BF 601 - Introduction to Bioinformatics

Instructor: Abigail S. Newsome

Office Location: Science and Tech Center Museum
E-mail Address: asnewsome@mvsu.edu

Class Meetings Location/Time: M/W 1 PM – 2:15 PM
Bioinformatics Computer Lab

Phone: 662-254-3395
Office Hours: 7AM-1PM/M, 7AM-10AM/W and others by appt

All email communication must include the words "BF601" in the subject line. I cannot guarantee an answer, or that I will even see any message that does not contain this information in the subject.

COURSE DESCRIPTION

Advances in molecular biology techniques have brought forth a wealth of information waiting to be mined. Bioinformatics and Computational Biology is a multidisciplinary field bringing together experts in computer science, molecular biology, physics, mathematics, chemistry and other disciplines to form synergistic collaborations. Introduction to Bioinformatics will focus on the current state of the art programs designed for sequence alignment, database searching, RNA structure prediction, microarray sequence analysis, gene prediction, repeat detection, and protein folding prediction. This course will also focus on the basic concepts of bioinformatics and how to identify, obtain, establish, maintain and exchange research information in biology.

PREREQUISITES

None

PURPOSE

This course introduces the emerging topic of bioinformatics. It is designed primarily for life science students who do not have an extensive background in mathematics, statistics, or computer science but who are interested in survey-level knowledge of bioinformatics and its techniques. The course is also suitable for students in mathematics, computer science, or statistics who wish to learn how methods drawn from their discipline are applied as bioinformatics. It is a 'hands on,' 'how to' lecture, demonstration, and laboratory (optional) course that includes both sequence- and structure-based methods. Guest faculty members may lecture in their individual areas of expertise and knowledge.

GENERAL COURSE GOALS

The following general course goals are established to meet the purpose of BF601:

1. Gain an understanding computational tools needed for a wide range of genomics problems.
2. Gain an understanding of working in interdisciplinary teams of biologists, biochemists, medical r researchers, geneticists, and computer engineers.
3. Gain an understanding of basic bioinformatics problems and their solutions, including: nucleotide and protein sequence comparison, complexity analysis, sequence search, alignment, assembly, and gene clustering.
4. Gain an understanding of computational biology modeling problems including: gene prediction and transcriptome interpretation.
5. Gain an understanding of computer Markov Model building and searching of large databases.
6. Gain an understanding of Phylogenetics, Genetic linkage analysis, and map construction.
7. Gain an understanding of gene expression, and the current state of understanding of the mechanisms controlling regulation of gene expression. Understand computational methods for analysis of microarray technologies, and interpretations of gene expression from this data.
8. Using compelling biological examples, sequence analysis problems are shown and solved.

COURSE OBJECTIVES

Bioinformatics has been defined as the science of examining the structure and function of genes and proteins through the use of computational analysis, statistics, and pattern recognition. A number of recent workforce studies have shown that there is a high current and unmet demand for people trained to various levels of expertise in bioinformatics, from technicians and technical librarians to developers of new and improved methodologies and applications. National estimates of needed positions in the field in the next four to five years is about 20,000. Bioinformatics is a rapidly evolving and developing field both in terms of breadth of scope of useful applications and in terms of depth of what can be accomplished.

This course is designed to introduce bioinformatics at a level appropriate for biology majors having completed the lower-division core, and for chemistry, computer science, and math majors with an interest in biology. This course is designed so that the content and curricula can rapidly adjust as required to meet changing circumstances during the course of the semester and to evolve with the topics of interest in bioinformatics over time. Students will learn to use conventional software, web-based applications, and software which they download to their machine. By using the well-tested and successful approach of problem-based learning, students will learn through applying the strategies and tools used in bioinformatics to topical problems drawn from ongoing research and applications in a variety of fields. There is to be an integration of the basics of computation and analysis along with chemistry and biology throughout the course.

A well designed course should prepare students to meet the expected outcomes of having taken that course. As a minimum at the end of this course, there should be a solid understanding of the scope of bioinformatics. It is expected that the basic knowledge can be applied in many different fields of interest. Students should gain substantial competency in content, skills, and awareness within the field of bioinformatics. Many of the problems presented in the course will serve as launch points for further inquiry and exploration as students move on into other courses.

The fields of study in molecular and cell biology continue to grow rapidly. The field of bioinformatics changes so rapidly that it is nearly impossible to keep up with all the different areas of progress. It is therefore necessary to learn how to access new information and how to assimilate it into the whole, in order to continue to learn beyond the limits of this course. It is also necessary to become aware of the tools and techniques used in research and in applications. Therefore, while learning the essentials of bioinformatics, you will also begin to learn to use some of these tools. In doing so, you should find that they help you in learning beyond this course and to see the connections in other courses.

COURSE CONTENT

Required Text

Mukhopadhyay, Chandra. 2018. *Basic Applied Bioinformatics.* John Wiley and Sons, Inc. ISBN 9781119244332.

Supplemental Text

Agostino, Michael. 2012. *Practical Bioinformatics.* Garland Science, Taylor & Francis Group, LLC.

Course Project

During the month of Septembr, you will be given an unknown genomic sequence to analyze. As we learn new techniques, you should apply them to your sequence to try and understand what your sample DNA codes for. You should also explore the web for any kind of analysis you might find useful. During the term you should stop in regularly to talk with me (or email me) about your progress and to discuss what to do next -- this should be a collaborative effort! At the end of the term, you will hand in a report detailing what you have learned about your sequence. The report should include specifics on the methods/algorithms you used, what the results of each analysis was and what conclusions you were able to draw about your sequence from each analysis. More details are available on the unknowns in an additional handout.

Major areas of study for BF601 will include:

	Topic/Event	Homework (Due Date)
August 20	Introduction and Chapter 1	
August 22	Chapter 1	1.6
August 27	Chapter 2	2.4
August 29	Chapter 3	3.4
Sept 3	Chapter 4	4.4
Sept 5	Chapter 5	
Sept 10	Chapter 6	6.9
Sept 12	Exam 1	
Sept 17	Chapter 7	
Sept 19	Chapter 8	8.6 #2
Sept 24	Chapter 9	9.4 #3 & #5
Sept 26	Chapter 9	
Oct 1	Chapter 10	10.4 #3
Oct 3	Chapter 11	11.6
Oct 8	Chapter 12 & 13	12.4 #5/13.4 # 4 & 5
Oct 10	Midterm Exam	
Oct 15	Chapter 14-16	
Oct 17	Chapter 21 and 23	21.6 #1/23.5 #1,2 & 4
Oct 22	Chapter 27	27.6
Oct 24	Chapter 38	38.5
Oct 29	Chapter 39	39.8
Oct 31	Chapter 40	40.5
Nov 5	Exam 3/Chapter 36 and 37	
Nov 7	Chapter 41	41.5
Nov 12	Chapter 42	42.4
Nov 14	Chapter 43	43.3
Nov 19-Nov 23	Fall Break	
Nov 26	Chapter 44	44.5
Nov 28	Chapter 44	
Dec 3-7	Final Examination	

USE OF TECHNOLOGY

Technology is integrated into the course to enhance and facilitate the promotion and development of the four attributes of HTM. The technology used includes but is not limited to:

1.Computerized library searches;
2.Internet searches; and
3.Bioinformatics computer applications - free and commercial.

EVALUATION AND GRADING PROCEDURES

Grading:

Tutorial Exercises (100 Points):

There will be several tutorial sessions during the semester that will complement this class. These sessions will focus on the applications taught during the lecture. Tutorial exercises will usually be due one week after the tutorial. While the tutorial sessions and lecture discussions are interactive, these exercises should be done individually. Copying another student's answers is considered cheating and will result in a failure for the exercise.

Class Participation (150 Points):

Lecture sessions and discussions are designed to be interactive. Your active participation will be considered as part of your final grade. Active participation includes asking questions in lecture, contributing to lecture discussions, being prepared with course book, pen/pencil and paper, and interacting effectively with others during the group projects. These 150 points will be divided amongst each class meeting. There are 29 class meetings (5.17 points each meeting). You cannot receive the points if you do not attend prepared for class and engaged (writing utensil/paper-1, book -2 , and participation – 2.17).

Course Project (200 Points):

Please review the section above regarding the project. This project will consist of an analysis of the given sequence using the tools that have been covered (100 points) and a written report (100 points).

Homework (150 Points)

Exercises are to be completed and submitted prior to the start of class on the date noted on the syllabus. NO late assignments will be allowed. A point of zero will be given immediately if the assignment is not submitted. You may submit assignments early.

Exams (400 Points)

There will be four exams administered (100 points each

Points Grading Scale

905	— 1000	= A
805	— 904.9	= B
705	— 804.9	= C
605	— 704.9	= D
Below	605	= F

Policies Regarding Attendance, Examinations, Grading and Supplies

Every student is expected to attend each lecture taking notes where applicable and recording observations. There will be no make-up lectures. If you should miss either a lecture due to illness or other circumstances, you must supply a written excuse from the dean of students. Attendance will be taken daily. Three unexcused absences in lecture will lower your final grade by ten points. You will receive a failing grade if you miss nine lectures without a written, approved excuse from the dean of students. This will be strictly enforced.

All students are responsible for reading all assigned materials prior to their discussion in class. Each student will be held responsible for completing all exercises. Classroom distractions are an annoyance to everyone and they interfere with the learning process. Chronic lateness, side conversations, unnecessary exits, cell phones or pagers are all considered unwanted distractions. Most Mississippi Valley State University students conform without reservation to the expected levels of classroom etiquette and I encourage students in this class to follow this example.

Adherence to the University's code of academic honesty, as indicated in the undergraduate catalog, is expected. Please refer to the catalog for an explanation of this code and what is expected. It is your responsibility to be familiar with activities considered to constitute academic dishonesty that may comprise your own intellectual and moral development. Any instance of academic dishonesty - cheating and/or plagiarism - will be reported to the Office of Academic Affairs along with the materials associated with the act of academic dishonesty.

Services for Students with Disabilities (SSD) Disabilities Statement

Mississippi Valley State University is committed to providing reasonable accommodations for students with a documented disability. If you feel you are eligible to receive accommodations for a covered disability (medical, physical, psychiatric, learning, vision, hearing, etc.) and would like to request it for this course, you must be registered with the Services for Students with Disabilities (SSD) program administered by University College. It is recommended that you visit the Disabilities Office located inside the EMAP Computer Lab in the Technical Education (IT) Building to register for the program at the beginning of each semester. For more information or to schedule an appointment, please contact Mr. Billy Benson, Jr. via phone or email at 662-254-3005 or billy.benson@mvsu.edu.

Note

This syllabus is not a binding contract between the instructor and the student. It is intended to serve as a guide for the course. All deviations will be announced and submitted to students in writing.

Computer Lab

Please check the hours of operation of the computer lab. You are welcome to use the facility in a responsible manner at any unoccupied time between the hours of 8 AM and 5 PM on Monday - Thursday and between 8 AM and 4 PM on Fridays. The lab is for Bioinformatics students only. Should you be found allowing others to accompany you in the use of the lab or be a visitor of yours in the lab, your usage privileges will be terminated. Drinking and eating foods are not allowed past the doors of the lobby.

Recommended Reading List

Sequence Alignment

Deleeuw L, Tchernatynskaia AV, Lane AN. (2008) Thermodynamics and specificity of the Mbp1-DNA interaction. Biochemistry. 2008 Jun 17;47(24):6378-85.

Waterhouse *et al.* (2009) Jalview Version 2--a multiple sequence alignment editor and analysis workbench. *Bioinformatics* **25**:1189-91.

Kemena & Notredame (2009) Upcoming challenges for multiple sequence alignment methods in the high-throughput era. *Bioinformatics* **25**:2455-65.

Benítez-Páez *et al.* (2012) A practical guide for the computational selection of residues to be experimentally characterized in protein families. *Brief Bioinformatics* **13**:329-36.

Phylogenetics

Swofford, D. L. et al. (1996). Phylogeny Reconstruction. in *Molecular Systematics* 2nd Edition. D. M. Hillis, C. Moritz and B. K. Mable eds. Sunderland, MA, Sinauer Assoc., Inc.: pp407-514.

- for those who want the details

Page, R. D. M. and E. C. Holmes (1998). *Molecular Evolution: A Phylogenetic Approach*. Oxford, Blackwell Science Ltd.

- an excellent text book with very clear descriptions of phylogenetic techniques

Databases

Tomb et al. (1997) The complete genome sequence of the gastric pathogen *Helicobacter pylori*

Motif and Profiles

Durbin *et al.* 1998 Biological Sequence Analysis: Probabilistic models of proteins and nucleic acids. Cambridge Univ. Press, Cambridge - chapter 5.

Zdobnov and Apweiler (2001) Bioinformatics 17:847-848.

Gene predicition and annotation

Allen et al. (2004) Computational gene prediction using multiple sources of evidence. Genome Res 14:142-8

Microarray Data Analysis

Behr et al. (1999) Science 284:1520-1523. Comparative Genomics of BCG Vaccines by Whole-Genome DNA Microarray
Harmer et al. (2000) Science 290:2110-2113. Orchestrated Transcription of Key Pathways in Arabidopsis by the Circadian Clock
Marton et al. (1998) Nat Med 4:1293-1301. Drug target validation and identification of secondary drug target effects using DNA microarrays
Sorlie et al. (2001) PNAS 98:10869-74. Gene expression patterns of breast carcinomas distinguish tumor subclasses with clinical implications
Thijs et al. (2002) JCB 9:447-464. A Gibbs Sampling Method to Detect Overrepresented Motifs in the Upstream Regions of Coexpressed Genes
Tusher et al. (2001) PNAS 98:5116-21. Significance Analysis of Microarrays
Zhu et al. (2003) Plant Biotech J 1:59-70. Transcriptional control of nutrient partitioning during rice grain filling

Motifs

Sigrist et al. (2010) PROSITE, a protein domain database for functional characterization and annotation. Nucleic Acids Res 38:D161-6.
Bendtsen et al. (2004) Improved prediction of signal peptides: SignalP 3.0. J Mol Biol 340:783-95.
Emanuelsson et al. (2007) Locating proteins in the cell using TargetP, SignalP and related tools. Nat Protoc 2:953-71.
SignalP 4.0: discriminating signal peptides from transmembrane regions. Nat Methods 8:785-6.

Disorder

Deng et al. (2012) A comprehensive overview of computational protein disorder prediction methods. Mol Biosyst 8:114-21.
Kozlowski & Bujnicki (2012) MetaDisorder: a meta-server for the prediction of intrinsic disorder in proteins. BMC Bioinformatics 13:111.

Secondary Structure

Pirovano & Heringa (2010) Protein secondary structure prediction. Methods Mol Biol 609:327-48.
Rackham et al. (2010) The evolution and structure prediction of coiled coils across all genomes. J Mol Biol 403:480-93.

Transmembrane Helices

Benchmark of Membrane Helix Predictions From Sequence
Wang et al. (2012) Improving transmembrane protein consensus topology prediction using inter-helical interaction. Biochim Biophys Acta 1818:2679-86.
Hennerdal & Elofsson (2011) Rapid membrane protein topology prediction. Bioinformatics 27:1322-3.
Briesemeister et al. (2010) YLoc--an interpretable web server for predicting subcellular localization. Nucleic Acids Res 38:W497-502.
Punta et al. (2007) Membrane protein prediction methods. Methods 41:460-74.

Localization

Yu et al. (2010) PSORTb 3.0: improved protein subcellular localization prediction with refined localization subcategories and predictive capabilities for all prokaryotes. Bioinformatics 26:1608-15
Briesemeister et al. (2010) YLoc--an interpretable web server for predicting subcellular localization. Nucleic Acids Res 38:W497-502.

Repeats

Heger & Holm (2000) Rapid automatic detection and alignment of repeats in protein sequences. Proteins 41:224-37.

Pellegrini et al. (2012) Ab initio detection of fuzzy amino acid tandem repeats in protein sequences. BMC Bioinformatics 13 Suppl 3:S8
Biegert & Söding (2008) De novo identification of highly diverged protein repeats by probabilistic consistency. Bioinformatics 24:807-14.

Integrated tools

Rost et al. (2004) The PredictProtein server. Nucleic Acids Res 32:W321-6
Cong & Grishin (2012) MESSA: MEta-Server for protein Sequence Analysis. BMC Biol 10:82.
Ooi et al. (2009) ANNIE: integrated de novo protein sequence annotation. Nucleic Acids Res 37:W435-40.
http://smart.embl-heidelberg.de/
http://toolkit.tuebingen.mpg.de/quick2_d/