Course Syllabus

BIOL 358 Computer Skills for Biotechnology

Spring 2007

Class Hours  Monday 17:00-20:15 (Sec 01; CRN 21358)
Classroom  SCI2 308
Instructor  Patrick Sebrechts, Department of Biology
Office and Office Hours  SCI2 237; (M & W: 10:30-12:30; Thurs: 12:00-14:00; By Appointment)
E-mail  patrick@csusm.edu
Home Page  http://www.csusm.edu/patrick
Phone  760.943.9121

Biotechnology Web:  http://www.csusm.edu/biotechnology

(The above web site was developed by two former BIOL 358 students as their class project!)

Important Dates

First day of class: Monday, 22 January 2007
Spring Break: Monday, 26 March 2007
Last Day of Class: Monday, 7 May 2007

Materials:

All lecture notes and other course materials, including a list of recommended books will be made available as part of this syllabus.

Course Objectives:

This course is designed to introduce and explain the application of computational and analytical methods to address problems in biotechnology. Many popular software tools employed in biotechnology and bioinformatics research will be covered. The theoretical basis governing the use and importance of these tools will be explored.
**Disabled Student Services:**

Students with disabilities who require academic accommodations must be approved for services by the Office of Disabled Student Services (DSS) located in Craven Hall 5205. DSS can be contacted by phone at 760.750.4905 or TTD 760.750.4909. Students authorized by DSS to receive accommodations should meet with me in a private setting to ensure your confidentiality.

**Grades:**

Grades will be assigned based on homework, class participation and a final project. There will be no mid-term and no final; however students are required to present their final projects during the last sessions of class.

**Books**

The following list of books is *recommended* reading for the class.

- **Developing Bioinformatics Computer Skills**
  by Cynthia Gibas, Per Jambeck
- **Beginning Perl for Bioinformatics**
  by James Tisdall
- **Bioinformatics for Dummies**
  by Jean-Michel Claverie, Cedric Notredame
- **Structural Bioinformatics**
  by Philip E. Bourne, Helge Weissig

Looking for more advanced books on each topic? You might find this list helpful:

- **Linux in a Nutshell**
  by Ellen Siever, Aaron Weber, and Stephen Figgins
- **Learning Perl, Third Edition**
  by Randal L. Schwartz and Tom Phoenix
- **Programming Perl, Third Edition**
  by Larry Wall, Tom Christiansen, and Jon Orwant
- **The Definitive Guide to MySQL, Second Edition**
  by Michael Kofler
- **Biological Sequence Analysis: Probabilistic Models of Proteins and Nucleic Acids**
  by Richard Durbin, Sean R. Eddy, Anders Krogh, and Graeme Mitchison
  by Andreas D. Baxevanis and B. F. Francis Ouellette

---

**Lectures**
I. Essential Basics

1. Computer Systems Math & Logic

   Computer Mathematics
   Logic and Truth Tables
   ASCII Code
   Logic Gates
   Lecture 1

2. Computer Systems & the Internet Part 1

   Internet -- An Introduction
   FrontPage Overview

   Web Site Construction Part 1

   Garden Web Site Tutorials
   FrontPage Data Files
   Solutions to FrontPage Exercises

   Site Exposure

   Web Link Article
   Lecture 2

3. Computer Systems & the Internet Part 2

   Web Site Construction Part 2

   Garden Web Site Tutorials
   FrontPage Data Files
   Solutions to FrontPage Exercises

   Search Engines

   Examples: Google, Yahoo, AltaVista, Ask, Wisenet ...

   Boolean operators: AND, OR, NOT
Assignment

Team selection

Develop Website foundation and outline

Hyperlink to J. Craig Venter Institute

Purpose of the institute?

Links to San Diego?

What is the Sorcerer II?

What is genome shotgun sequencing?

Lecture 3

4. Databases

Review

Flat File

Relational

File -- (Table) -- Record -- Field

Microsoft Access®

Incorporation into Web Project.

Lecture 4

5. Programming

Project Web Site Assignments

Fundamentals of Programming

Creating a First Program

Reference

Lecture 5

6. Web Publishing, UNIX, LINUX, vi Text Editor, CGI Programming, Perl
Microsoft FrontPage enabled server

LINUX server

Accessing LINUX

SHH (Secure Shell -- replaces FTP)

Telnet/SSH client -- PuTTY

WS_FTP Home 2006 (30 day trial)

UNIX/LINUX

UNIX

The UNIX Primer

UNIX Tutorial for Beginners

UNIX Tutorial

A Basic UNIX Tutorial

LINUX

vi Text Editor

vi Text Editor

vi Lesson Document

vi Intro

Other Editors

Pico

Emacs

CGI Programming

CGI Programming

Perl

Perl -- An Introduction

Learning Perl
A Simple Perl Script

7. Using Bioinformatics

Bioinformatics Resources

10 Major Bioinformatics Databases

10 Major Bioinformatics Software Programs

10 Major Bioinformatics Resource Locators

PubMed


Public version of Medline.

PubMed Tutorial

Search Field Descriptions and Tags

Entrez: The National Center for Biotechnology Information (NCBI) querying system that is best known in the context of the PubMed/Medline bibliographical databases.

Expasy, the SWISS-PROT database home page.

http://www.expasy.org/sprot/

Protein information, protein sequence database, analytical tools, links.

Lectures 6 & 7

Homework Exercises

Team work previews.

8. BLAST; ClustalW; The Genographic Project; Using Nucleotide Sequence Databases

BLAST

Basic Local Alignment Search Tool.

**BLAST Tutorial**

ClustalW

http://pir.georgetown.edu

Multiple alignments.

**Tutorial**

The Genographic Project - a study that charts the migration of humankind through history.

https://www3.nationalgeographic.com/genographic/

GenBank

Go to http://www.ncbi.nlm.nih.gov/Education/ for the following GenBank tutorials:

<table>
<thead>
<tr>
<th>BLAST Information</th>
<th>Entrez tutorial</th>
<th>PubMed tutorial</th>
<th>NCBI News</th>
</tr>
</thead>
<tbody>
<tr>
<td>Resource publications</td>
<td>Map Viewer exercises</td>
<td>Structure tutorial</td>
<td>NCBI Handbook</td>
</tr>
</tbody>
</table>

In addition NCBI, (National Center for Biotechnology Information), offers many educational resources including a science primer "to gain an understanding of our resources and explore our databases and tools to see what we can do for you."
9. **Common Uses of Databases -- review.**

   **Homework Exercises**

   **WebCT submissions**

Gene-Centric Databases


Whole Genome Databases

   **Complete Viral Genomes**


   **Complete Bacterial Genomes**


TIGR (The Institute for Genome Research)


Exploring the Human Genome

   **Ensembl project**

SNPs (Single Nucleotide Polymorphisms)

http://www.ensembl.org/

Other tools

The Genome Bioinformatics Laboratory at the University of California Santa Cruz

http://genome.cse.ucsc.edu/

NCBI


10. Bioinformatics: The Use of Computer Systems to Understand Biology

Bioinformatics Applications

PDF version

Video

Giant Brains

Using Protein and Specialized Sequence Databases

Swiss-Prot

http://www.expasy.ch/sprot/

11. Database Reviews

Critique of Web Sites
13. Introduction to Structural Bioinformatics

Structural Bioinformatics is a specialized area of bioinformatics where tools are developed to navigate between sequences and 3-D structures.

FASTA-formatted .txt file for GenPep NP_360043, Rickettsia conorii sequence

PSIPRED: a prediction tool for secondary protein structure.

Example using GenPep NP_360043, Rickettsia conorii sequence TolB

PredictProtein: probably the most comprehensive site for protein structure analysis.

Protein Data Bank (PDB) (3D Structure)

Rasmol
The Protein Data Bank & Visualization with Rasmol [HTML]

Links: PDB | Rasmol | Other Graphics Tools

Structure Determination & Structure/Function Assignment

Secondary Structure Calculation & Structure Classification

14 & 15. Presentations

Roster