INSTRUCTOR CONTACT INFORMATION:
Dr. Greg Krukonis Office: CAV 003A
Email: gkrukonis@angelo.edu (preferred contact)
Office hours: Posted on BlackBoard

LEARNING OBJECTIVES AND NATURE OF THE COURSE:
Lecture: Introduction to methods for acquiring, analyzing, and employing biological sequence information. Topics will include – Theory and process of PCR, mass spectroscopy, and DNA microarrays. Algorithms for searching and clustering sequences. Applications of bioinformatic data to questions such as the geographical movement of Zika virus, horizontal gene transfer in bacterial viruses, and changes in human gene expression in response to disease and treatment.

Lab: Students will access remote sequence databases (NCBI, EMBL-EBI) and analyze sequences with open source bioinformatics software running natively, in a Linux virtual machine, and on remote servers. Analyses will include protein structure prediction, phylogenetics using molecular data, and genome annotation. Students will annotate a novel viral genome and submit the completed annotation to NCBI.

Lecture Course Materials (required):
by Barry G. Hall

<table>
<thead>
<tr>
<th>Component</th>
<th>Maximum Points</th>
<th>Grading Scale</th>
</tr>
</thead>
<tbody>
<tr>
<td>Class participation Activities &amp; Homework</td>
<td>150</td>
<td>A = 90 to 100% (900-1000 pts)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>B = 80 to 89.9% (800-899.9 pts)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>C = 70 to 79.9% (700-799.9 pts)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>D = 60 to 69.9% (600-699.9 pts)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>F = &lt;60% (0-599.9 pts)</td>
</tr>
<tr>
<td>Exam 1</td>
<td>50</td>
<td></td>
</tr>
<tr>
<td>Exam 2</td>
<td>100</td>
<td></td>
</tr>
<tr>
<td>Exam 3</td>
<td>100</td>
<td></td>
</tr>
<tr>
<td>Lab</td>
<td>400 (200 for Phendrix Annotation)</td>
<td></td>
</tr>
<tr>
<td>Oral presentations</td>
<td>200</td>
<td></td>
</tr>
<tr>
<td>Total Points***</td>
<td>1000</td>
<td></td>
</tr>
</tbody>
</table>

All grades will be calculated in the same way, regardless of extenuating circumstances or any reason not related to your actual performance in the course. However much I may sympathize with your personal circumstances, I never consider them to be a basis for grade assignments. The activity and homework points serve as an extremely generous, built-in curve. I strongly encourage you to take advantage of them when they become available because once assigned they cannot be made up. Therefore you should always attend class and strive to do your best, so that you may earn the grade you want. It is your responsibility to keep up with your point total. Don’t worry I will help you, if you just ask!
Laboratory: This portion of the course offers you the opportunity to explore and apply concepts to answer research questions. Success in the laboratory involves teamwork in designing and conducting analyses report writing. In addition, you will conduct activities designed to develop and improve critical thinking and problem-solving skills related to the topics discussed in lectures.

STUDENT RESPONSIBILITIES:

Attendance: You are expected to attend all scheduled class meetings. You are expected to arrive on time and stay for the entire period. Missed lecture activity points CANNOT be made up. Attendance will be checked at each class meeting via the Top Hat system at random. Please inform me well ahead of time if you will need to be absent for any reason including religious holidays. NOTE: You are NOT automatically dropped if you stop attending class. November 1 is the last day to drop a course.

Academic Honesty and the ASU Honor Code: Angelo State University expects its students to maintain complete honesty and integrity in their academic pursuits. Students are responsible for understanding the Academic Honor Code and the ASU policies on academic dishonesty, which is contained in both print and web versions of the Student Handbook. The penalty for ANY act of dishonesty in this class, including any form of cheating or plagiarism: 1) is a grade of ZERO on the assignment and, 2) disciplinary action as warranted in accordance with university guidelines. Please do NOT jeopardize your career; it’s not worth it.

Accommodations for students with disabilities: All students at Angelo State must have the capacity to undertake, with reasonable assistance from the faculty and administration, the academic challenges necessary to fulfill the academic requirements for the degree or certification programs that they are pursuing. If you have a disability and need special accommodations of any nature, you should contact the Student Life Office (Garden Level, University Center, (325) 942-2191 or Student.Life@angelo.edu). I will be happy to make accommodations for you based on the recommendations from the Student Life Office. Please make your request early in the semester to allow time for appropriate arrangements.

Religious Holy Day: A student who intends to observe a religious holy day during the semester should make that intention known in writing to the instructor during the first week of the semester and one week prior to the absence. If this submission is completed, a student who is absent from classes for the observance of a religious holy day shall be allowed to take make up missed exams or assignments scheduled for that day in accordance with syllabus policy.

Class Preparation ASU email: Since class announcements will be routinely distributed via email, you will need to regularly check your ASU email account. Please check you ASU email daily. All course correspondence will be through your ASU email only (I will not respond to email from other accounts). Please see the email policy in Bb for more details. ASU provides Internet and email services to you at any of the computer labs on campus. Call 942-2911 to set this up if necessary.

Lecture: A typical class meeting will combine mini-lectures, discussions, group activities, multimedia presentations, and other demonstrations and activities to give you an opportunity to learn biological concepts in as active a manner as possible. Each segment of the course is structured around one or more conceptual units that can be interpreted or solved by applying selected biological concepts. You can accumulate up to 150 points toward your final semester grade from unannounced group or individual in-class activities (no make-ups) or homework assignments. We will also use Top Hat questions for in class activities.

<table>
<thead>
<tr>
<th>Week/ Date (approx)</th>
<th>Tentative lecture/lab outline</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 Aug. 27-31</td>
<td>Introduction to bioinformatics, Introduction to DNA MASTER, Genome analysis and annotation of bacteriophage phage Phendrix</td>
</tr>
<tr>
<td>2 September 3-7</td>
<td>Exam 1 Methods and analysis from: Jagger et al An Overlapping Protein-Coding Region in Influenza A Virus Segment 3 Modulates the Host Response Introduction to Mega 7 and identifying and retrieving sequences from online databases</td>
</tr>
<tr>
<td>Date</td>
<td>Topic</td>
</tr>
<tr>
<td>------------</td>
<td>------------------------------------------------------------------------</td>
</tr>
<tr>
<td>September 3</td>
<td>Synonymous Virus Genome Recoding as a Tool to Impact Viral Fitness</td>
</tr>
<tr>
<td></td>
<td>Phendrix annotation HHPRED</td>
</tr>
<tr>
<td>September 4</td>
<td>DNA sequencing techniques</td>
</tr>
<tr>
<td></td>
<td>Phendrix annotation</td>
</tr>
<tr>
<td>September 5</td>
<td>Protein structure prediction</td>
</tr>
<tr>
<td></td>
<td>Phendrix annotation with PHYRE THMM</td>
</tr>
<tr>
<td>October 1</td>
<td>Identifying horizontal gene transfer</td>
</tr>
<tr>
<td></td>
<td>Dot plot analysis</td>
</tr>
<tr>
<td>October 7</td>
<td>The Unix environment</td>
</tr>
<tr>
<td></td>
<td>Installation of Linux virtual machine, HHMI software package, Phamerator</td>
</tr>
<tr>
<td>October 8</td>
<td>Phylogenetic methods, tree reconstruction</td>
</tr>
<tr>
<td></td>
<td>Building trees with influenza genes</td>
</tr>
<tr>
<td>October 9</td>
<td>Epigenetics and DNA methylation</td>
</tr>
<tr>
<td></td>
<td>Comparing Phendrix with other Gordonia viruses</td>
</tr>
<tr>
<td>October 10</td>
<td>Exam 2</td>
</tr>
<tr>
<td></td>
<td>Newbler, Consed and genome assembly</td>
</tr>
<tr>
<td>November 1</td>
<td>ChIP-chip, and ChIP-seq</td>
</tr>
<tr>
<td></td>
<td>Working with gene expression data</td>
</tr>
<tr>
<td>November 11</td>
<td>Bioinformatics of vaccine design</td>
</tr>
<tr>
<td></td>
<td>Analysis of reassortment events in Influenza A</td>
</tr>
<tr>
<td>November 12</td>
<td>Horizontal gene transfer and Cancer</td>
</tr>
<tr>
<td>November 13</td>
<td>Mass Spectroscopy theory and application</td>
</tr>
<tr>
<td></td>
<td>Analysis of Mass spec data from</td>
</tr>
<tr>
<td>November 14</td>
<td>Oral Presentations</td>
</tr>
<tr>
<td>November 15</td>
<td>Exam 3</td>
</tr>
<tr>
<td></td>
<td>Final preparation and submission of Phendrix Genome to NCBI</td>
</tr>
</tbody>
</table>

**Topic Learning Goals / Sample Learning Objectives (subject to change)**

Understand the role of computation in hypothesis driven discovery processes within the life sciences.

Explain the role of wet-bench techniques in verifying computational results in life science research.

Compare and contrast computer-based research with wet-lab research.

Read a scientific article and evaluate how bioinformatics methods were employed by the authors to explore a particular hypothesis.

Given a scientific question, develop a hypothesis and propose computational approaches that could be used to explore the hypothesis.

Evaluate the social, legal, and ethical implications of computational approaches to understanding biology.

Define the term algorithm.

**Explain the difference between a heuristic (approximate) algorithm and an exact algorithm.**
What statistical concepts are important in bioinformatics?
Perform elementary statistical analysis on an “omics” dataset (e.g. using Excel or Weka).
Interpret the biological significance of an e-value.

Where are data about the genome found (e.g., nucleotide sequence, epigenomics) and how are they stored and accessed?
Describe how nucleotide sequence data are represented (FASTA, FASTQ, GenBank).
Describe the nucleotide databases available at NCBI.
Describe how protein sequence data are represented (e.g., FASTA, GenBank, etc.)
Describe the different protein databases available at NCBI (sequence, structure, function).
Describe how the NCBI nucleotide databases intersect with other nucleotide databases (EBI, DDBJ, UniProt, etc.).
Compare and contrast the data contained in different nucleotide databases.
Search for a sequence record in a nucleotide database with a given accession number.
Create a collection of nucleotide sequence records that meet a specified criterion (e.g., gene name or symbol).

Determine the DNA methylation state of a particular region of a genome.
Describe the types of metadata that accompany sequence data to make for useful biological interpretation (e.g. biological source, accession number, GeneID, journal articles, etc.).

How can bioinformatics tools be employed to analyze genetic information?
Calculate the alignment score between two DNA sequences using a provided scoring matrix.
Perform a BLASTN search and interpret the results.
Create and interpret a multiple sequence alignment (e.g., T-COFFEE, MUSCLE, etc.).

For a genomic region of interest (e.g., the neighborhood of a particular gene), use a genome browser to view nearby genes, transcription factor binding regions, epigenetic information, etc.
Describe how the NCBI databases intersect with other databases (e.g., EBI, DDBJ, UniProt, etc.).
Compare and contrast data contained in different databases (e.g., EBI, DDBJ, UniProt, etc.).
Search for a protein record in a database with a given accession number.
Create a collection of records that meet a specified criterion (e.g., gene name or symbol).
Describe the types of metadata that accompany sequence, structure, and function data to make useful biological interpretation (e.g. biological source, accession number, UniProt number, journal articles, etc.).

Define the BLASTP, BLASTX, tBLASTn, tBLASTx algorithms for protein sequence information
How can bioinformatics tools be employed to examine protein structure and function?
Query a dataset with a specific protein sequence to learn about potential functions (e.g. Pfam, CDD, SwissProt, UniProt, etc.).

View and interpret the structure output from Protein Data Bank (e.g. Cn3D, Jmol, etc.).
Propose potential functions for a give protein structure.

Explain the outputs from protein-folding algorithms to predict structure from sequence.
Create and interpret a multiple sequence alignment (e.g., T-COFFEE, MUSCLE, etc.).

Describe the components of a phylogenetic tree (e.g., root, node, leaf).
Explain the various types of phylogenetic trees (e.g., rooted, unrooted).
Interpret a phylogenetic tree (e.g., which organism is most closely related to a given organism in the tree)
Use bootstrapping to assess the quality of a phylogenetic tree.
Create a phylogenetic tree for a set of related sequences {nucleotide or amino acid} (e.g., MEGA).
Use a phylogenetic tree to address a research question.