Statistics 590:
Introductory Bioinformatics (3 credit hours)

NC State University
Fall Semester 2017

Mondays and Wednesdays 3-4:15 PM, Daniels 214

Course Syllabus

INSTRUCTOR	OFFICE HRS
Dr. Gavin Conant (gconant@ncsu.edu) Assoc. Prof. Biological Sciences 356 Ricks Hall	By appointment

---

Course Content

1. **Course Description:** Almost every aspect of modern biology involves large-scale datasets and computational analyses. In this course, we will cover some of the basic theoretical and practice background needed to understand and use computational tools for biological analyses. The course will feature a mixture of lecture, activity-based and hand-on computational analyses using the LINUX operating system.

2. **Course Objectives:**
Upon completion of this course, students will meet goals in:

   I. Integrative skills:
      i. Explain the different ways in which computing is used in modern biology
      ii. Differentiate between computing approaches that automate task, perform statistical analyses, make evolutionary inferences and others

   II. Conceptual skills in bioinformatics:
      i. Define biological homology, orthology and paralogy
      ii. Explain the factors that make genome assembly a challenging problem
      iii. Explain the basic algorithm and assumptions of pairwise sequence alignment, including understanding the difference between local

and global alignment and the difference between exact and heuristic algorithms for alignment

iv. Discuss various methods of phylogenetic analysis and explain their assumptions and trade-offs, such as the difference between algorithms for tree inference and optimality criteria for tree search

v. Understand the concept of a biology network and explain why this concept represents an abstraction

III. Technical skills for carrying out computational and lab-based analyses in molecular evolution, including the:

i. Operation of basic sequence assembly software

ii. Performance of sequence database searches with BLAST

iii. Construction of multiple sequence alignments

iv. Calculation of diversity indices including evolutionary distances and measures of nonsynonymous and synonymous divergence in protein-coding sequences

v. Use of command-line tools for maximum likelihood and maximum parsimony phylogenetic analyses

vi. Limited script creation in perl

3. Suggested Textbook: Li and Graur – *Fundamentals of Molecular Evolution*. (**NOT REQUIRED)** All required reading will come from assigned scientific articles (posted on Moodle and noted during lectures).

Prerequisites

None: A background in biology or computer science is recommended

---

Assignments and Grading

Grading:

Exams (two, 21% each)
6 homework assignments (48%)
Class participation/In-class exercises (10%; lowest dropped)

All components of the class are required to pass (i.e., you must take the exams, complete the homework assignments and participate in class).

Late assignments: All assignments are due at the beginning of class on the date indicated. Late assignments will be penalized 15% for each 48 hours they are

late. (Thus the minimum penalty for a late assignment will be 15% if it is turned in before the beginning of class the day after it is due).

Grading Scale. Possible grades will be A, B, C, or F.

Class Policies

Computer Resources

Course Moodle page: https://wolfware.ncsu.edu

The computational assignments for the class will be available through the NCSU **Virtual Computing Laboratory** (VCL). We will cover using this system in class. You will need access to a computer to complete your assignments (Windows/Macintosh/LINUX are all acceptable), and I strongly recommend bringing a laptop to class for all class meetings.

Attendance

Attendance is an important part of the class and counts toward your participation grade. You will be responsible for obtaining notes for any unexcused absences from a classmate. There will be no make up exams for any unexcused absences.

Academic Integrity

All exams are individual assignments, unless otherwise stated in writing. Evidence of cheating, plagiarism, or other violations of the Code of Student Conduct will be investigated and, if appropriate, referred to the Office of Student Conduct for disciplinary review. You are free to collaborate on your in-class assignments, but each student must submit her or his own unique version for credit.

The Code of Student Conduct can be found at:
http://www.ncsu.edu/policies/student_services/student_discipline/POL11.35.1.php

Information about policies and procedures of the Office of Student Conduct can be found at:
http://www2.ncsu.edu/student_affairs/osc/

Inclement Weather

The class will follow the University's closure policy. If classes are not cancelled, I will make every effort to be in class on time, and so should you. Please do not send me email

asking whether class is going to meet. Instead, check the University website or the weather hotline (513-8888).

Students with Disabilities

Reasonable accommodations will be made for students with verifiable disabilities. In order to take advantage of available accommodations, students must register with Disability Services for Students at 1900 Student Health Center, Campus Box 7509, 515-7653. http://www.ncsu.edu/provost/offices/affirm_action/dss/

For more information on NC State's policy on working with students with disabilities, please see: http://www.ncsu.edu/provost/hat/current/appendix/appen_k.html

Schedule of Topics

The dates and topics covered may change given student needs and progress. Homework and assignment due dates *will not* change.

Date	Topic	To do
Week 1 August 16,	• Introduction to biological information • Goals of bioinformatics • Setup & test VCL	• Bring Laptop to class
Week 2 August 21, August 23	• Using LINUX • Discuss genetic drift • Introduction to Perl • Modeling drift with PERL	• Submit final drift perl script to GCC Wednesday (Ex#1)
Week 3 August 28, August 30	• Genome assembly exercise • Intro. to genome assembly • Simulated assembly example • Biological Homology	• For Monday: Read Nagarajan and Pop, 2013 • For Wednesday: ○ Read Matias Rodrigues &Wagner 2009 ○ Scan Prigambada et al., 1995 • On Wednesday: Hand in homology exercise (Ex#2)

Week 4 September 6	• Sequence Alignment • Example pairwise alignments	• **Wednesday: Homework #1 due** • Submit local and global Alignment scores (Ex#3)
Week 5 September 11, September 13	• BLAST and database search • E-values and scoring matrices • Start Tree thinking and phylogenetics	• For Wednesday: Read Baum et al., 2005
Week 6 September 18, September 20	• Example phylogenetic analysis with multiple alignment and tree inference • Orthologs and paralogs • BLAST exercise	• **Monday: Homework #2 due** • Submit tree file from example phylogenetic analysis (Ex#4) • For Wednesday: Read Koonin 2005
Week 7 September 25, September 27	• Modeling evolution, take 2 • Codon models, K_a & K_s, selection tests • Start analysis of gene expression	• Work on Homework #3! • Submit your FDR corrected P-values (Ex#5)
Week 8 October 2, October 4	• Analysis of gene expression • Example use of BowTie • Multiple testing/exercise	• **Wednesday: Homework #3 due** • Submit results of expression analysis (Ex#6)
Week 9 October 9, October 11	• **Monday: Exam #1 in class** • Variation/population genomics	• For Wednesday: Read Rosenburg et al., 2002
Week 10 October 16, October 18	• Variation/population genomics • Linkage analysis/GWAS	• **Monday: Homework #4 due** • Submit Genehunter results (Ex#7) • Wednesday, read:

		○ Klein et al, 2005 ○ Mackay et al., 2009
Week 11 October 23, October 25	• Databases	• Submit results of SQL query for maximum number of reactants per rxn (Ex#8)
Week 12 October 30, November 1	• Introduction to Networks • Protein interaction Networks	• For Monday: Read Zhu et al., 2007 • Submit the shortest path between your actor and Kevin Bacon (Ex#9)
Week 13 November 6, November 8	• Metabolic networks • Metabolic modeling	• **Monday: Homework #5 due** • For Wednesday: Read Kacser and Burns 1981 • Submit results of FBA analysis (Ex#10)
Week 14 November 13, November 15	• Metagenomics	• For Monday: Read Hess et al., 2011
Week 15 November 20	• Parallel Computing	
Week 16 November 27 November 29	• Parallel Computing (cont) • **Wednesday: Exam 2 in class**	• **Monday: Homework #6 due** • Submit calc_pi timings (Ex#11)