BIO 562:
Fundamentals of Bioinformatics
(3 credit hours)

NC State University
Fall Semester 2023

102 David Clark Labs
Tuesdays and Thursdays 1:30-2:45 PM,

Course Syllabus

INSTRUCTOR	OFFICE HRS
Dr. Gavin Conant (gconant@ncsu.edu) 919-515-8164 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 python

3. **Optional 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).

Course format

The course will comprise a mixture of lectures, group discussions and exercises (primarily of primary literature articles) and in-class computational assignments that the students will complete in parallel with the instructor. The exercises can be done in groups, but it is expected that each student will bring a laptop computer to access the required computational tools. (See **Grading** and **Attendance** below).

Prerequisites

None: A background in biology or computer science is recommended

Health and Well-Being Resources

These are difficult times, and academic and personal stress is a natural result. Everyone is encouraged to take care of themselves and their peers. If you need additional support, there are many resources on campus to help you:

- Counseling Center (https://counseling.dasa.ncsu.edu/)
- Health Center (https://healthypack.dasa.ncsu.edu/)
- If the personal behavior of a classmate concerns or worries you, either for the classmate's well-being or yours, we encourage you to report this behavior to the NC State CARES team: (https://advising.dasa.ncsu.edu/resources-for-advisors/advisors-toolkit/cares/)
- If you or someone you know are experiencing food, housing or financial insecurity, please see the Pack Essentials Program (https://dasa.ncsu.edu/pack-essentials/).

Technology Requirements

This course requires that you possess a computer and have internet access. Be sure to review the syllabus for these expectations, and see go.ncsu.edu/syllabus-tech-requirements to find out more about technical requirements for your course. If you need access to additional technological support, please contact the Libraries' Technology Lending Service: https://www.lib.ncsu.edu/devices.

Assignments and Grading

Grading:

Exams (two, 21% each)
6 homework assignments (48%)
"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: Homework assignments and in-class exercises are due at the date indicated by the instructor and in Moodle. 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 1 minute after it is due). Homework and in-class assignments will be submitted through Moodle (with the exception of one "on-paper" assignment due at the end of the class where it is completed.

The "in-class" exercises will require you to attend scheduled class meetings. As such, these exercises will be due 1 to 2 days after the class in which we complete them. I will announce each deadline in these meetings and post them to Moodle. These exercises will be submitted through Moodle.

Exam format. Exams will be conducted with Moodle at the times and dates below. You will be expected to take the exam ***in person in class*** unless you have made prior arrangements with me. **Taking the exam outside of our designated classroom (e.g., at home or elsewhere via Moodle) is grounds for failure unless you have received prior written permission from me.**

Missed exams: **If you miss an exam without an excused absence you will be able to make it up by taking the optional final exam.** This exam will occur at the time scheduled for the course final exam, but will constitute a third exam of a similar length and difficulty to the first two, save that it will cover material from the entire course. Exams will be conducted online through Moodle.

Grading Scale. Possible grades will be A, B, C, or F. Cutoffs for each letter grade will be no higher than 90% (A), 80% (B), and 70% (C): I may choose to adjust these cutoffs down if needed.

Class Policies

Health policies.

Students should stay current with campus public health measures through the Protect the Pack website (https://www.ncsu.edu/coronavirus/). The sections below provide expectations and conduct related to these measures.

Health and Participation in Class

We are most concerned about your health and the health of your classmates and instructors/TAs.

- If you test positive for COVID-19, or are told by a healthcare provider that you are presumed positive for the virus, you should not attend any hybrid or face-to-face (F2F) classes and work with your instructor on any adjustments necessary; also follow other university guidelines,.
- If you feel unwell, even if you have not been knowingly exposed to COVID-19, please do not come to a F2F class or activity.
- If you are in quarantine, have been notified that you may have been exposed to COVID-19, or have a personal or family situation related to COVID-19 that prevents you from attending this course in person (or synchronously), please connect with your instructor to make alternative plans, as necessary.

- If you need to make a request for an academic consideration related to COVID-19, such as a discussion about possible options for remote learning, please talk with your instructor.

If applicable, university policies regarding mask use will be enforced in this course. You are more than welcome to wear a mask at all times regardless of these rules if you feel more comfortable doing so.

Class meetings

Attendance is expected: in-case exercises will require you to attend the lectures in question in order to complete them. *I do not view the livestream/recordings as a general replacement for attending class unless you are ill or have contacted me in advance.* The lowest exercise will be dropped. It is permissible to work with another student to complete and submit a missed in-class assignment. If for any reason you are not comfortable attending class in person, please contact me in advance so that we can make arrangements.

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. Chromebooks are **not** compatible with the VCL).

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 and homework, 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. 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. Exam dates *will not* change, but homework and exercise due dates may moved *later*. (In other words, the dates below are the earliest possible due dates for the homework and exercises.)

Date	Topic	To do
Week 1 August 22, 24	• Introduction to biological information • Goals of bioinformatics • Setup & test VCL/ Using LINUX • Discuss genetic drift • Introduction to Python • Modeling drift with Python	• Learn to use PuTTy and the VCL with your computer
Week 2 August 29, 31	• Simulating drift (continued) • Genome assembly exercise • Intro. to genome assembly • Simulated assembly example	• August 29th Submit final drift python script to GCC (*Exercise #1*) • For August 29th: Read Nagarajan and Pop, 2013

Week 3
September 5,
September 7
	• Biological Homology
	• Sequence Alignment
	• Example pairwise alignments
	• BLAST and database search
	• E-values and scoring matrices
		• For September 5th:
		  ○ Read Matias Rodrigues &Wagner 2009
		  ○ Scan Prigambada et al., 1995
		• Sept 5th: Hand in homology exercise (*Exercise#2*)

Week 4 September 12, September 14	• Start Tree thinking and phylogenetics • Example phylogenetic analysis with multiple alignment and tree inference	• September 12th: Submit alignment scores (*Exercise#3*) • For September 12th: Read Baum et al., 2005 • For September 14th: Submit tree file from example phylogenetic analysis (*Exercise#4*)
Week 5 September 19, September 21,	• **September 19th: Wellness Day: NO CLASS** • Orthologs and paralogs • BLAST exercise	• **Sept. 19th: Homework #1 due** • For September 21st: Read Koonin 2005
Week 6 September 26, September 28	• Modeling evolution, take 2 • Codon models, K_a & K_s, selection tests • Start analysis of gene expression	• **September 26th: Homework #2 due** • Work on Homework #3! • For September 28th: Read Marioni et al., Gen Res 2008
Week 7 October 3, October 5	• Example use of BowTie • Differential expression • Multiple testing/exercise • Variation/population genomics • Linkage analysis/GWAS	• October 3rd: Submit your FDR corrected P-values (*Exercise#5)* • October 5th: Submit results of expression analysis (*Exercise#6*) • For October 3th: Read Rosenburg et

		al., 2002
		• For October 5th: read:
		  ○ Klein et al, 2005
		  ○ Mackay et al., 2009

Week 8 October 12	• **Fall Break: No classes on October 10th** • **Thursday October 12: Exam #1 1:30-2:45**	
Week 9 October 17&19	• Finish Population genetics/GWAS • Databases	• For October 17th: Submit Genehunter results (*Exercise#7*) • October 19th: Submit results of SQL query for maximum number of reactants per rxn (*Exercise#8*) • **October 19th: Homework #3 due**
Week 10 October 24&26	• Introduction to Networks • Protein interaction Networks • Metabolic networks	• For October 24th: Read Zhu et al., 2007 • October 24th: Submit the shortest path between your actor and Kevin Bacon (*Exercise#9*)
Week 11 October 31st & November 2nd	• Metabolic modeling	• For October 31st: Read Kacser and Burns 1981 • **November 2nd: Homework #4 due** • November 2nd: Submit results of FBA analysis (*Exercise#10*)

Week 12 November 7&9	• Metagenomics	• For November 7th: Read Hess et al., 2011
Week 13 November 14&16	• Parallel Computing	• November 16th: Submit calc_pi timings (*Exercise#11*) • **November 16th: Homework #5 due**
Week 14 November 21	• Protein structure prediction • **No class Nov. 24: THANKSGIVING!**	
Week 14 November 28, 30	• Topics arising	
Week 15 December 5th	• **Exam 2: December 5th.**	• **Tuesday December 5th: Homework #6 due**
FINALS	• **Optional Final exam, 12:00-2:30, December 12th.**	•