MB 668 Microbial Bioinformatics and Genome Evolution

4 credits
Spring, 2017

Instructors: T. Sharpton, R. Mueller and S. Giovannoni

Thomas Sharpton (Microbiology): thomas.sharpton@oregonstate.edu
  Office Hours: Monday, 3:00-4:00, Nash 530, or by appointment
Ryan Mueller (Microbiology): ryan.mueller@oregonstate.edu
  Office Hours: Monday, 3:00-4:00, Nash Hall 448, or by appointment.
Stephen Giovannoni (Microbiology); steve.giovannoni@oregonstate.edu
  Office Hours: Monday, 3:00-4:00, Nash Hall 248, or by appointment.

Lectures: MW   10-11, Cordley 3003

Lab: F   1:00-4:00, Cordley 3003

Course Description: Theoretical and practical issues in microbial genomics. This course teaches the bioinformatic theory and methods used to assemble, annotate, and analyze microbial genomes, with an emphasis on evolutionary theory and comparative analysis of microbial genome sequences. This includes techniques for searching biological sequence databases, aligning sequences, constructing phylogenies, inferring function, and predicting metabolism. It also explores applications of genomics and allied tools to microbial populations and communities, including population genomics, metagenomics, metaproteomics, and metatranscriptomics.

Course Format: Monday and Wednesday lectures, and Friday computer-based laboratory. Specific prerequisites are not enforced, but students should enter the course with a basic knowledge of computers and molecular biology. Course projects consisting of objectives or hypotheses selected by the student, original bioinformatics analyses, literature review, and written and oral presentations, will be central to this course.

Required reading will be from the original scientific literature (see *schedule*, below). PDFs are available on Canvas.

Student Assessment: Lab reports and homework assignments will be reviewed in class, and will count towards the student's participation grade, to provide each student with opportunities to assess their learning.

Student Evaluation:
1. 20% of the grade (100 points) will be based on a take-home midterm, that students will be expected to complete independently, using any forms of published information. The exam will be passed out at the end of the fifth week of classes and will be due exactly one week later, as a hard copy handed in at the end of class.
2. 20% of the grade (100 points) will be based on student participation throughout the course. This will include completion of short homework questions and lab reports based on computational labs.
3. 60% of the grade will be based on oral (20%; 100 points) and written presentations (40%; 200 points) of a course project.

Course project: By Wednesday of the second week a one-half page summary of your class project is due. The summaries will be reviewed by the instructors and an appointment will be arranged if the instructors feel you need guidance. See comments about the project summaries below. A hypothesis or question should be clearly articulated. You are encouraged to use your own or provided software programs in your project, whenever appropriate. Half-way through the semester, students will complete a one-half page progress summary of their work to date. Students will produce a written scientific report of their project. The report should be written in the form of a scientific publication and include an Introduction/Background, Methods, Results, and Discussion sections. This report should be at least 10

double-spaced pages in length. Relevant figures, tables and references should be included in addition to the 10 pages of text. All students will present their projects orally in the final week of class (12 minute presentation, 3 minutes for questions).

   Choosing a good course project begins with formulating a question or hypothesis that can be tested. In your short description of your project, handed in by Wednesday of week two, be sure to indicate not just what your activities will be (i.e., "I want to write a program to analyze RNA seqdata"), but indicate what your objectives are, and why these objectives are significant. This is the same process that many students go through writing a short proposal for their oral prelim. The course project summaries will be evaluated by the course instructors by Friday of week two. If the ideas are vague we will provide suggestions and ask for it to be re-written.

Student Learning Outcomes: At the completion of the course, you should be able to:
1. Apply basic knowledge of the procedures used to assemble genomes, identify genes and predict their function.
2. Analyze problems that involve the above information and concepts and to reach the correct conclusion.
3. Analyze genome sequence data and communicate scientific concepts clearly and concisely in writing.
4. Describe advanced concepts that are exemplified in scientific literature assigned as reading. Comprehension is displayed by written answers to problems that require an understanding of experimental results and theories and their application to new situations.

University and Departmental Policies:

Students with Disabilities: Accommodations are collaborative efforts between students, faculty and Disability Access Services (DAS). Students with accommodations approved through DAS are responsible for contacting the faculty member in charge of the course prior to or during the first week of the term to discuss accommodations. Students who believe they are eligible for accommodations but who have not yet obtained approval through DAS should contact DAS immediately at 737-4098.

Expectations for Student Conduct: The Department of Microbiology follows the university rules on civility and honesty. These can be found at:
http://oregonstate.edu/studentconduct/http:/%252Foregonstate.edu/studentconduct/code/index.php
Behaviors disruptive to the learning environment will not be tolerated and will be referred to the Office of Student Conduct for disciplinary action. Cheating or plagiarism by students is subject to the disciplinary process outlined in the Student Conduct Regulations.

<div align="center">

Schedule (2017):

</div>

Week 1	**Class 1**	**4/3M**	**Introduction to the course** **Lecture 1: Genome Sequencing Strategies** [Sharpton] **Suggested reading:** Edwards&Holt, 2013, *Beginner's guide to comparative bacterial genome analysis using next-generation sequence data*; Metzker, 2010, *Sequencing technologies — the next generation*.
Week 1	**Class 2**	**4/5W**	**Lecture 2: Sequence Alignment** [Sharpton] **Required reading:** Fitch and Smith, 1983, *Optimal Sequence Alignments* **Suggested Reading:** Chapter 2 of Durbin, Eddy, Krogh, Mitchison *Biological Sequence Analysis;* Li (2010) A survey of sequence alignment algorithms for next-generation sequencing.
Week 1	**Class 3**	**4/7F**	**LAB 1: Introduction to Biocomputing and Genome Assembly** **[Sharpton]** **Required reading:** Loman, 2013, *So you want to be a Computational Biologist?* **Suggested Reading:** Pop, 2009, *Genome assembly reborn: recent computational challenges*; Compeau et al. (2011). *How to apply de Bruijn*

graphs to genome Assembly; Nagarajan, 2009, *Sequence assembly demystified.*

Week 2   Class 4   4/10M   **Lecture 3: Protein Evolution [Giovannoni]**
Required reading: Fitch 2000 *Homology: a personal view on some of the problems*
Suggested reading: Sander and Schneider (1991); *The Statistics of Sequence Similarity Scores* (http://www.ncbi.nlm.nih.gov/BLAST/tutorial/Altschul-1.html); Eddy (2004) *What is a hidden Markov model?* Lagunda (2009) *Finding Homologs in Amino Acid Sequences Using Network BLAST Searches*

Week 2   Class 5   4/12W   **Lecture 4: Finding Homologs with Sequence Searches [Giovannoni]**
Required reading: Fitch 2000 *Homology: a personal view on some of the problems*
Suggested reading: Sander and Schneider (1991); *The Statistics of Sequence Similarity Scores* (http://www.ncbi.nlm.nih.gov/BLAST/tutorial/Altschul-1.html); Eddy (2004) *What is a hidden Markov model?*. Lagunda (2009) *Finding Homologs in Amino Acid Sequences Using Network BLAST Searches*

Due: One-half page statement of project aims/hypotheses and methods.

Week 2   Class 6   4/14F   **LAB 2: Genome Annotation [Sharpton]**
Required reading: Mavromatis (2009) *The DOE-JGI Standard Operating Procedure for the Annotations of Microbial Genomes*

Annotation report: On Friday of the third week you will hand in a report on the uncharacterized fragment of DNA you selected in Lab 2, one typed page in length. The Annotation report should include lists of likely homologues for the gene(s) on the fragment (i.e., the top BLAST scores), information about multiple protein domains, if they are present, and information about conserved motifs. The report should also include COG assignments, for CDS's, Pfam assignments, and the results of any other analyses needed to identify homologues.

Week 3   Class 7   4/17M   **Lecture 5: Phylogenetics [Sharpton]**
[Office1]**Required reading:** Yang and Rannala (2012) *Molecular phylogenetics: principles and practice*.
Suggested Reading: James (2006) *Reconstructing the early evolution of Fungi using a six-gene phylogeny*; Faith (1992) *Conservation evaluation and phylogenetic diversity*; Hafner (1988) *Phylogenetic trees support the coevolution of parasites and their hosts.*

Week 3   Class 8   4/19W   **Lecture 6: Phylogenomics [Sharpton]**
Required reading: Delsuc et al. (2005) *Phylogenomics And The Reconstruction of The Tree Of Life.*
Suggested reading: Wolf (2002) *Genome trees and the tree of life;* Eisen (1998) *Phylogenomics: improving functional predictions for uncharacterized genes by evolutionary analysis*

Week 3   Class 9   4/21F   **LAB 3: Phylogenetic Analysis of 16S Ribosomal RNA Genes [Sharpton]**
Required readings: Rajendhran (2011) *Microbial phylogeny and diversity: Small subunit ribosomal RNA sequence analysis and beyond.*

Week 4   Class 10 4/24M   **Lecture 7: Comparative Microbial Genomics [Sharpton]**

Required reading: *Tettelin (2005). Genome analysis of multiple pathogenic isolates of* Streptococcus agalactiae*: implications for the microbial pan-genome.*
Suggested reading: Rocha (2008) *The organization of the bacterial genome*

Week 4   Class 11 4/26W   Lecture 8: Microbial Population Genetics [Giovannoni]

Week 4   Class 12 4/28F   LAB 4: Comparative Genomic Analysis of Protein Families [Sharpton]
Required reading: Edwards (2013) *Beginner's guide to comparative bacterial genome analysis using next-generation sequence data.*

Week 5   Class 13 5/1M   Lecture 9: Microbial Communities: Taxonomic Diversity [Mueller]
Required Reading: Lozupone & Knight (2008) *Species divergence and the measurement of microbial diversity.*
Suggested Reading: Shapiro & Polz (2014) *Ordering microbial diversity into ecological and genetically cohesive units*

Week 5   Class 14 5/3W   Lecture 10: Metagenomics and Functional Diversity [Mueller]
Required Reading: Wooley, et al. (2010) *A primer on metagenomics*, Prakash & Taylor (2012) *Functional assignment of metagenomic data: challenges and applications*

Week 5   Class 15 5/5F   LAB 5: Metagenomic Fragment Recruitment to Genome Sequences [Mueller]

Due: One-half page summary of progress report on student research project.

Take home midterm handed out, due one week later.

Week 6   Class 16 5/8M   Lecture 11: Single Cell Genomics [Giovannoni]

Week 6   Class 17 5/10W   Lecture 12: Protein Structure Prediction [Giovannoni]
Required viewing: Roy et al., *A Protocol for Computer-Based Protein Structure and Function Prediction http://www.jove.com/video/3259/a-protocol-for-computer-based-protein-structure-function.*
Required reading: Sleator (2012) *Prediction of Protein Functions.*
Suggested Reading: Podar et al. (2005) *Evolution of a microbial nitrilase gene family: a comparative and environmental genomics study*; Karplus et al. 2003.

Week 6   Class 18 5/12F   LAB 6: Database Searches Based on Protein Structure [Giovannoni]
Suggested reading: Roy et al, (2011) *A Protocol for Computer-Based Protein Structure and Function Prediction***.**

Take home midterm due. (Please upload onto Canvas, under *assignments***).**

Week 7   Class 19 5/15M   Lecture 13: Metabolic Reconstruction [Giovannoni]
Required reading: Francke, et al. (2005) *Reconstructing the metabolic network of a bacterium from its genome*.
Suggested reading: Thiele and Palsson (2010) *A protocol for generating a high-quality genome-scale metabolic reconstruction.*

Week 7 Class 20   5/17W   Lecture 14: Genomics of Microbial Eukaryotes [Sharpton]

Required Reading: Yandell (2012) *A beginner's guide to eukaryotic genome annotation.*
Suggested Reading: Galagan (2005) *Genomics of the fungal kingdom: Insights into eukaryotic biology.*

Week 7   Class 21 5/19F   LAB 7: Metabolic Reconstruction from a complete genome sequence [Giovannoni]
Suggested reading: Markowitz et al. (2012) *IMG: the integrated microbial genomes database and comparative analysis system.*

Week 8   Class 22 5/22M   Lecture 15: Regulation [Giovannoni]

Week 8   Class 23 5/24W   Lecture 16: Transcriptomics and Proteomics [Mueller]
Suggested Reading: Schneider and Riedel (2010) *Environmental proteomics: Analysis of structure and function of microbial communities.* Ozsolak & Milos (2011) *RNA sequencing: advances, challenges and opportunities.*

Week 8   Class 24 5/26F   LAB 8: Identifying Functional Differences between Microbial Communities [Mueller]
Required Reading: Parks & Beiko (2010) *Identifying biologically relevant differences between metagenomic communities.*

Week 9   Class 25 5/29M   Memorial Day Holiday

Week 9   Class 26 5/31W   Lecture 17: Phage Genomics [Mueller]
Required Reading: Kristensen, et al. (2012) *Orthologous gene clusters and taxon signature genes for viruses of prokaryotes.*
Suggested Reading: Lawrence, et al. (2002) *Imbroglios of viral taxonomy: Genetic exchange and failings of phenetic approaches.*

Week 9   Class 27 6/2F   LAB 9: Detecting Genomic Signatures of Phage Defense [Mueller]
Required Reading: Barrangou, et al. (2007) *CRISPR provides acquired resistance against viruses in prokaryotes.*

Week 10 Class 28 6/5M   Student Project Presentations: Students will prepare a 15-minute scientific presentation that of their projects to the class.

Week 10 Class 29 6/7W   Student Project Presentations: Students will prepare a 15-minute scientific presentation that of their projects to the class.

Week 10 Class 30 6/9F   Student Project Presentations: Students will prepare a 15-minute scientific presentation that of their projects to the class.

Week 11   6/15H   Student Project Written Report Due

(Please upload onto Canvas, under *assignments*).