MICR4203/5203 – Principals of Bioinformatics

Instructor: Robert L. Burnap, Department of Microbiology & Molecular Genetics

Email: rob.burnap@okstate.edu (allow 1 business day for response); Please use office hours for routine discussion of the course material and mechanics.

Office: 130P Henry Bellmon Research Center,

Office Hours: in person, by appointment.

Course Website: D2L (https://oc.okstate.edu/d2l/home/1244636)

Lectures: Tuesdays –Thursday 9:00-10:20 am 113 Life Sciences East

Required Textbook: Zvelebil & Baum *Understanding Bioinformatics* 798 pages 2007 (ISBN 9780815340249)

note: Fall 2017 this will change to *Concepts in Bioinformatics and Genomics* 2016 First Edition
Jamil Momand, Alison McCurdy, Contributions by Silvia Heubach, and Nancy Warter-Perez
ISBN: 9780199936991

Course Description: The course provides life scientists and pre-professionals a foundation in the concepts of biological sequence information and the inferential techniques used to assign structure, function, and evolutionary relationships among genes and proteins. It introduces the theory and application of available software used to assign functions to genes and gene networks and understand network characteristics such as metabolism from genomic analysis of predicted enzymes. It is inherently multidisciplinary and involves, most notably, molecular biology and computer science. The first part of the course emphasizes the mechanics of molecular evolution and how bioinformatic tools, notably scoring matrices, are designed to simulate these processes in a way that allows predictive powers for things such as sequence alignments and inferring phylogeny. Second part of the course recognizes that assigning functions to genes and the gene networks of genomes remains a huge challenge. These challenges are clarified and approaches to their solution are covered. Although computers and biology have been united for decades now, the impetus for bioinformatics derives from the need to make biological sense out of the explosion of genetic sequence data that floods in from modern high-throughput sequence projects. This course is oriented to the intelligent understanding and application of computers to the problems of modern molecular biology. Although the course does not require extensive computer programming, it covers basic algorithms used, for example, for analyzing sequence similarity focusing on the development of scoring matrices and introduces file manipulation using exercises in the Python language. This course is the foundation course for the OSU Bioinformatics Certificate Program.

Prerequisites: Cell & Molecular Biology MICR3033 or Biochemistry BIOC3653 or equivalent. An excellent understanding of the fundamentals of molecular genetics (i.e. DNA->RNA->protein) and the basics of protein structure is assumed. A good familiarity with the operation of a personal computer is also assumed.

Graded Material:

Homework assignments= Assigned Project Tasks: A major component of the class is to perform a set of homework tasks relating to the course. Reading assignments from your textbook and primary literature will be given in class. Since the text does not cover all of the topics covered in the course, additional readings in the form of handouts or library materials will be required.   *Homework will often be given as an email message that you will receive before 5 pm the day before class*. **Pay attention to your Okstate email.**

Online Discussion: There will be weekly questions for discussion in the D2L 'Discussion' section at the D2L website. Please adhere to 'Netiquette' guidelines for your posting this will influence the grade you received for this weekly contribubions. Please note that full credit will be assigned for thoughtful and informative comments and not simply stating agreement or disagreement with the existing posts.

Quizzes: Quizzes may be unannounced and will typically cover material covered in reading assignments or to assess homework. For the latter, homework will be maintained in your digital notebook, but the actual assessment on whether comprehension was obtained will be made using a quiz covering recent homework.

Exams: Three exams (2 midterm 1 comprehensive final exam) will be given. Each exam will consist of a combination of short answer, computation, and short essay type questions. Questions will be based upon material from lectures, discussion, and reading assignments. Questions based upon material in reading assignments, but not lectured on, may appear on exams. No make-up exams will be given - if you have an <u>excused</u> absence you must make arrangements prior to the scheduled exam time.   Excused exam absences are granted only for truly emergency conditions. Final Exam is scheduled for Tuesday December 6, 8:00-9:50. **Note: arrangements for final examinations at earlier dates will not be made.**

Graduate Credit:

Graduate student are highly encouraged to take the class, but will be required to complete a project that employs the basic principals of the course.   Students must consult with the instructor to define the project, which, minimally, will be comprised of a well-researched statement of a bioinformatics problem and a documentation of a set of completed tasks aimed at its solution. Typically, this will involve a project on the student's thesis topic. The final project will be given in digital notebook form according to the initial specifications, which typically follow the fours main tasks for the assigned gene sequence. The graduate project will be worth an additional 100 points.

Graded Points:

Exams (3) at 90, 110 and 150pts	350 pts
Quizzes/Discussion (both online and in-class)	50 pts
Assigned Project Tasks (homework)	100 pts
Total Points for Course	**500 pts**

Grades for tests and submitted items will be posted at the D2L site within one week. Every effort will be made by the instructor to ensure the latest grade information is available prior to key course dates (exams, 'W'' withdrawal dates).

Letter Grade Assignments:
450-500 pts	= A
400-449 pts	= B
350-399 pts	= C
300-349 pts	= D
Below 300 pts	= F

Late Policy for Assignments: Late assignments will have 10% possible points on the first day, and an additional 10% deducted for each of the subsequent days the assignment is late.

Drop Policy
Information about university drop policy and dates is at this website:
http://registrar.okstate.edu/
Click on "class schedules," and "short, internet, and outreach courses"
To drop this course, contact the Registrar's office, (405) 744-6876, or drop through SIS (Student Information Services).

Academic Honesty: Cheating on exams or any form of academic dishonesty will not be tolerated. Flagrant violations may result in an automatic "F" for the entire course. It is our hope and your responsibility that this situation does not arise. Participating in a behavior that violates academic integrity (e.g., unauthorized collaboration, plagiarism, multiple submissions, cheating on examinations, fabricating information, helping another person cheat, unauthorized advance access to examinations, altering or destroying the work of others, and fraudulently altering academic records) will result in your being sanctioned. Violations may subject you to disciplinary action including the following: receiving a failing grade on an assignment, examination or course, receiving a notation of a violation of academic integrity on your transcript (F!), and being suspended from the University. You have the right to appeal the charge. Contact the Office of Academic **Affairs, 101 Whitehurst, 405-744-5627,** http://academicintegrity.okstate.edu/**.**

Withdrawals and Incompletes: Please consult university schedules and policy. Incompletes ("I") will be given only in truly exceptional circumstances and students must be passing the course.

Internet Netiquette Guidelines (adapted from OSU Preparing Online Instruction course)
A melding of the words "network" and "etiquette", **netiquette** refers to the manner in which communication is conveyed in an electronic environment.

Here are some guidelines for communication within this course:
- REFRAIN FROM USING ALL CAPS. It is considered SHOUTING when communicating online.

- Do not post or forward offensive or racially insensitive jokes or comments.

- Be careful with humor and sarcasm.
- Don't respond to personal attacks: Contact the instructor for action and referral.
- Always add in the subject line a concise statement describing the email or discussion post.
- Respect others' opinions. If you disagree with what another has said, post your thoughts in an objective, respectful manner. Do not make remarks that can be taken personally.
- Reflect upon the text you have entered before posting.
- Keep the discussion within the scope of the course material.
- Communication should be grammatically correct. Adhere to correct sentence structure, grammar, and spelling conventions. Proofread for errors before posting a message.
- Before you respond to a threaded message, read all the messages related to that message that have been previously posted.

- Send out an email to a group using the blind carbon copy field – BCC does not allow your recipients to view who was sent the email.

Outline of the Course:

The course consists of ten conceptual modules that introduce the principals of bioinformatics:

1. **Algorithmic thinking and expression for molecular genetics**

2. **Databases and Molecular File Structure**

3. **Biological Sequence Alignments**

4. **Pairwise Alignments and Database Searches**

5. **Patterns, Profiles, and Multiple Alignments**

6. **Inferring Evolutionary History: Phylogenetic Methods**

7. **Analysis of Genomes**

8. **Protein Structure/Function Prediction**

9. **Proteome and Gene Expression Analysis**

10. **Systems Biology & Metabolomic Reconstruction**

Tentative Schedule

	Date	Topic	Reading	Comments
1	Aug 16	**Introduction and background:**	Install test editors discuss some available software	
2	Aug 18	**Special instruction: Dr. Dana Brunson, Director of OSU High Performance Computing Center: BLAST**	Website: https://hpcc.okstate.edu/content/new-user-tutorial	Bring laptops with terminal application
3	Aug 23	**Setting up with iPython**	Chapters 1-2	
4	Aug 30	**Databases and Molecular File Structure**	Chapter 3	
5	Sep 1	**Biological Sequence Alignments and the statistical/structural development of scoring matrices**	Chapter 4	
6	Sep 6	**Multiple sequence alignments**	Selection of atoms, residues, 3D structural alignments, PyMol scripting	
7	Sep 8	*PyMOL:* **understanding scripts and algorithms**	Be sure you have already installed PyMol!	
8	Sep 13	**Pairwise sequence alignments and database searching**	Chapter 5: skip page 134 to end of chapter	
9	Sep 15	**Concepts of dynamic programming**	See PowerPoint on the Dynamic Programming and corresponding textbook	
10	Sep 20	**Midterm Exam 1**		
11	Sep 22	**BLAST re-visited,** Psi-BLAST		
12	Sep 27	**Patterns, Profiles, and Multiple Alignments**	Chapter 6 skip: 172-180, skip 185-end	
13	Sep 29		***Task 1 assignment***	

14	Oct 1	**Hidden Markov Models**		
15	Oct 4	**Using Python with BioPython libraries**	Exercise in Fasta header manipulation	
16	Oct 6	**Inferring Evolutionary History: Phylogenetic Methods**	Chapter 7 read all; Chapter 8 skip; be sure you have already installed MEGA	
17	Oct 11			
18	Oct 13			
19	Oct 18	**Midterm Exam 2**		
20	Oct 20	**Proteins Structure/Function Prediction**	Chapter 11: pp. 411-430, 436-442, Chap 12 skip	
21	Oct 25	**Modeling protein structure**		
22	Oct 27	**Domain structure, rearrangements, and COGs**		
23	Nov 1	**Analysis of genomes and comparisons of related genomes**		
24	Nov 3			
25	Nov 8			
26	Nov 10	**Clustering techniques: Proteome and Gene Expression Analysis**		
27	Nov 15	**Python algorithms for clustering**		
28	Nov 17	**Systems Biology**		
29	Nov 22	**Metabolic Reconstruction/Flux Balance Analysis**	Chapter 15 skip 604-605	
30	Nov 29			Grad Projects Due Dec 1
31	Dec 1			
	Dec 6	**Final Exam**		8:00 - 9:50 am