Bio 4342: General Course Information (Spring 2012)

Instructors
Sarah C R Elgin  131 McDonnell Hall  935-5348  selgin@biology.wustl.edu
Elaine Mardis  Genome Center  286-1805  emardis@watson.wustl.edu
Jeremy Buhler  Jolley Hall 530  935-6180  jbuhler@cse.wustl.edu
Chris Shaffer  McDonnell 112 (Danforth)  935-6837  shaffer@biology.wustl.edu

Teaching Assistants
Wilson Leung  McDonnell 112 (Danforth)  935-6837  wleung@artsci.wustl.edu
Lee Trani, Holly Kotkiewicz, Jenn Hodges, Ian Hartzel (sequence improvement)  Genome Center
Katie Geist  katherine.geist@wustl.edu
Suchita Rastogi, Hashini Gunasinghe
Sarah Chavez (Writing Instructor)  sarahvanvicklechavez@wustl.edu

Class Schedule
Lecture and lab will function together. The class will meet from 1:30 to 5:00 PM on Monday and Wednesday, and from 1:30 to 2:30 PM on Friday; occasionally the Friday session will extend to 3:30. Students who elect the writing-intensive option (Bio 434W) will have ca. 5 additional hour-long meetings to focus on writing, scheduled for Friday 2:30-3:30 PM. Attendance is required. Because this is a laboratory course, true make-up sessions are often not possible. Students who must miss a class due to ill health, a death in the family, or a med school/grad school interview should inform Dr. Shaffer prior to the class session to obtain a bye. If you miss a class, you are responsible for obtaining notes and information from your lab partner, and consulting with the instructor and/or TA as necessary to gain an understanding of the material covered and catch up on your work as needed.

Meeting Sites
During the first six weeks of the semester, class will meet at the WU Genome Institute, 4444 Forest Park Parkway, with the exception of some Fridays (see schedule). The Center is ca. 2 blocks from the West End Metro stop. Class will meet in the 5th floor classroom. During the last eight weeks of the semester (starting March 1), class will meet in the Biology Department, Life Sciences 202, on the Danforth Campus.

Texts
There are no required texts. The following books may be useful, depending on your background. These books will be on reserve in the Biology Library.

“Bioinformatics and Functional Genomics” by R. Pevsner, 2009 (2nd ed.), J. Wiley & Sons, NJ (recommended for Bio majors if you would like more introduction to the computer tools we use).


“Genomes 3” by Terence A. Brown, 3rd ed, 2006, Garland Science (ISBN 0815341385) (recommended if no equivalent book, such as the texts for Bio 2970, are in hand).
In addition, you can access “Current Protocols in Bioinformatics” as a link from the Biology Library Webpage at [http://library.wustl.edu/subjects/life/ej/c.html](http://library.wustl.edu/subjects/life/ej/c.html).

**Web Site**
All course information, announcements, reading assignments, etc. will be posted on the Bio 4342 web site, maintained by the Biology Department through the NSLC. The latter portion of the web site is password protected. This has copies of all of the recommended and required reading. Most of the teaching materials used in the course can be found at the Genomics Education Partnership web page ([http://gep.wustl.edu](http://gep.wustl.edu)) under Curriculum. Examples of student papers from previous years are also found there.

**Student Responsibilities, Grading**
Grades will be assigned based on the following components: participation in discussions, summary papers on reading, 15%; five graded computer-based problem sets, 25%; report on finishing a Drosophila fosmid clone (written and oral), 25%; report on genes/pseudogenes, 10%; final report on individual Drosophila fosmid (analysis and annotation) (written and oral), 25%.

Students who elect the Writing Intensive version of the course will have an introductory writing assignment that constitutes 5% of the grade.

**Lab Overview: Sequencing / Finishing**
We anticipate either 1) video and/or a demonstration, where noted below, or 2) having students perform hands-on, the following types of activities during the first half of the semester:
- Direct sequencing techniques for plasmid DNA—both manual and automated (video)
- QC procedures—how/when they are performed and why this is important
- Looking at DNA sequence data—trace viewing and examining graphical representations of large data sets/comparative graphical views
- Finishing process—calling sequencing primers from fosmid or genomic DNA, adding additional data, making joins and editing; methods for assessing quality of finished sequence

**Lab Overview: Analysis / Annotation**
We anticipate that students will become familiar with the use of Phred/Phrap/Consed to assemble and evaluate sequence reads (see above); commonly used DNA databases; model organism websites; genome browsers; RepeatMasker; Genscan and other gene prediction tools; BLAST, BLAT searches for similarity; Clustal for comparative analysis. As time permits and the research dictates, we may explore other databases and comparative tools.

**Computers**
We will provide Mac laptops for your use during the course; you will be responsible for returning these in good condition at the end of the semester. If you prefer, you can use your own portable computer. However, we recommend that only Macs be used during our work on sequence improvement (first six weeks of the course), as Consed (the key software) is available only in a Mac version.
Bio 4342, Research Explorations in Genomics

Spring 2012 Draft

M, W 1:30-5:00; F 1:30-2:30 (occasionally 3:30), Writing Intensive group 2:30-3:30 when scheduled

Class meets at The Genome Center, Fourth Floor Conference room.

1/18 W  
Course structure, research problem (S. Elgin, 30 min).
Lecture: Overview of DNA sequencing (goals); the pipeline; different sequencing strategies (E. Mardis).
Lab: Intro to laptop computers, intro to Unix, hand out HW1 (C. Shaffer).

1/20 F  
Lecture: The D. ananassae fosmids (C Shaffer) (Danforth Campus, LS 202)
Lab (optional): Computer help session, HW1.

1/23 M  
Tour of the The Genome Center (watch Virtual Tour and Next Generation Technologies videos on line prior to class; be sure to wear long pants, closed shoes) (Cherilynn Shadding, 30 min.)
Lab: Introduction to Consed [read Guide to Consed (notebook)] (WU GC finishers Lee Trani).

1/25 W  
Lecture: The challenges in generating finished sequence (C. Shaffer; alumni)
Lab: Using Consed Graphically (inc. tear/assemble); hand out HW2 (Drosophila problems) (L Trani) **HW1 due.**

1/27 F  
Lab (**2 hr**) Work on HW2 (Finishers available).

1/30 M  
Lecture: Finishing/autofinish (L. Trani); alternative chemistries (C Shaffer).
Lab: **HW2 due.** Introduce finishing report requirements (including Finishing Checklist), obtain sequence file for own D. grimshawi or D ananassae fosmid, run Phred, Phrap, Consed, begin analysis (W. Leung; Finishers).

2/1 W  
Lab: Prefinishing own fosmid with help from GSC finishers available, must **order reads** today (use alternative chemistries). Discussion of Phred, Phrap, Consed and assessment of sequence quality (L. Trani)

2/3 F  
**WI subgroup:** discussion on scientific writing, peer review.

2/6 M  Lecture: Chromatin structure (J. Majors)
Lab: Finishing (NS, LT); student calls vs. Autofinish calls

2/8 W  Lab: Finishing own fosmid; receive new data, assemble; **order additional reads** today, GC finishers available for help.

2/10 F  Lecture: Heterochromatin / silencing (J. Majors) (Danforth campus, LS202).
**HW3 due. WI subgroup:** 1000 word intro paper due.

2/13 M  Lecture: Drosophila dot chromosomes, GEP findings (J Majors)
Lab: Finishing (L Trani, C. Shaffer) – consideration of restriction digests.

2/15 W  Lab: Receive additional data, assemble, finish. **Final day for ordering reads if needed.** Bio 4342 alumni (H. Yang and S. Spencer) present prototypes of report.

2/17 F  Lecture: modENCODE: a chromatin map of the *D. melanogaster* genome

**WI subgroup:** peer review of intro paper

2/20 M  Lecture: Using Genomics: Complex Traits (**B. Cohen**)
Lab: Preparation of reports; finishing checklist. Discussion of HW3 (J Majors)

2/22 W  Lab: Receive any additional data, assemble, finish. Final day for help from GC Finishers. Prepare reports, PowerPoint presentation.

2/24 F  Lecture: The Human Microbiome (**G. Weinstock**)

2/27 M  **Class reports** (10 min each) “Finishing my fosmid; problems identified and solved; comparing my analysis to the computer; meeting GSC standards.” Class meets in Life Sciences 202, Danforth Campus

2/29 W  Lecture/Demo: Review UNIX (C. Shaffer), introduce Goose; introduce BLAST (P. Gonzales).
Lab “Introduction to BLAST” (scripted walk-through)

3/2 F  Lab: Complete “Introduction to BLAST” **WI subgroup:** peer review of finishing papers.

3/5 M  Lecture: Gene finding: detecting and interpreting genetic homology (J Buhler) Lab: Begin CS HW1

3/7 W  Continue CS HW1 Lecture: Visualizing BLAST output and using EST’s (J
Buhler) Lab: Begin CS HW2. **CS HW1 due.**


**3/12 – 3/16 Washington University Spring Break**

3/19 M Lecture: Introduction to gene predictors. Gene prediction tutorial (train with chimp Contig95, Genscan) (C Shaffer)  
Lab: continue with gene finding lab as group  
**CS HW2 due**

3/21 W Lab: begin annotation of fragments of the chimp genome (CS HW3 data sets; work with partners)

3/23 F Lecture: New sequencing technologies and new explorations (**E Mardis**)  
Lab: optional help session.

3/26 M Lecture: Genes & pseudogenes in chimp; Bio 4342 alumni (H. Yang and J. Wong) present prototype reports  
Lecture: Introduction to *ab initio* gene finding, Drosophila browser, Flybase (W. Leung)

Lab: A Simple Annotation Problem (scripted walk-through)

3/30 F Lecture: Medical Sequencing (**R Wilson**)  

4/2 M **HW3 class reports** (10 min each) “Genes and pseudogenes in chimp.”  
Lab: begin work with own Drosophila fosmid to identify genes.

4/4 W Lecture: Introduction to multiple sequence alignments; use of Clustal (P. Gonzales)  
Lab: Clustal analysis of one feature (translated portion and regulatory element)

4/6 F Lecture: Finding regulatory elements in yeast (**B. Cohen**)  
**WI subgroup:** peer review of gene/pseudogene papers.

4/9 M Lecture: design and use of RepeatMasker; other ways to find repeats (J Buhler)  
Lab: Continue analysis and annotation of own fosmid, particularly repeats

Lab: Continue analysis and annotation of own fosmid
4/13 F  Lecture: The “Comsic Genome” Project (A. Templeton)  
**WI subgroup:** gene/pseudogene rewrites due.

4/16 M  Lecture: Preparation of annotation reports; H. Yang and J. Ma present prototypes  
Lab: Preparation of annotation reports; synteny

**Draft report on one gene due.**

4/18 W  Lab: Preparation of annotation reports.

4/20 F  Lecture: SNPs, indels, and rearrangements: diversity in human genomes (E Mardis)  
**WI subgroup:** peer review of annotation drafts.

4/23 M  Lecture: DNA methylation, another epigenetic mark (Ting Wang)  
Lab: Preparation of annotation reports; functions of one gene

4/25 W  Lab: Preparation of annotation report

4/27 F  Lab: Final consultations on annotation report

5/1 Tu  Final annotation reports: your fosmid, with a map of genes, repetitious elements, and alignment to *melanogaster*, including a Clustal analysis. Complete annotation of all exons; include a map and discussion of synteny. As time permits, investigation of the nature of the repetitious elements; search for noncoding conserved regions; search for endpoints of genes; search for regulatory elements; etc. 1:30 pm in LS 202.

5/2 W  **Course Assessment (time TBA):** Follow-up session on course evaluation  
(http://evals.wustl.edu, GEP web site, and Bio 4342 surveys/suggestions).

**Reading**

The scheduled papers will be provided on the class website. Reading these papers prior to the date specified will provide greater understanding of the lecture material and will help you to develop an understanding of the foundations of the experimental approaches we are using, as well as providing the antecedents for the scientific problem we are studying. You will not have time to read all of these papers in depth, but will be asked to read and summarize five papers over the course of the semester; these key papers are marked “R*”. Papers marked “R” are required background reading, but should not be used for your “reporting” assignments.

Background material on BLAST and other computer programs can be found in the recommended texts, and on-line through our subscription to “Current Protocols in Bioinformatics,” available at http://library.wustl.edu/subjects/life/ej/c.html.
Group #1


1/24 M Sequencing a Genome (view on line):

and “Next generation Sequencing: Genome Center Video Tour” (view all four segments online):
http://gep.wustl.edu/curriculum/course_materials/gsc_next.php R


Group #2


Mardis ER (2010) The $1,000 Genome, the $100,000 Analysis. Genome Medicine 2:84.

Group #4/#5


The following papers may be helpful during the second half of the course:


