BINF 6201/BINF 8201: Molecular Sequence Analysis

SYLLABUS, Spring

<table>
<thead>
<tr>
<th>Instructor: Dr. Zhengchang Su</th>
<th>Contact Information: <a href="mailto:zcsu@uncc.edu">zcsu@uncc.edu</a></th>
</tr>
</thead>
<tbody>
<tr>
<td>Office Hours: Tue and Thur: 2:00~3:00pm</td>
<td>Time/Location of Course: Bioinformatics 217</td>
</tr>
</tbody>
</table>

Main textbooks: Higgins and Teresa K. Attwood, Bioinformatics and Molecular Evolution

Other Reference Textbooks
1. Molecular Evolution, Wen-Hsiung Li, Sinauer Association
2. Biological Sequence Analysis, Richard Durbin et. al. Cambridge Univ. Press

Additional readings from the current literature may be assigned as appropriate. Lectures will be based on relevant chapters in these four textbooks.

A.) COURSE DESCRIPTION
This course introduces the basic computational methods and open sources software commonly used in molecular sequence analysis. The course covers biological sequence data formats and major public databases, concepts of computer algorithms and complexity, introductions to principle components analysis and data clustering methods, dynamics of genes in populations, evolutionary models of DNA and protein sequences, derivation of amino acid substitution matrices, algorithms for pairwise sequence alignments and multiple sequence alignments, algorithms for fast sequence database search, methods for molecular phylogenetic analysis, hidden Markov models and neural networks for sequence pattern and family recognition, and introductions to genome evolution and omics data analysis. (3 credit hours)

B.) PRE/CO-REQUISITES
Co-requisite: BINF6201L/8201L. Prerequisite: BINF6200/8200 or equivalent. Students should also have basic concepts in statistics or take BINF6200/8200 simultaneously when taking this course.

C.) LEARNING OBJECTIVES
Having successfully completed this course, the student will be able to:

1. Understand the mathematical models of the dynamics of genes in populations.
2. Understand the mathematical models of changes in DNA and protein sequences.
3. Understand the pair-wise and multiple sequence alignment algorithms and common database search algorithms.
5. Use the open source implementations of these algorithms to solve real-world sequence analysis problems.
6. Lay a solid foundation for the subsequent more advanced bioinformatics and computational courses.

D.) INSTRUCTIONAL METHODS
The course is presented in a lecture format which will include the following elements as appropriate: presentation of factual material in a standard lecture format, interactive demonstrations of methods to be applied in assignments, and opportunities for student to ask questions.

E.) GRADING
Students will be evaluated on their ability to answer factual questions regarding material presented in the class and assigned texts, to correctly solve problems using methods presented in the class, and to participate productively in critical discussion of course material. Weekly or bi-weekly homework assignments (30%), midterm exam (25%), final (35%) and lab and classroom participation will count for 10% of the grade.

Grades will be assigned on the following scale:
A: 90-100%
B: 75-89%
C: 60-74%
U: 0-59%

F.) TENTATIVE SCHEDULE (WEEKLY, OR TOPICS)

<table>
<thead>
<tr>
<th>Date (Week)</th>
<th>Subject/Topic</th>
<th>Assignments Due</th>
</tr>
</thead>
<tbody>
<tr>
<td>Week 1</td>
<td>Lecture 1: Introduction</td>
<td></td>
</tr>
<tr>
<td>Week 1</td>
<td>Lecture 2: Basics of molecular biology</td>
<td></td>
</tr>
<tr>
<td>Week 2</td>
<td>Lecture 3: Introduction to principle components analysis</td>
<td></td>
</tr>
<tr>
<td>Week 2</td>
<td>Lecture 4: Introduction to clustering analysis</td>
<td></td>
</tr>
<tr>
<td>Week 3</td>
<td>Lecture 5: Dynamics of genes in populations 1</td>
<td></td>
</tr>
<tr>
<td>Week 3</td>
<td>Lecture 6: Dynamics of genes in populations 2</td>
<td></td>
</tr>
<tr>
<td>Week 4</td>
<td>Lecture 7: Evolutionary change in DNA sequences 1</td>
<td></td>
</tr>
<tr>
<td>Week 4</td>
<td>Lecture 8: Evolutionary change in DNA sequences 2</td>
<td></td>
</tr>
<tr>
<td>Week 5</td>
<td>Lecture 9: Protein evolution models</td>
<td></td>
</tr>
<tr>
<td>Week 5</td>
<td>Lecture 10: Biological sequence databases</td>
<td></td>
</tr>
<tr>
<td>Week 6</td>
<td>Lecture 11: Sequence alignment algorithms 1</td>
<td></td>
</tr>
<tr>
<td>Week 6</td>
<td>Lecture 12: Sequence alignment algorithms 2</td>
<td></td>
</tr>
<tr>
<td>Week 7</td>
<td>Lecture 13: Searching sequence databases 1</td>
<td></td>
</tr>
</tbody>
</table>
Week 7 | Lecture 14: Searching sequence databases 2
Week 8 | Lecture 15: Searching sequence databases 3
Week 8 | Lecture 16: Phylogenetic methods 1
Week 9 | Lecture 17: Phylogenetic methods 2
Week 9 | Lecture 18: Phylogenetic methods 3
Week 10 | Lecture 19: Phylogenetic methods 4
Week 10 | Lecture 20: Patterns in sequences
Week 11 | Lecture 21: Probabilistic methods and machine learning 1
Week 11 | Lecture 22: Probabilistic methods and machine learning 2
Week 12 | Lecture 23: Probabilistic methods and machine learning 3
Week 12 | Lecture 24: Probabilistic methods and machine learning 4
Week 13 | Lecture 25: Neural networks for sequence analysis
Week 13 | Lecture 26: Further topics in molecular evolution
Week 14 | Review

G.) POLICIES AND PROCEDURES

a. ACADEMIC INTEGRITY
   All students are required to read and abide by the Code of Student Academic Integrity. Violations of the Code of Student Academic Integrity, including plagiarism, will result in disciplinary action as provided in the Code. Definitions and examples of plagiarism are set forth in the Code. The Code is available from the Dean of Students Office or online at: http://www.legal.uncc.edu/policies/ps-105.html. A set of links to various resources on plagiarism and how to avoid it is available at the UNCC Library website: http://library.uncc.edu/display/?dept=instruction&format=open&page=920.

b. ATTENDANCE
   Attendance at lecture is required, although exceptions will be made for reasons such as illness or family emergency. Excessive absences will result in a reduced classroom participation score at the instructor’s discretion, and will negatively impact the overall course grade.