ST 591: Introduction to Quantitative Genomics
Syllabus

General Information
Instructor: Thomas Sharpton
Email: thomas.sharpton@oregonstate.edu
Office: 530 Nash Hall
Phone: (541) 737-8623
Office Hours: TBD
Teaching Assistant: TBD

Course credits: 3

Class meetings:
Meeting Time: Tuesday & Thursday 10 - 11:20 am
Location: TBD

Enforced Prerequisites
None

Recommended Prerequisites
ST 411/511 or an equivalent

Course Description:
This is a lecture and discussion course that provides an overview of how genomic data is generated and analyzed. It focuses on the underlying biological motivation, theoretical concepts, and analytical challenges associated with genomic research, especially the generation of statistics that summarize genomic data. The class is organized as a combination of lectures and group literature review discussions. Students are expected to actively participate in the class. Students from diverse backgrounds, including quantitative, biological, and computations sciences, are encouraged to enroll. *This is a companion course to “Statistical Methods for Genomics Research” (ST 591, Winter in Odd Years, Jiang).*

Learning Outcomes

Upon completing this course, students should be able to:

1. Demonstrate an understanding of the language and terms associated with genomic research
2. Describe genomic methods that solve specific biological problems and their caveats
3. Identify appropriate analytical methods that can be used to quantify specific genomic signals and describe their caveats
4. Demonstrate an understanding of the basic statistical concepts underlying with these methods (e.g., the role of stochasticity, noise, bias, variance)
5. Define data structures relevant to genomic analysis

Organizational Format

This course is composed of rapid-paced lectures that cover a diverse array of topics as well as class discussions on landmark and cutting-edge scientific literature that is pertinent to course material.

This class will begin with a high-level introduction to the biological and statistical concepts needed understand the subsequent lectures and participate in discussions. Then, the class will focus on topical modules in quantitative genomics (e.g., Genomes, Population and Comparative Genomics, Functional Genomics, and Metagenomics). For each module, the instructor will provide lectures that clarify the topic’s biological background and motivation as well as relevant quantitative concepts and challenges. Modules will end with a
class review and discussion of research manuscripts. **Students are expected to actively participate in these discussions and will be responsible for leading at least one discussion topic over the course of the quarter.**

**Assessments, Assignments, and Examinations**

**Written assessments.** Students will be expected to complete assessments of their learning comprehension. *These mandatory assessments will not be graded, but will contribute to the student’s class participation score.*

**Journal clubs, Literature reports, and Discussion leadership.** At the end of each module, the class will collectively discuss two scientific papers during a class ‘journal club’. To facilitate discussion and peer interaction, students will form small groups (4-5 students) at the start of the term that should meet periodically through the term to review these papers prior to the class discussion. Upon arriving to journal club classes, *each group will hand in a one-page summary of one of the journal club papers assigned during that module.* This summary should briefly articulate the biological question being investigated, the genomic methodology used to conduct the analysis, the conclusions and interpretation of the work and, most importantly, limitations and caveats of the analysis (even if not identified by the authors). These literature reports will contribute to each student's grade.

Participation class journal club discussions is expected and will contribute to the class participation grade. *Each group will be expected to lead one class discussion on one paper over the course of the class. Groups are not expected to submit a literature report during those journals where they are leading a discussion.*

**Examinations.** Students will complete a midterm and final examination. These tests will be open book, but students are expected to work independently on these exams.

**Grading**

Participation (class discussions and weekly assessments): 30%
Literature Reports: 20%
Midterm Examination: 25%
Final Examination: 25%

**Disability Access Services**

Accommodations for students with disabilities are determined and approved by Disability Access Services (DAS). If you, as a student, believe you are eligible for accommodations but have not obtained approval please contact DAS immediately at 541-737-4098 or at http://ds.oregonstate.edu. DAS notifies students and faculty members of approved academic accommodations and coordinates implementation of those accommodations. While not required, students and faculty members are encouraged to discuss details of the implementation of individual accommodations.

**Academic Honesty**

The students are expected to be honest and ethical in their academic work. Please read the text about the University Policy through the following link: http://studentlife.oregonstate.edu/sites/studentlife.oregonstate.edu/files/final_code_of_student_conduct_update_d_1_8_18.pdf

Any incident of academic dishonesty will be handled according to the University's Academic Regulations (http://catalog.oregonstate.edu/ChapterDetail.aspx?key=75#Section2883).

**Course Content**

**Module: Background**

Class 1: Introduction to Genome Biology
Suggested reading:
- Molecular biology for bioinformatics – a very short primer

If you need extra help:
- Molecular Biology Primer
- Preparata, A Biology Primer for Computer Scientists
- Zien, A Primer on Molecular Biology

For fun:
- DNA seen through the eyes of a coder (http://ds9a.nl/amazing-dna/)

Class 2: Introduction to Quantitation
Suggested reading:
- Hon, An Introduction to Statistics

If you are interested in R:
- Seefeld & Linder, Statistics Using R with Biological Examples

Class 3: Molecular evolution and sequence alignment
Suggested reading:
- Chapter 2 of Durbin, Eddy, Krogh, Mitchison Biological Sequence Analysis

Class 4: DNA Sequencing, Error, and Quality Control
Suggested reading:
- Metzker, 2010, Sequencing technologies— the next generation.

Class 5: Journal Club
1-page literature report due
Required reading:
- Ross (2013) Characterizing and measuring bias in sequence data
- Lou (2013) High-throughput DNA sequencing errors are reduced by orders of magnitude using circle sequencing
- Buchfink (2015) Fast and sensitive protein alignment using DIAMOND

Module: Genomes

Class 6: Genome Assembly
Suggested reading:
- Compeau et al. (2011). How to apply de Bruijn graphs to genome assembly
- Nagarajan, 2009, Sequence assembly demystified

Class 7: Genome Annotation
Suggested reading:
- Stein (2001) Genome annotation: from sequence to biology

Class 8: Protein Annotation
- Fitch (2000) Homology: a personal view on some of the problems
- Friedberg (2006) Automated protein function prediction – the genomic challenge

Class 9: Journal Club
1-page literature report due
Required reading:
- Price (2016) Deep annotation of protein function across diverse bacteria from mutatnt phenotypes
Midterm Exam

Class 10: Midterm Exam
The midterm exam is a take-home, open-book exam due by the end of the class period. Students will have the class period to work on it.

Module: Comparative and Population Genomics

Class 11: Population genomics
Suggested reading:
• Lulkart et al (2003) *The power and promise of population genomics: from genotyping to genome typing*

Class 12: Comparative Genomics
Suggested reading:
• Koonin et al (2003) *Comparative genomics, minimal gene-sets and the last universal common ancestor*

Class 13: Genome Wide Association Studies (Guest Lecture by Dr. Jiang)
Suggested reading:
• Frayling (2014) *Genome-wide association studies: the good, the bad and the ugly*

Class 14: Journal Club
1-page literature report due
Required reading:
• Ellison (2011) *Population genomics and local adaptation in wild isolates of a model microbial eukaryote*
• Zhang (2014) *Comparative genomics reveals insights into avian genome evolution and adaptation*

Module: Functional Genomics

Class 15: Inferring Function from Conservation
Suggested reading:

Class 16: Transcriptomics
Suggested reading:
• Wang (2009) *RNA-Seq: a revolutionary tool for transcriptomics*

Class 17: Epigenetics and ChlP-Seq
Suggested reading:
• Furey (2012): ChIP-seq and beyond: new and improved methodologies to detect and characterize protein-DNA interactions

Class 18: Journal Club
1-page literature report due
Required reading:
• Lindblad-Toh (2011) *A high-resolution map of human evolutionary constraint using 29 mammals*
• Bruno (2010) *Comprehensive annotation of the transcriptome of the human fungal pathogen Candida albicans using RNA-seq*
• Zhang (2012) *Epigenetic variation creates potential for evolution of plant phenotypic plasticity*

No class on Thanksgiving Holiday

Module: Metagenomics

Class 19: Metagenomics
Suggested reading:
- Tringe & Rubin (2005) Metagenomics: DNA sequencing of environmental samples
- Sharpton (2014) An introduction to the analysis of shotgun metagenomic data
- The Human Microbiome Project Consortium (2012) Structure, function, and diversity of the healthy human microbiome
- Bikel S (2015) Combining metagenomics, metatranscriptomics, and viromics to explore novel microbial interactions: towards a systems-level understanding of the human microbiome

Class 20: Journal Club
1-page literature report due

Required reading:
- Howe (2013) Tackling soil diversity with the assembly of large, complex metagenomes
- Callahan (2016) DADA2: High-resolution sample inference from Illumina amplicon data

Final Exam

*Finals Week: The final exam will be a take-home, open-book examination. It is due TBD by 11:59 pm.*