NORTHERN ARIZONA UNIVERSITY
College of Engineering, Forestry & Natural Sciences
DEPARTMENT OF ELECTRICAL ENGINEERING & COMPUTER SCIENCE

SYLLABUS FOR CS299

INTRODUCTION TO BIOINFORMATICS

General Information
- Fall 2011
- Tu/Th 2:20-3:35, 3 credit hours
- Professor: Dr. Gregory Caporaso (gregcaporaso+cs299@gmail.com)
- Office location: Engineering (building 69), Room 265
- Office hours: Tu 12:30-2pm, Th 4-5pm, and by appointment.
- Course website: http://www.cefns.nau.edu/~jgc53/cs299

Course Prerequisites
CS 126 or instructor consent.
Ideal prerequisites include introductory biology, introductory statistics, and introductory programming courses, but the course will cover the necessary material in these areas.

Course Description
Introduces computer scientists or biologists to Bioinformatics. This course will introduce python programming, relevant topics in biology, and tools and algorithms used in bioinformatics with the aim of teaching students how to develop their own bioinformatics software and pipelines to address biological questions. The course will be project oriented, giving students the tools to contribute to the many open questions in the field.

Topics covered will include python programming, biological sequence alignment, phylogeny construction, sequence clustering, working with "Next Generation" (i.e., 454 and Illumina) DNA sequencing data, cloud and cluster computing, and microbial ecology. Students will learn to use tools including BLAST, muscle, RAxML, uclust, and QIIME, but the focus of the course will be on the development of new tools or pipelines that make use of or extend these tools.

Student Learning Expectations/Outcomes for this Course
After taking this course, students should be able to develop small programs written in python, understand python code developed by others and make use of it in their own software, and have a basic understanding of the PyCogent bioinformatics toolkit. Students will also be able to find, install, and use existing bioinformatics software as well as critically evaluate its utility.

Course Structure/Approach
The course will begin with lectures covering basic concepts in molecular biology and python programming, and homework assignments will be focused on learning the Python programming language. These programming assignments will not be graded, but surprise quizzes will be given at random times and you will be expected to keep a notebook of your programming experiences which I'll collect (digitally) at random times. If you do the homework and the assigned reading the quizzes will be very easy (you may even recognize some of the questions!). Toward the middle of the semester we will begin having graded homework assignments and student presentations on existing bioinformatics software for 15 minutes. Students will work in groups on these

Revised 08/29/11

projects, with groups chosen by the instructor. Graded homework assignments will be done independently, and will primarily be programming tasks. Two exams will be given over the course of the semester: a midterm and a final exam.

Homework assignments and reading should be completed by the due date provided on the course schedule. Quizzing on the assignments due the same day is fair game.

Come to office hours! This course covers a lot of diverse material, and office hours are your chance to interact with me directly to ensure that you understand the concepts being covered in class. Coming to office hours is also a great way to boost your class participation score.

Textbook and Required Materials
The Processes of Life (ISBN: 978-0262013055)
This books covers basic biological concepts that we'll hit during the course. You can find chapter 1 online for free so you can keep up on the reading if you're waiting for a hard copy of the text. Chapter 1 is available here: http://mitpress.mit.edu/books/chapters/0262013053chap1.pdf.

Learn Python the Hard Way, 2nd edition (http://learnpythonthehardway.org/)
We will use this book extensively for learning python. There is a free html version, as well as a PDF ($2.99), and a hard copy ($15.99 paperback; $60 hardcover). If you choose to purchase this book, be sure to get the second edition.

Recommended References
Molecular Biology of the Cell (ISBN: 978-0815341055)
This book is an extremely useful reference. If you're planning to take additional biology classes, or if you'll continue to pursue Bioinformatics, a copy of this book will be invaluable. We may reference some sections of this text that are available for free online. There is also an e-book available which you can buy, rent, or buy by the chapter – you can get information here: http://store.vitalsource.com/show/9781136844423.

Links to other useful resources are available here:
http://www.cefns.nau.edu/~jgc53/cs299/books_and_links.html

Course Outline
http://bit.ly/mOASpI
A tentative course outline is provided at the link above. It is very likely that this outline will change over the course of the semester as we identify certain areas that we'll need to spend additional time on – for that reason the outline is provided in a dynamic format. The spreadsheet at the above link will always be the definitive source for the outline, reading assignments, and homework assignments.

Homework assignments listed as "LPTHW exercises …" refer to exercises in the *Learning Python the Hard Way* text. You can find these at this link: http://learnpythonthehardway.org/book/. Readings listed as "Hunter…" refer to sections in *The Processes of Life*.

Assessment of Student Learning Outcomes
- You will be assessed primarily thorough graded material (see *Grading System*). A small portion of your assessment will be based on participation. This allows me to determine if you understand the material well enough to ask questions and engage in discussion in class and in office hours.
- I will strive to return all graded materials within seven days, and hopefully sooner. Under some circumstances I may need a little more time. Check with me at any time during the semester on your current participation score.

Revised 08/29/11

Grading System
- 5% Participation
- 20% Surprise quizzes and programming notebook; lowest quiz score will be dropped
- 5% Application presentation
- 20% Graded homeworks; lowest homework score will be dropped
- 25% Midterm exam
- 25% Final exam

Grading Scale
> A: 90-100
> B: 80-90
> C: 70-80
> D: 60-70
> F: 0-59

Course Policy
- If you have a legitimate reason why you cannot make one of the exam times or why you won't be able to complete an assignment by the due date, talk with me early on and something can be arranged. Students who do not show up for exams without making prior arrangements will get a zero.
- Students are not graded on attendance, but there will be no make-ups of surprise quizzes so it is highly recommended that you plan to attend all lectures. If you have to miss a lecture and there is a quiz that day, it will count as your dropped quiz score.
- Plagiarism and cheating will not be tolerated. Any students found guilty of either will receive a failing grade in the class. We will discuss what counts as plagiarism of software on the first day of the course.
- No computers, cell phones, headphones, books or papers may be used during quizzes or exams. If a student is observed looking at or touching any of these items during a quiz or exam it will be considered cheating – this includes checking to see who just texted you! To be safe, I recommend turning off cell phones during exams.
- Computers may be used for course-related work during the class such note taking or working along with programming examples that are being done on the overhead. You may not however use computers for non-course-related work such as checking e-mail or Facebook. If I notice you doing this I will document it and count it against your participation grade in the course.

University Policies
The Safe Environment, Students with Disabilities, Institutional Review Board, Academic Integrity, Academic Contact Hour, Classroom Management and Professional Ethics and Code of Conduct policies are available at http://www4.nau.edu/avpaa/policy1.html. Students are responsible for reviewing and understanding these policies.