Bio 4342, Research Explorations in Genomics

Draft, Spring 2016

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 1/22 there will be a visit to the Washington University McDonnell Genome Institute. Please review our research problem (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.

1/20 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 on line 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.

1/22 Fri  **Tour** of the Genome Institute (be sure to wear long pants, closed shoes) (Cherilynn Shadding, TGI staff)
NOTE: meet in the elevator lobby on the Fourth Floor, 4444 Forest Park Parkway (1:14 pm train from Skinker to Central West End Metro stop).

1/25 Mon  **Lecture:** Overview of DNA sequencing (goals); the pipeline; different sequencing strategies (E Mardis, 45 min plus discussion)
**Lecture:** The basics on finishing *D. ficusphila* hybrid assemblies (30 min) (Shaffer)
**Lab:** Complete Using Consed Graphically; start GEP Hybrid Assembly Walkthrough (using navigators, assessing quality, making corrections) (Shaffer, Trani). Assign HW#1 (Consed)
**Reading:** “Sleeping dogs of the genome,” Gorbunova et al; start Ellison & Bachtrog + commentary (RR due 1/29).

1/27 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 file for own *D. ficusphila* project, begin analysis (Leung; Trani, Shaffer).
1/29 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, SCRE).
**Optional Lab** (~1 hr): Work with Consed, HW1

2/1 Mon **Lab**: HW#1 due. Finishing own project with help from TGI finishers available. Discussion of Miniassembly, Consed, and assessment of sequence quality as needed. Review of when to request additional sequencing, designing primers, as needed (Shaffer).
**Suggested reading**: Figures in Treangen & Salzberg, 2012.
**Presentation**: Bio 4342 alumni present prototypes of oral finishing report.

2/3 Wed **Lab meeting**: 9 min presentation + 3 discussion each (individual reports in groups of five, one group starts at 1 pm) “Finishing my project; problems identified and solved; remaining issues.” Continue to finish own project based on feedback.

2/5 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)
**Optional homework**: Cot curve analysis.

2/8 Mon **Lecture**: Heterochromatin/euchromatin (Elgin, 45’ + discussion).
**Lab**: Finishing own project (TGI staff); checklist for project submission; final day for consultation
**Reading**: Start Haynes et al. (RR#2 due 2/13).

2/10 Wed **Lecture**: Chromatin states (modENCODE); dot chromosome, GEP findings (Elgin)
**Lecture/Demo**: Introducing BLAST (Yu He).
**Lab**: Simple Introduction to BLAST (scripted walk-through).
**Finishing papers due**: Submit data files. (WI subgroup: exchange finishing papers)

2/12 Fri **Lecture**: Introduction to web databases and FlyBase (Leung)
**WI subgroup**: Rewrite of 1000 word paper due.
**Reading**: Webber & Ponting, 2004

2/15 Mon **Lecture**: Gene finding: detecting and interpreting genetic homology (Buhler)
**Lab**: Begin CS HW1.
**Optional homework** on Cot curves due.

2/17 Wed **Lecture**: Browser-Based Annotation and RNA-Seq Data (Buhler)
**Lab**: Begin CS HW2. **CS HW1 due** by end of class.
**Guest lecture**: Ting Wang: Insights from the human epigenome browser
2/19 Fri **Discussion**: Haynes et al, Yandim et al (Elgin) RR#2 due.  
**WI subgroup**: peer review of finishing papers due.

2/22 Mon **Lecture**: Introduction to gene predictors (Shaffer).  
**Lab**: Gene prediction tutorial (train with chimp Contig95, Genscan); organize chimp project teams. **CS HW2 due**.

2/24 Wed **Guest lecture**: John Edwards: The role of DNA methylation  
**Lab**: continue with chimp gene finding lab with partners.  
**Presentation**: Bio 4342 alumni present chimp genome oral reports

2/26 Fri **Presentation**: Bio 4342 alumni present chimp genome oral reports  
**Lab**: Prepare lab meeting presentation with partners  
**WI subgroup**: Rewrite of finishing papers due.

2/29 Mon **Lab**: Annotation of fragments of the chimp genome.  
**Lab meeting**: oral presentations (12 min each group, starting ~3 pm) “Genes and pseudogenes in chimp.”

3/2 Wed **Lecture**: Introduction to *ab initio* and evidence-based gene finding (Leung)  
**Lab**: Complete work on chimp annotation.

3/4 Fri **Lecture**: Hidden Markov Models (Yu He)  
**Chimp analysis paper due** (WI exchange papers)  
**Reading**: Eddy, S (2004a) What is a Hidden Markov Model?

3/7 Mon **Lecture**: Cancer genetics (E Mardis)  
**Lab**: Hidden Markov Models exercise (CS HW3)  
**Discussion**: start Leung et al (RR #3 due 3/25); overview discussion (Elgin, 20’)

3/9 Wed **Lecture**: Dynamic programming (similarity searches) (Yu He)  
**Lab**: Begin Dynamic programming exercise (CS HW4)  
**Reading**: Eddy, S (2004b) What is dynamic programming?  
**WI subgroup**: peer review of Chimp paper due, discussion.

3/11 Fri Optional help session  
**CS HW3 due**.

3/14 - 3/18 **Washington University Spring Break**
3/21 Mon  **Lecture**: An introduction to the annotation projects (Elgin, 20’)
**Lecture**: Efficient gene finding in Drosophila (Shaffer)
**Lab**: Annotation of a Drosophila gene (scripted walk-through).
CS HW4 due.

3/23 Wed  **Lab**: claim projects (Leung); begin work with own Drosophila project to identify genes (Leung, Shaffer, TAs).

3/25 Fri  **Discussion**: Leung et al 2015 (Elgin, Leung); *RR #3 due*

3/28 Mon  **Lecture**: Review of gene model checker (Leung)
**Lab**: Gene annotation; check first gene model by end of class.
**WI subgroup**: Rewrite of chimp analysis due

3/30 Wed  **Lecture**: Searching for transcription start sites (Leung)
**Lab**: Continuing annotation project; check for regulatory elements of first gene

4/1 Fri   **Lecture**: Motif finding (J Buhler); motif hunting exercise (begin CS HW5)

4/4 Mon   **Lab Meeting**: oral report on first gene (9’ + 3’ discussion; individual reports in groups of five, one group starts at 1 pm) “Annotation of the first gene in my project”

4/6 Wed   **Lecture**: RepeatMasker; other ways to find repeats (Buhler).
CS HW5 due.

4/8 Fri   **Lecture**: Finding repeats in Drosophila (Leung)
Draft report on one gene due (WI exchange papers)

4/11 Mon  **Lecture**: Targeting heterochromatin formation in Drosophila (Elgin)
**Lab**: Continue analysis and annotation of own project.
**Reading**: start #4 (motif hunting, regulation); RR #4 due 4/22

4/13 Wed  **Lecture**: Further characterization of genes/gene products: use of Clustal, use of FlyBase resources (Yu He)
**Lab**: Preparation of annotation reports; gene intron/exon structures
**Presentation**: Bio 4342 alumni present annotation project oral reports

4/15 Fri  **Guest lecture**: Rick Wilson, Human Genetics
**WI subgroup**: critiques of annotation drafts due, discussion.

4/18 Mon  **Guest lecture**: The RNAi system
**Lab**: Preparation of annotation reports; TSS determinations; functions of genes

4/20 Wed  **Lab**: Preparation of annotation reports; TSS determinations; functions of genes
**Lab Meeting**: oral report on finding TSSs (9’ + 3’ discussion; individual reports in
groups of five, first group at 1 pm)

4/22 Fri Discussion: Paper #4 (RR #4 due).

4/25 Mon Lecture: Silencing repeats in Drosophila (Elgin)
Lab: Preparation of annotation reports; start PPT preparation.

4/27 Wed Lab: Preparation of annotation reports

4/29 Fri Lab: Final consultations on annotation report

5/2 Mon 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 alignment 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. Use Clustal at least once. 10’ presentations (1 pm 3 pm OR 3 pm – 5 pm in LS 311).

5/3 Tue 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; 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. 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://gep.wustl.edu) 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):


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).


**Additional References:**

**Sequencing Technology**

Before class starts:

“Sequencing a Genome” (view on line, including review of chemistry if needed):

“Next generation Sequencing: Genome Center Video Tour” (view all four segments online):
http://gep.wustl.edu/curriculum/course_materials_WU/introduction_to_genomics/nextgen_video_tour


Chromatin Structure / Epigenetics


Yandim C, Natisvili T, Festenstein R. (2013) Gene regulation and epigenetics in Friedreich’s ataxia. J Neurochem.126 Suppl 1: 21-42. (This paper includes a review of background information as well as recent results in a mammalian system.)


**Human Genomics**

**Weeks 5+**


**Variation**


**General correlations**


**Cancer**


Finding Genes in Drosophila


Leung, W, CD Shaffer, T Cordonnier, J Wong, MS Itano, EE Slawson-Tempel, E Kellmann, DM Desruisseau, C Cain, R Carrasquillo, TM Chusak, K Falkowska,


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


Revised 01/12/2016