Teaching Bacterial Pathogenesis using Bioinformatics and Genomics Approaches

Janelle Hare
Morehead State University

Janelle Hare
Biological and Environmental Sciences Department
LA327-G Lappin Hall
Morehead State University
Morehead, KY 40351
(606)783-2951, FAX 783-5002
jm.hare@moreheadstate.edu

ABSTRACT

"Changing Concepts in Pathogenesis" is a 300-level elective course I developed for biology undergraduates who have not taken microbiology. To make this course attractive and accessible, I structured it to have a conceptual, not organismal, focus on how scientists research and understand microbial pathogens with new molecular, genomic, and bioinformatic technologies. To support our classroom activities, we used computer lab projects. We examined how societal changes results in new pathogens, and how new technologies help us learn about pathogens. We first used the Bacterial Pathogenesis text to learn about microbes, host defenses and bacterial genetics. Later, we read the primary and secondary literature to understand recent advances such as *in vivo* technologies and microarrays. We used the Lasergene Navigator software suite from DNASTAR, which is free for educational use, runs on both PC and Macintosh platforms, and contains a variety of DNA and protein analysis programs. However, because many bioinformatics texts focus on eukaryotic DNA analyses, I designed computer lab exercises that focused on bacterial genetics. Students learned the Lasergene software, navigated publicly available bacterial genome databases, and designed and conducted research projects using the software. Their topics included homologies among Type IV secretion systems, genes involved in crossing the blood-brain barrier, superantigens, toxin homologs, and differences in quorum sensing systems of pathogens *vs.* non-pathogens. Post-semester assessments indicated that students enjoyed the novel, *in silico* approach to studying microbiology, and learned new skills that they could apply to future studies in biology or computing.

INTRODUCTION: Syllabus Course Description & Organization

~~etiology, symptoms, disease, prevention, treatment, epidemiology of a wide variety of infections...~~

As fun as that is, **this course is for upper-level biology students who may not have had microbiology.** ⇒

concepts of: pathogenicity & pathogen-host interactions, by investigating genetic bases of virulence.

This course examines both traditional notions of pathogenicity and newer, expanded models of pathogenicity that have resulted from recent molecular biology, genome sequencing, and bioinformatics technologies. Some themes include:

- *the role of horizontal gene transmission in pathogen evolution*
- *Using "in vivo" techniques to see how host cells and pathogens interact*

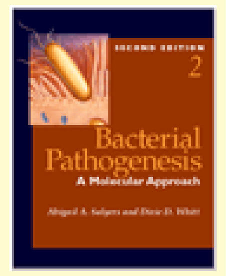
These concepts are increasingly relevant in a world replete with emerging infectious diseases, the potential for bioterrorism, and interactions between immunocompromised hosts and pathogens. The main focus is on bacterial pathogens, but could also include fungal, viral and parasitic pathogens. This class is taught in a workshop format with lecture, class discussions, oral presentations, and computer workshop activities. Computer bioinformatic analyses of DNA sequences from publicly available databases forms the basis of student research projects.

METHODS: Course Objectives

- To become familiar with a variety of virulence mechanisms and common virulence genes
- To learn how bacteria acquire and regulate DNA sequences and thus evolve as pathogens
- To examine recent molecular techniques used to investigate pathogen genomes & virulence
- To learn how to analyze DNA sequences and describe gene structure, organization and content
- To explore (some of!) the genome sequencing data available for microorganisms

METHODS: Building the Basics

Readings for first quarter of the course



Koch, 1884. The etiology of tuberculosis (Koch's postulates).

Bacterial Pathogenesis (Salyers & Whitt 2nd ed.)

Smith, 1995. The Revival of Interest in Mechanisms of Bacterial Pathogenicity. *Biol. Rev*. 8:288-317.

Falkow, 1988. Molecular Koch's postulates applied to microbial pathogenicity. *Rev. Infect. Dis.* 10 Supl. 2:S274.

METHODS: Pathogens Emerge

Readings for second quarter of the course

Alternate models of pathogenicity

P.W. Ewald. 1998. The Evolution of Virulence and Emerging Diseases. *J. Urban Health.* 73:480-491.

P.W. Ewald. 1996. Guarding Against the Most Dangerous Emerging Pathogens: Insights from Evolutionary Biology. *Emerging Infectious Diseases*. 2:245-257.

Readings and Powerpoint presentations on emerging infections

EMERGING INFECTIOUS DISEASES — A Peer-Reviewed Journal Tracking and Analyzing Disease Trends

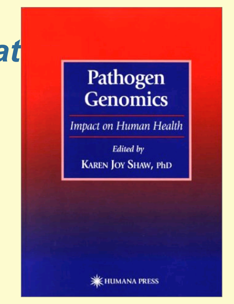
http://www.cdc.gov/ncidod/eid/

Microbiographies

Students use primary literature to research one pathogen, and write a microbiography of that organism, describing its pathogenic lifestyle and virulence genes.

METHODS: Technologies aiding understanding

We'll look at recent advances gained through bioinformatics techniques, genome sequencing and new molecular methods that dissect interactions of the pathogen and host.



Readings for second half of the course

Pathogen Genomics: *Impact on Human Health* (excerpts)
Primary literature on: bioinformatics, genomics, proteomics, microarrays, animal model systems, Type 3 secretion systems, *Escherichia coli, Yersinia pestis* & general pathogen evolution, and *in vivo* strategies for identifying virulence genes (IVET, STM, DFI, & SCOTS).
For specific references, see link to syllabus ("Resources").

METHODS: Learning *in silico* research techniques: DNA/Protein Analysis software

Understanding the genetic basis of virulence: During the last half of the course, we'll spend part of each class learning to conduct various bioinformatic analyses on DNA sequences, using DNAStar Lasergene software.

DNASTAR, Inc — Lasergene®

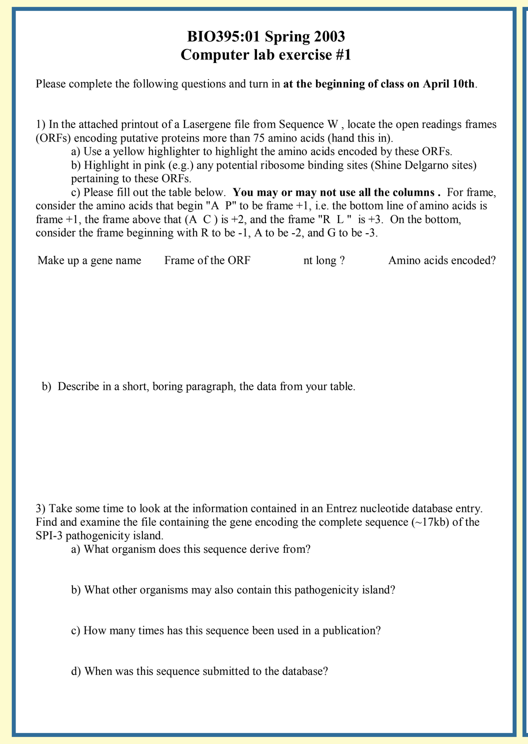
"Lasergene is a comprehensive suite of easy-to-use sequence analysis software for Windows and Macintosh computers. Lasergene provides multiple analysis tools for DNA and protein sequences including sequence alignment, contig assembly, gene discovery, primer design, restriction mapping, and protein structure prediction. Lasergene also integrates BLAST and Entrez searching for easy sequence downloading and comparisons. Systems are available on a standalone basis or networked and they are backed by expert technical support. Sign up to try our fully functional Lasergene software demonstration by clicking here." http://www.dnastar.com

I demonstrated, on a video projection screen :

- use of the various software capabilities
- the NCBI website and associated genomic databases

Each student used a wireless notebook computer in class. They also received a copy of the Lasergene software for use on their personal home computer, to practice and conduct research outside of class.

Sample exercises I designed are shown below.



BIO395:01 Spring 2003
Computer lab exercise #1

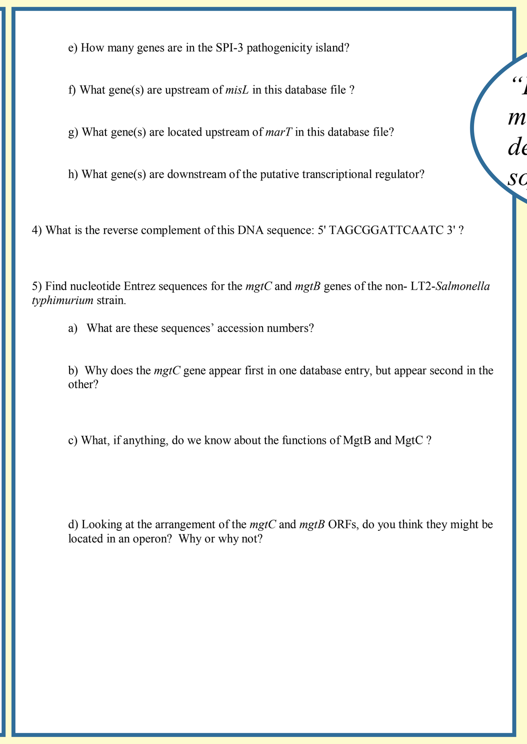
Please complete the following questions and turn in **at the beginning of class on April 10th**.

1) In the attached printout of a Lasergene file from Sequence W , locate the open readings frames (ORFs) encoding putative proteins more than 75 amino acids (hand this in).
a) Use a yellow highlighter to highlight the amino acids encoded by these ORFs.
b) Highlight in pink (e.g.) any potential ribosome binding sites (Shine Delgarno sites) pertaining to these ORFs.
c) Please fill out the table below. **You may or may not use all the columns .** For frame, consider the amino acids that begin "A P" to be frame +1, i.e. the bottom line of amino acids is frame +1, the frame above that (A C) is +2, and the frame "R L " is +3. On the bottom, consider the frame beginning with R to be -1, A to be -2, and G to be -3.

Make up a gene name	Frame of the ORF	nt long ?	Amino acids encoded?

b) Describe in a short, boring paragraph, the data from your table.

3) Take some time to look at the information contained in an Entrez nucleotide database entry. Find and examine the file containing the gene encoding the complete sequence (~17kb) of the SPI-3 pathogenicity island.
a) What organism does this sequence derive from?
b) What other organisms may also contain this pathogenicity island?
c) How many times has this sequence been used in a publication?
d) When was this sequence submitted to the database?

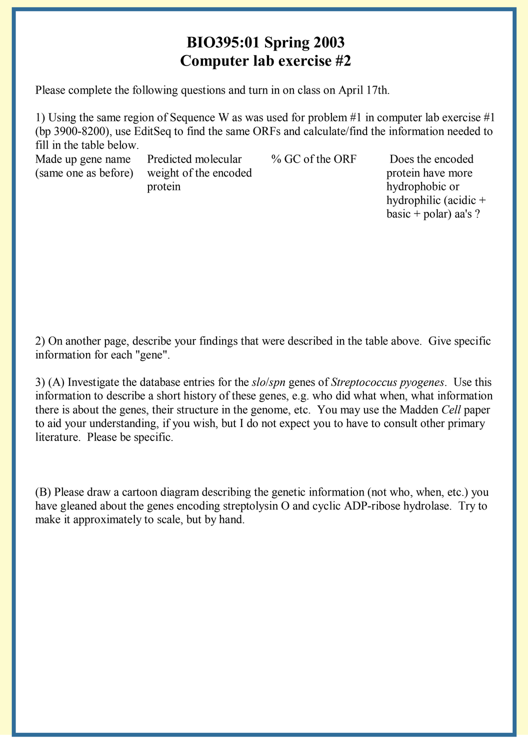


e) How many genes are in the SPI-3 pathogenicity island?
f) What gene(s) are upstream of *misL* in this database file ?
g) What gene(s) are located upstream of *marT* in this database file?
h) What gene(s) are downstream of the putative transcriptional regulator?

4) What is the reverse complement of this DNA sequence: 5' TAGCGGATTCAATC 3' ?

5) Find nucleotide Entrez sequences for the *mgtC* and *mgtB* genes of the non- LT2-*Salmonella typhimurium* strain.
a) What are these sequences' accession numbers?
b) Why does the *mgtC* gene appear first in one database entry, but appear second in the other?
c) What, if anything, do we know about the functions of MgtB and MgtC ?
d) Looking at the arrangement of the *mgtC* and *mgtB* ORFs, do you think they might be located in an operon? Why or why not?

"Daily computer assignments helped a great deal. It was good to see software in action."



BIO395:01 Spring 2003
Computer lab exercise #2

Please complete the following questions and turn in on class on April 17th.

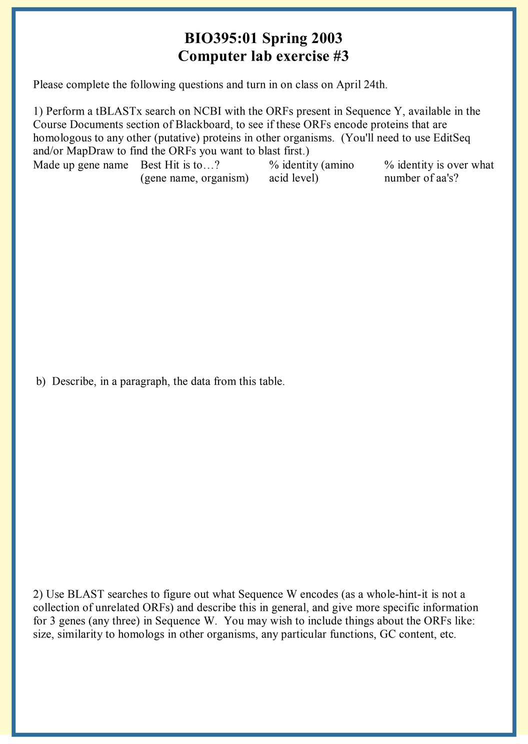
1) Using the same region of Sequence W as was used for problem #1 in computer lab exercise #1 (bp 3900-8200), use EditSeq to find the same ORFs and calculate/find the information needed to fill in the table below.

Made up gene name (same one as before)	Predicted molecular weight of the encoded protein	% GC of the ORF	Does the encoded protein have more hydrophobic or hydrophilic (acidic + basic + polar) aa's ?

2) On another page, describe your findings that were described in the table above. Give specific information for each "gene".

3) (A) Investigate the database entries for the *slo/spn* genes of *Streptococcus pyogenes*. Use this information to describe a short history of these genes, e.g. who did what when, what information there is about the genes, their structure in the genome, etc. You may use the Madden *Cell* paper to aid your understanding, if you wish, but I do not expect you to have to consult other primary literature. Please be specific.

(B) Please draw a cartoon diagram describing the genetic information (not who, when, etc.) you have gleaned about the genes encoding streptolysin O and cyclic ADP-ribose hydrolase. Try to make it approximately to scale, but by hand.



BIO395:01 Spring 2003
Computer lab exercise #3

Please complete the following questions and turn in on class on April 24th.

1) Perform a tBLASTx search on NCBI with the ORFs present in Sequence Y, available in the Course Documents section of Blackboard, to see if these ORFs encode proteins that are homologous to any other (putative) proteins in other organisms. (You'll need to use EditSeq and/or MapDraw to find the ORFs you want to blast first.)

Made up gene name	Best Hit is to…? (gene name, organism)	% identity (amino acid level)	% identity is over what number of aa's?

b) Describe, in a paragraph, the data from this table.

2) Use BLAST searches to figure out what Sequence W encodes (as a whole-hint-it is not a collection of unrelated ORFs) and describe this in general, and give more specific information for 3 genes (any three) in Sequence W. You may wish to include things about the ORFs like: size, similarity to homologs in other organisms, any particular functions, GC content, etc.

METHODS: Using other technologies

Blackboard-assisted discussions of daily readings:
What do you see as the main point(s) of the reading?
What did you find most interesting, and why?
What was the most confusing thing you read?

"The project was interesting and indicated how much work goes into studying pathogens at the genetic level."

RESULTS: Student Research Papers

You'll then do an in silico research project, again on either an organism or infection process of your choice, using genome sequence data from public databases on the internet.

Students identified both the genomic databases for their organism(s) and the "seed" genes with which to query these database. They then investigated similarities among and between their results, and attempted to interpret the significance of these matches.

"Type IV Secretion Systems: Identification of Putative Coding Regions using Conserved Sequence for *virB4* and *virB11*"

"A Bioinformatic Approach to Quorum Sensing Systems of Pathogenic and Non-pathogenic Bacteria

"Searching for Homology in Genetic Sequences of Superantigen-producing Bacteria: The value of Inter- and Intra-Species Genetic Homology"

"Homologies found between toxins in *Clostridium* species and non-toxic *C. Acetobutylicum"*

"Comparative Study of the Bacterial Techniques to Cross the Blood-Brain Barrier"

"It is one of the only courses here that deals with things that are actually applicable in the medical world."

CONCLUSIONS

This was a very challenging class, and worked well in this group of 12 students. It would also be suitable for a guided independent study or interdisciplinary class co-taught by faculty in computer science and biology.

Improvements suggested:

- Carefully monitor student computer use while demonstrating software on video projection screen, as it's difficult for them to concentrate while their computer is showing their own work. The Robotel system would be great for this application.
- Make field trips to labs where microarrays, flow cytometers, etc. are used, to aid in understanding these more abstract techniques.
- Make sure students understand the reasons for meticulous genetic analysis and the possible phenotypic repercussions of altered genetic information. Critical analysis is particularly important when using unpublished/genomic databases.

RESOURCES/ACKNOWLEDMENTS

A copy of the syllabus used for this class can be found at:
http://people.morehead-st.edu/fs/jm.hare/, under "Courses".

I thank: DNA STAR, Inc., for allowing us to use the Lasergene Navigator program for this class, members of the BIO395 class, Leslie Gregg-Jolly for her input into the development of this class, and ASM for an Early Career Faculty Travel Grant to attend this meeting.