Genomics

As DNA sequencing technology becomes cheaper and more efficient, the amount of sequencing data available to biologists is exploding. Not only has this allowed us opportunities to compare gene structure in many diverse organisms, but also to study and compare the structure of entire genomes. The sheer volume of information we are now coping with is unprecedented, and new tools and ways of thinking are not only possible, but increasingly necessary. This course will be an exploration of the techniques used to create genomic DNA libraries, to sequence the resulting DNA fragments, and to analyze the sequences of these fragments, at both the gene and genome levels. Students will gain familiarity with the computer programs used to assemble and annotate genomic sequence data as they use them to analyze their own raw data from the Washington University Genome Sequencing Center. This course will be extensively computer-based. We will be working with large (ca. 40-kb) sections of genomic DNA *in silico*: by the end of the semester, each student will have finished improving the sequence quality of one of these 40-kb clones to a publishable level and extensively annotated another, indicating the locations of genes, repeat sequences, and other sequence motifs.

"Finishing" in the field of genomics means to take a partially assembled sequence and correct errors in assembly and sequence to a set level of accuracy; we will be improving our sequences to the NHGRI "mouse standard" (fewer than 1 error per 1000 basepairs). This process will involve:

1. learning to use Consed, a software package used by professional sequence finishers
2. examining and evaluating sequence fragments obtained from The Genome Institute at Washington University
3. reassembling computer-derived errors as appropriate
4. determining which regions are of inadequate quality and need to be re-sequenced
5. selecting reaction parameters and informing The Genome Institute, which will carry out those reactions and return the results to you
6. incorporating the new data and returning to step 3

As you can see, this is an iterative process, and will take several cycles to complete. By the end of it you will have improved your chosen sequence significantly, will have become familiar with the logic and methodology of sequence improvement, and will have prepared both a written and an oral presentation of your finishing work.

Annotation is the process of determining the positions of every significant structure in a particular region of DNA. We will review eukaryotic gene structure and the structures and functions of repeated sequence elements before examining individual projects, sections of unannotated genomic sequence roughly 40,000 bases in size (a "contig"). Your goal will be to accurately identify the location of every open reading frame, every intron/exon splice junction, and every repeat sequence in your contig, as well as homologies to genes in other species. You will evaluate patterns of synteny (are homologous genes found in the same order and arrangement in other species, or have they become "scrambled" over evolutionary time?) and construct a phylogenetic tree for one of the genes in your contig as well.

You will learn several software programs to carry out various aspects of this work, over the course of the second half of the semester, completing the annotation of your contig in that time as well as preparing both a written and an oral presentation of your annotation work.

Classes

Classes are held in Memorial 201 (the Mac lab) for the first half of the semester; for the second half we'll be in PPHAC 331.

We meet Monday and Wednesday evenings, 6:30 pm to 9:30 pm

Accessing the Lab

Feel free to work on your projects whenever you'd like outside of class hours. Memorial 201 is accessible until the building is locked up by Campus Safety at approximately 9 pm. (If you need to get in after they've locked up, call Campus Safety and they will let you in.) Do note however that there is a class held in Memorial 201 on Thursdays from 4 pm to 7 pm.

Genomics Spring 2013

- Home
- Objectives
- Components
- Policies
- Syllabus
- Homework
- Prof. Jones

How Do I…?

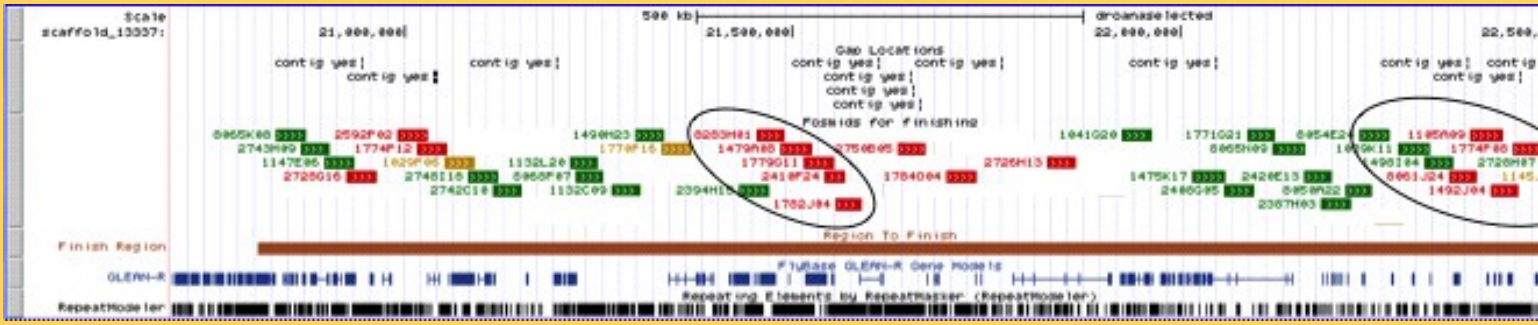
- create a virtual machine?
- format my virtual hard drive?
- work on my project with a USB drive?
- add reads to my project?
- update my restriction digest file?
- use EST data to identify intron/exon boundaries?

Text

The text for this course is *Introduction to Genomics*, 2nd ed., by Arthur Lesk (Oxford University Press). Online resources available for the book can be found here. In addition, readings from the primary research literature will be assigned as the semester progresses.

Finishing fosmid map

Here's a map of the contigs I've selected for us to work on this semester; they're the red ones (of course) in the circles:



Objectives

The primary objective of this course is to give you extensive familiarity with finishing and annotating genomic DNA sequences. This will come as a result of using the strategies and computer tools used by professional genomics researchers to both finish and annotate raw sequence data generated by the Genome Sequencing Center at Washington University, St. Louis, and by reading and discussing articles on genomics-related topics throughout the semester.

Over the course of the semester you will become painfully familiar with the admittedly specialized software used in finishing and annotation, which you may well never encounter again (and by the end of the semester you will no doubt be very grateful that that is the case!). Much more important is the increased knowledge you'll have about gene and genome structure, how knowledge in the field of genomics is acquired and integrated, and the value of genomics research in the larger scientific scheme of things.

Our readings will expose you to discoveries and challenges in recent research articles. Although (depending on your future career) the specifics will not be especially important to you, I hope that our readings and discussions will help you become better at effectively gaining valuable insights from the literature of whatever area you find yourself in in future.

You will also gain experience in writing reports describing the process and results of your experiences and in presenting those results orally. In addition, you will evaluate the oral presentations of your peers, another skill more widely applicable than in just this field.

Finally, all students who participate fully in the course will be included as authors on the (eventual) papers which result from this work; participating in the pulse-pounding excitement that is manuscript revision is an adventure not to be missed!

To summarize, after successfully completing this course, students will have:

- acquired extensive familiarity with computer programs used professionally to assemble and analyze genomic sequence data
- learned how such data is interpreted in the primary literature
- become familiar with the impact of genomics research on current questions in biology and medicine
- improved their ability to present the results of their own research, both in writing and orally
- gained critical insights into the criteria for successful presentations by evaluating the presentations of their peers
- earned the right to be included as an author on any publications incorporating data derived from their work in this course

Genomics Spring 2013

Home

Objectives

Components

Policies

Syllabus

Homework

Prof. Jones

How Do I…?

create a virtual machine?

format my virtual hard drive?

work on my project with a USB drive?

add reads to my project?

update my restriction digest file?

use EST data to identify intron/exon boundaries?

Components

First, a note on grading. I do not grade on a curve, so I hope that each of you will do your best to help your fellow students: if they benefit, it does you no harm. In fact, one of the best ways to learn something is to explain it to someone else, so *talk to your classmates*.

Finishing

There will be three graded items derived from the finishing portion of the course: you will prepare both a written report and an oral presentation on your finishing project, each of these is worth 100 points. In addition, you will be expected to participate in the assessment of your peers' oral presentations; this participation is worth an additional 100 points.

Annotation

There will similarly be written and oral reports, with the latter being peer-evaluated, for the annotation portion of the course, for another 300 points.

Discretionary points

Some of you may feel that you'll do better with your oral presentations than your written reports (or vice versa); you each have 100 points to distribute as you see fit among the oral and written reports. You must notify me of your allocation *before* our class on Monday, March 11 (that's our first meeting after spring break).

Other points

There will be occasional homework assignments over the semester, worth a total of approximately 100 points. Readings from the text and articles which I assign will be discussed in class; these discussions will be worth another 100 points. Finally, overall class participation will be worth 100 points.

Summary

There will be a total of approximately 1000 points possible over the course of the semester; your final grade will be calculated as a simple fraction of points earned over the total possible.

Genomics Spring 2013

- Home
- Objectives
- Components
- Policies
- Syllabus
- Homework
- Prof. Jones

How Do I…?

- create a virtual machine?
- format my virtual hard drive?
- work on my project with a USB drive?
- add reads to my project?
- update my restriction digest file?
- use EST data to identify intron/exon boundaries?

Policies

Below you will find various course policies, including:

- Reading
- Late Assignments
- Extra Credit
- Food
- Cell Phones and their ilk
- Academic Honesty
- Accommodations

Attendance

This is an upper-level elective course, and as you will discover there is a lot of hands-on work to be done. While some, particularly in the latter half of the course, may be done on your own, I think you'll find that working in class along with your compatriots will greatly improve the experience. That said, I don't require attendance for this course. (Although part of your grade for the courses will be based on participation, peer evaluation, and class discussion, so excessive absence will certainly not help you there.)

Grading

Grades for this course will be determined as described in the Components section of this site.

I'm going to be using the point system for this course, so you don't have to worry about calculating percentages for individual components. Given my grading scale (see below), you can therefore calculate your own grade in the course at any time.

In summary, then:

Finishing reports (oral and written)	200 points
Peer finishing evaluations	100 points
Annotatian reports (oral and written)	200 points
Peer annotation evaluations	100 points
Discretionary points	100 points
Homework	100 points
Class Discussions	100 points
Participation	100 points
Anticipated Total	**1000 points**

I reserve the right to tweak these distributions as I see fit.

Here is the grading scale I use in all my classes:

numeric grade	letter grade
93.3 - 100	A
90.0 - 93.2	A-
86.7 - 89.9	B+
83.3 - 86.6	B
80.0 - 83.2	B-
76.7 - 79.9	C+
73.3 - 76.6	C

Genomics Spring 2013

- Home
- Objectives
- Components
- Policies
- Syllabus
- Homework
- Prof. Jones

How Do I...?

- create a virtual machine?
- format my virtual hard drive?
- work on my project with a USB drive?
- add reads to my project?
- update my restriction digest file?
- use EST data to identify intron/exon boundaries?

70.0 - 73.2	C-
66.7 - 69.9	D+
63.3 - 66.6	D
60.0 - 63.2	D-

Just to review, this is what the Student Handbook has to say about grades:

A and A-
: These grades are given for achievement of the highest caliber. They reflect independent work, original thinking, and the ability to acquire and effectively use knowledge.

B+, B, and B-
: These grades are given for higher than average achievement. Evidence of independent work and original thinking is expected.

C+, C, and C-
: These grades are given when the student has devoted a reasonable amount of time, effort, and attention to the work of the course and has satisfied the following criteria: familiarity with the content of the course, familiarity with the methods of study of the course, and active participation in the work of the class.

D+, D, and D-
: These grades are given for unsatisfactory work, below the standard expected by the College. They indicate work which in one or more important aspects falls below the average expected of students for graduation. The work is, however, sufficient to be credited for graduation, if balanced by superior work in other courses.

Note that there isn't a whole lot of emphasis on *knowledge* here (as traditionally defined). Which is odd, but in any event, these descriptions are guidelines, not absolute criteria for a given grade. If you work independently, think originally, and are able to acquire and effectively use knowledge, but don't know squat about genetics at the end of the course, that's simply not "A" work. Context is important, people.

Reading

In this course, the reading is an important important supplement to our class activities. Part of your grade will come from readings which we will discuss in class. My recommended strategy for tackling reading in an area you're not familiar with is to skim the entire article or paper early on, then come back to it later for a more thorough perusal. Probably more than once. Don't try to memorize all of the content, there are no content-based examinations in this course. My goal is to help you become better consumers of scientific literature and to gain familiarity with issues, strategies, and discoveries in modern genomics research.

Late Assignments

Assignments turned in late will not be accepted. Period.

Extra Credit

There is no opportunity for extra credit in this course.

Food

No eating in class, unless you can convince me it's medically necessary. The computers don't deal well with spilled comestibles and (especially) potables.

Cell Phones

Cell phones, like television, are tools of Satan. They are without significant positive value in my world and while I don't expect you to share my view of them, I expect you to spare me from being rudely reminded of their existence. If you are expecting an ***urgent*** phone call while in class, alert me to that fact ahead of time. Otherwise, if your cell phone goes off in class, you can expect me to penalize you some number of points, based entirely on my whim. Someday this will be looked on as one of my lovable eccentricities, but until then you'll just have to put up with my sociopathy.

My antipathy extends to Blackberries and other such devices as well. If you want to text, tweet, IM, email, browse the web, check your stocks, shop, or do anything else that you feel is more important than paying attention in class, go for it. But don't do it in this class: you've got a computer right in front of you, but it's important that you stay focused on the task(s) at hand. I have had students in years past who thought they could multi-task, holding IM chats and I-don't-know-what-all while in class. They didn't do nearly as well as they might have, had they been paying more attention to what they were supposed to be doing. Our time in this class is very limited: don't waste it.

Academic Honesty

I adhere to the Academic Honesty policy of the College. There is nothing more important to me than personal integrity — not happiness, not power, not even genomics, nothing — and I conduct myself and all of my classes in that spirit. If you're not familiar with College policy, you should be.

Accommodations

Per Moravian College policy: "Students who wish to request accommodations in this class for a disability should contact Elaine Mara, assistant director of learning services for academic and disability support, at 1307 Main Street or by calling 610-861-1510. Accommodations cannot be provided until authorization is received from the Academic Support Center."

Syllabus

Meeting number	Date	In-Class	Background Reading
1	Mon., Jan. 14	Organizational meeting project overview	
2	Wed., Jan. 16	discuss chapter 1 Consed orientation & exploration	text chapter 1
3	Mon., Jan. 21	no class (MLK Day)	
4	Wed., Jan. 23	discuss chapter 2 work on finishing projects	text chapter 2
5	Mon., Jan. 28	work on finishing projects	
6	Wed., Jan. 30	work on finishing projects	
7	Mon., Feb. 4	discuss chapter 3 work on finishing projects	text chapter 3
8	Wed., Feb. 6	work on finishing projects	
9	Mon., Feb. 11	discuss chapter 4 work on finishing projects	text chapter 4
10	Wed., Feb. 13	work on finishing projects	
11	Mon., Feb. 18	work on finishing projects	
12	Wed., Feb. 20	work on finishing projects	
13	Mon., Feb. 25	discuss chapter 5 work on finishing projects	text chapter 5
14	Wed., Feb. 27	work on finishing projects	
	Mon., Mar. 4	No class (Spring Break)	
	Wed., Mar. 6	No class (Spring Break)	
15	Mon., Mar. 11	discuss chapter 6 work on finishing projects	text chapter 6
16	Wed., Mar. 13	Finishing Project presentations	
17	Mon., Mar. 18	discuss chapter 7 discuss annotation	text chapter 7
18	Wed., Mar. 20	work on annotation projects	
19	Mon., Mar. 25	discuss chapter 8 work on annotation projects	text chapter 8
	Wed., Mar. 27	work on annotation projects	
20	Mon., Apr. 1	discuss chapter 9 work on annotation projects	text chapter 9
21	Wed., Apr. 3	work on annotation projects	
22	Mon., Apr. 8	discuss chapter 10 work on annotation projects	text chapter 10
23	Wed., Apr. 10	work on annotation projects	
24	Mon., Apr. 15	discuss chapter 11 work on annotation projects	text chapter 11

Genomics Spring 2013

- Home
- Objectives
- Components
- Policies
- **Syllabus**
- Homework
- Prof. Jones

How Do I…?

- create a virtual machine?
- format my virtual hard drive?
- work on my project with a USB drive?
- add reads to my project?
- update my restriction digest file?
- use EST data to identify intron/exon boundaries?

25	Wed., Apr. 17	work on annotation projects
26	Mon., Apr. 22	work on annotation projects
27	Wed., Apr. 24	Annotation Project presentations

Life is fluid, so this syllabus is subject to change. I may have to change the syllabus as the semester progresses, but this is certainly preferable to rigidly adhering to some timetable in lockstep. So come to class and you'll always know what's going on with the syllabus; changes will of course also be posted here, but you should be in class anyway!

Professor Jones

If you ever have questions that you can't answer yourself, realize that there are a lot of resources available to you: if your classmates can't help you, feel free to ask me. This is part of my job, and one which I don't shirk. Depending on what the problem is, the most reliable method is probably email (I sometimes don't realize I have voicemail for a day or two). My email address is cjones [at] moravian [dot] edu and my office (and lab) phone number is 610-861-1614.

If you need to speak with me sometime when I'm not on campus (a rare event!), call me at home any time between 9 am and 9 pm. Students often tell me they don't feel comfortable calling me at home because they think I mind. Consider the logic here: there's nothing that says I have to give you my home phone number, yet I have done so in class. So ***why*** would I give you that number if I didn't want you to use it? Note that "use" is not the same thing as "abuse": don't call me at 3 am the day before an assignment is due and expect much sympathy (or pleasant conversation)!

Here's a copy of my current class schedule. My official office hours this semester are from 9:30 pm to 10:30 pm Mondays, right after this class. That said, official hours are all but irrelevant to me. If I'm not in my office (Room 310, Collier Hall of Science), try my lab (Room 233, Collier Hall of Science — between the elevator and the loading dock on the main floor). Feel free to get hold of me any time; if I can't spare the time to talk then, I'll tell you so, and we can set up an appointment at our mutual convenience.

Genomics Spring 2013

- Home
- Objectives
- Components
- Policies
- Syllabus
- Homework
- Prof. Jones

How Do I…?

- create a virtual machine?
- format my virtual hard drive?
- work on my project with a USB drive?
- add reads to my project?
- update my restriction digest file?
- use EST data to identify intron/exon boundaries?