during re-checking, marks will be reduced.

No mercy pleas for better grades after the final exam will be entertained.

**Home assignments**

Practical bioinformatics assignments will be provided. These will be graded.

**Make-up policy**

Quiz: If you miss one quiz (out of five conducted), you can be scored on the remaining four (but you lose the opportunity of ignoring your worst quiz). If you miss more than one, you will lose the corresponding marks. No make-up quizzes will be conducted, even if you have official leave sanctioned.

Attendance: Attendance will be taken on randomly selected days. No make-up or prorating will be possible, even if you have official leave sanctioned.

Team debates: The debates will be scheduled in advance to ensure all members of the debating teams can be present. If a student still misses it, he/she will lose corresponding marks and no make-up will be allowed.

Mid-sem: If a student is absent in the mid-sem with prior approval of the instructor, the end-sem can be weighted more to cover mid-sem marks (but remember that the end-sem will be harder as it covers more syllabus than the mid-sem).

End-sem: Missing end-sem is not allowed, and any circumstance that necessitates it should be discussed in advance with the instructor.

Assignments: Assignments will be given with plenty of time for submission, and thereafter no excuse for late submission will be entertained.

**Code of Conduct**

Any instance of cheating or any other academic fraud will invite very STRICT punishment (reporting to SSAC, and change of grade or deregistration from the course). No leniency will be shown in such a case, even if it is a first-time offense.

**Reference Book**

1. "Introduction to Bioinformatics Algorithms" by Neil Jones and Pavel Pevzner.

2. "Bioinformatics" by David Mount (2nd edition).

The books are not compulsory, but students are encouraged to read them for learning the subject in more detail. Both books are available in the library, and can also be found online.