

Bio 4342, Research Explorations in Genomics

Draft, Spring 2017

M, W 1:30-5:00; F 1:30-2:30 (occasionally 3:30); Writing Intensive group F 2:30-3:30 when scheduled. Meet in LS 311, Danforth Campus; on Friday 1/20 there will be a visit to the Washington University McDonnell Genome Institute. Please review our research problem (on Blackboard and on the course website at <http://www.nslc.wustl.edu/courses/Bio4342/bio4342.html>) and read "A Guide to Consed" on the GEP website at http://gep.wustl.edu/curriculum/course_materials_WU/sequence_improvement/tutorials_and_walkthroughs) prior to the first class.

- 1/18 Wed **Course structure**; research problem overview; assembly challenge (Elgin, 45 min).
Lab: Intro to computers; UNIX commands to get you started; begin work on *Using Consed Graphically* (navigation; adding reads) (Shaffer, Trani).
Lab: Pre-course assessment, GEP survey and quiz (45 min) (<http://gep.wustl.edu>).
Reading: Please watch Genome Center Virtual Tour and Next Generation Sequencing Video Tour online prior to class on Friday.
http://gep.wustl.edu/curriculum/course_materials_WU/introduction_to_genomics/
The associated worksheet (DiAngelo+SCRE, in folder) will be collected at the end of the tour on Friday.
- 1/20 Fri **Tour** of the McDonnell Genome Institute (be sure to wear long pants, closed shoes) (Cherilynn Shadding, Laura Courtney, MGI staff).
NOTE: meet in the elevator lobby on the Fourth Floor, 4444 Forest Park Parkway.
1:34 pm train from Skinker to Central West End Metro stop; tour starts at 2 pm.
Video Tour worksheets due.
- 1/23 Mon **Lecture**: Overview of DNA sequencing: the pipeline; different sequencing strategies; the basics for finishing *D. eugracilis* and *D. ficusphila* hybrid assemblies (50 min plus discussion) (Shaffer).
Lab: Complete *Using Consed Graphically*; start *GEP Hybrid Assembly Walkthrough* (using navigators, assessing quality, making corrections) (Trani) (HW#1, Consed).
Reading: "Sleeping dogs of the genome," Gorbunova et al; start Ellison & Bachtrog + commentary (RR#1 due 1/27).
- 1/25 Wed **Lecture**: The challenges in generating finished sequence; questions on HW#1 (Shaffer).
Lab: Continue *Finishing a Drosophila Hybrid Assembly* (corrections, resolving gaps, PCR primer design) (Trani, Shaffer).
Introduce reference materials and Finishing Report requirements (including *Finishing Checklist*), obtain sequence files for own *D. eugracilis* / *D. ficusphila* project, begin analysis (Leung; Trani, Shaffer). **HW#1 due** at end of class.

- 1/27 Fri **Discussion:** Ellison & Bachtrog; RR#1 due (Elgin, ~60 min.)
WI subgroup: 1000 word paper due (extended RR#1); discussion on scientific writing, peer review (~30 min) (Bednarski, Elgin).
Optional Lab (~1 hr): Work with Consed.
- 1/30 Mon **Lab:** Finishing own project with help from LT, CS, & WL as needed. Discussion of Miniassembly, Consed, and assessment of sequence quality; review of when to request additional sequencing, designing primers, as needed (Shaffer).
Suggested reading: Figures in Treangen & Salzberg, 2012.
Presentation: Bio 4342 alumni present examples of oral finishing report.
- 2/1 Wed **Lab meeting:** 9 min presentation + 3 min discussion each (individual reports to groups of five), "Finishing my project; problems identified and solved; remaining issues." Continue to finish own project based on feedback.
- 2/3 Fri **Lecture:** Eukaryotic genomes/chromatin structure (Elgin, 45' + discussion).
Optional help session (~1 hr).
WI subgroup: Critique of 1000 word paper due; discussion.
Reading: Eddy (2012) *The C-value paradox...*; and the modENCODE page on chromatin, including the vignette on fly chromatin
(<http://modencode.sciencemag.org/chromatin/introduction>)
- 2/6 Mon **Lecture:** Heterochromatin/euchromatin (Elgin, 45' + discussion).
Lab: Finishing own project: checklist for project submission; final day for consultation (LT, CS, & WL).
Reading: Start Haynes et al. (RR#2 due 2/15).
- 2/8 Wed **Lecture:** Chromatin states (modENCODE); dot chromosome, GEP findings (Elgin)
Lecture/Demo: *Basics of BLAST* (Kailong Mao).
Lab: *An Introduction to NCBI BLAST* (scripted walk-through).
WI subgroup: Rewrite of 1000 word paper due.
- 2/10 Fri **Lecture:** Introduction to web databases and FlyBase (Leung)
Reading: Webber & Ponting, 2004
Finishing papers due. Submit data files. (WI subgroup: exchange finishing papers.)
- 2/13 Mon **Lecture:** *Detecting and Interpreting Genetic Homology* (Buhler)
Lab: CS HW#1.
- 2/15 Wed **Lecture:** *Browser-Based Annotation and RNA-Seq Data* (Buhler)
Lab: CS HW#2.
Discussion: Haynes et al. (Elgin, 60 min) RR#2 due.

- 2/17 Fri **Guest Lecture:** Chris Miller: Cancer genetics
Reading: Wartman (2015) *A case of me*.
Lab: Optional help session on CS HW#1 and CS HW#2.
WI subgroup: peer review of finishing papers.
- 2/20 Mon **Lecture:** Efficient gene finding in *Drosophila* (Shaffer)
Lab: *Annotation of a Drosophila Gene* (scripted walk-through)
CS HW#1 and CS HW#2 due.
- 2/22 Wed **Lab:** continue *Annotation of a Drosophila Gene* (scripted walk-through).
Lecture: An introduction to the annotation projects (Elgin, 20')
Lab: Claim projects (Leung); begin work with own *Drosophila* project to identify genes (WL, CS, TAs).
- 2/24 Fri **Guest lecture:** Ting Wang: Insights from the human epigenome browser
WI subgroup: Rewrite of finishing papers due.
- 2/27 Mon **Lab:** Analysis of the first gene in your *Drosophila* project
Lecture/demo: Review of the Gene Model Checker (Leung)
- 3/1 Wed **Lab:** Gene annotation; check gene model by end of class.
Presentation: Reporting on a gene (Undergraduate TA's, 10')
- 3/3 Fri **Discussion:** Writing your paper on annotation of Gene #1 (due 3/10) (Elgin)
Lab: Optional help session on gene annotation, presentations.
- 3/6 Mon **Lab Meeting:** oral report on first gene (9' + 3' discussion; individual reports in groups of five) "Annotation of the first gene in my project."
- 3/8 Wed **Presentation/Discussion:** introduction to Leung et al (RR#3 due 3/25) (Elgin, 30')
Lab: Continue work on annotation, paper on gene #1.
- 3/10 Fri Optional help session
Paper due: "Annotation of the first gene in my project."
- 3/13 - 3/17 **Washington University Spring Break**
- 3/20 Mon **Lecture:** *Searching for Transcription Start Sites in Drosophila* (Leung).
Lab: Continue annotation project; check for regulatory elements of first gene.
- 3/22 Wed **Lecture:** *Introduction to Motifs and Motif Finding* (Buhler).
Lab: Continuing annotation project; motif hunting exercise (CS HW#3).
- 3/24 Fri **Discussion:** Leung et al 2015 (Elgin, Leung) RR#3 due.
WI subgroup: peer review of Annotation of Gene #1 papers.

- 3/27 Mon **Lecture:** Clustal Omega for annotation and motif finding (K Mao).
Lecture: Synteny.
Lab: Continue annotation; CS HW#3 due.
- 3/29 Wed **Guest Lecture:** Doug Chalker: RNAi.
Lab: Continuing annotation project; check for TSS, regulatory elements of genes.
- 3/31 Fri **Lecture:** *Hidden Markov Models* (K Mao); CS HW#4.
Reading: Eddy, S (2004a) What is a Hidden Markov Model?
WI subgroup: rewrite of Annotation of Gene #1 due.
- 4/3 Mon **Lab Meeting:** oral report on first Transcription Start Site (9' + 3' discussion; individual reports in groups of five).
Lab: Continuing annotation projects.
- 4/5 Wed **Lecture:** Targeting heterochromatin formation in *Drosophila* (Elgin).
Lab: Continuing annotation projects. CS HW#4 due.
- 4/7 Fri **Lecture:** RepeatMasker; other ways to find repeats (Buhler).
Paper due: first TSS annotation.
- 4/10 Mon **Lecture:** Finding repeats in *Drosophila* (Leung).
Lab: Continue analysis and annotation of own project.
Reading: start #4 (motif hunting, regulation, TBN); RR#4 due 4/21
- 4/12 Wed **Lecture:** Further characterization of genes/gene products: use of FlyBase resources; protein threading (K Mao).
Lab: Exploration of favorite gene, possible functions.
Presentation: Bio 4342 alumni present annotation project oral reports.
- 4/14 Fri **Guest lecture:** Marty Israel: Fragile X (provide feedback on Fragile X website).
WI subgroup: critiques of TSS annotation drafts due, discussion.
- 4/17 Mon **Lab:** Preparation of annotation reports; TSS determinations; functions of genes.
Guest lecture: John Edwards: The role of DNA methylation
- 4/19 Wed **Lab:** Preparation of annotation reports; TSS determinations; functions of genes
Lab Meeting: oral report: Further gene characterization (9' + 3' discussion; individual reports in groups of five).
- 4/21 Fri **Discussion:** Paper #4 (RR#4 due). Alumni visits?
WI subgroup: rewrite of TSS annotation due.

- 4/24 Mon **Lecture:** Preparing annotation files for project submission (Leung).
Lab: Preparation of annotation reports; start PPT preparation.
- 4/26 Wed **Lab:** Preparation of annotation reports.
- 4/28 Fri **Lab:** Final consultations on annotation report.
- 5/2 Tues **Final written and oral annotation reports:** Submit final paper on your project, with a map of genes (including estimates of transcription start sites), repetitious elements, and alignments to *D. melanogaster*, including a discussion of synteny. Complete annotation of all exons, all isoforms. Include results of searches for TSS candidate sites and regulatory elements. As time permits, exploration of one gene on FlyBase, expanding on gene features, regulation, and function, potentially including protein threading. Use Clustal at least once. 10' presentations (1 pm – 3 pm OR 3 pm – 5 pm in LS 311).
- 5/3 Wed Course Assessment: Follow-up session on course evaluation, submission of final files, return of computers, etc. (12 noon lunch – 2 pm, LS 311).
[\(http://evals.wustl.edu/;](http://evals.wustl.edu/) GEP web site; and Bio 4342 surveys/suggestions).

Reading

We will read and discuss four papers over the course of the semester, centered on the theme of genome organization and evolution in *Drosophila*, with an emphasis on the role of repetitious elements. [If you have not read scientific papers before, look at “How to Read a Scientific Paper” by Mary Williams, pp 1-5 (on the Bio 4342 website) before starting.] These papers are listed below; for each paper you will turn in a “reading reflection” (~2 pages, double-spaced, typed) that summarizes the big idea and proposes the next experiment. In addition, we have assembled a list of papers that are pertinent to the material we will be discussing, including papers recommended by our guest lecturers, on the Bio 4342 website (<http://www.nslc.wustl.edu/courses/Bio4342/elgin/4342cm.html>). Among these, papers marked “R” are highly recommended background reading. 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://onlinelibrary.wiley.com/book/10.1002/0471250953>. Background information on many scientific terms is available through the Genomics Education Partnership Glossary (http://community.gep.wustl.edu/repository/introduction_to_genomics/GEP_Glossary.pdf) and information on terms and techniques is available through Wikipedia (generally a good source, but be cautious!).

Required Reading (paper copies provided; reading reflections due):

1. Ellison CE, Bachtrog D. (2013) Dosage compensation via transposable element mediated rewiring of a regulatory network. *Science* 342: 846-50. (See also Chuong EB & Feschotte C (2013) Evolution: Transposons up the dosage. *Science* 342: 812-13.)
2. Haynes KA, Caudy AA, Collins L, Elgin SCR (2007) Element *1360* and RNAi components

contribute to HP1-dependent silencing of a pericentric reporter. *Curr Biol* 16: 2222-7. (See also Grewal & Elgin (2002) for background on concepts tested in this paper.)

3. Leung W et al. (2015) Muller F elements maintain a distinct set of genomic properties over 40 million years of evolution. *G3 Genes|Genomes|Genetics* 5: 719-40. Focus your review and experiment either on genome organization (figures 1-4, 9) OR on properties of genes (figures 5-8).

4. Alekseyenko AA, Peng S, Larschan E, Gorchakov AA, Lee OK, Kharchenko P, McGrath SD, Wang CI, Mardis ER, Park PJ, Kuroda MI (2008) A sequence motif within chromatin entry sites directs MSL establishment on the *Drosophila* X chromosome. *Cell* 134: 599-609. **(Tentative; TBN)**