

Bioinformatics Course Proposal

Audience: Both non-biologists and molecular biology mavens.

Goals:

1. Ability to use and understand the major features of the NCBI website.
2. Understand and effectively use sequence searching parameters.
3. Understand and observe postulated evolution of proteins/genes.
4. Understand and observe the interconnectedness of genetic diseases, gene alleles, proteins booboos, protein functions, amino acid sequences, and nucleotide sequences.
5. Study an amalgamation of disease/gene/protein using the NCBI website.
6. Perform basic bioinformatics procedures regarding—homologues and orthologs, SNPs, cladograms, e-PCR, 3D protein models, and gene expression.
7. Course should be amenable to the online format.

Un-Goals: No pretension of becoming experts in computer programming, search techniques, sequence alignment strategies, protein folding, mechanics of cell biology, By this disclaimer I mean to specify that the biologists are not to be burdened with the mechanics of implementing search or alignment algorithms but rather with understanding the meaning of alignment scores and the effects alignment parameter selections. Likewise the non-biologists are presented with the information flow perspective of molecular biology rather than the mechanisms the cell uses to achieve its mission.

Proposed Organization of the Course:

Five Parts: a vocabulary and central dogma part for the non-biologists, a sequence alignment section, a using the NCBI website section, a special topics section (SNPs, gene expression, orthologs, viruses, prokaryotes, etc), a student projects/notebook section.

Outline:

Part One—Vocabulary and Central Dogma

A. Overview

1. Graphic of central dogma (DNA→mRNA→Protein) with hot linked vocabulary.
2. Lecture or text.
3. Fill in the blank exercise.
4. Quiz

B. DNA Structure

1. Graphic with hot linked vocabulary.
2. Lecture or text.
3. Fill in the blank exercise.
4. Quiz

C. Transcription

1. Graphic with hot linked vocabulary.
2. Lecture or text.
3. Fill in the blank exercise.
4. Quiz

D. Post transcription processing

1. Graphic with hot linked vocabulary.
2. Lecture or text.
3. Fill in the blank exercise.
4. Quiz

E. Translation

1. Table of amino acids with properties.
2. Codon translation table.
3. Open reading frame text/lecture.
4. Quiz.

F. Summary of sequence DNA→pre-mRNA→mRNA→Protein

1. Example of sequences on NCBI showing all four parts and their alignments.
2. Review text/lecture
3. Quiz

G. Proteins

1. Level one structure—Amino Acid sequences
2. Level two structures—alpha-helix and beta-sheets
3. Level three structures—3D folded molecules
4. Level four structures—Domains and Motifs
5. Quiz

H. Sources of Genetic Diversity

1. Sexual reproduction
2. Chromosome alignment in metaphase
3. Crossover and linkage disequilibrium
4. Spontaneous mutation
5. Other

I. Mitochondrial Genomics

Part Two—Comparing and Aligning Sequences

- A. Statistics, Probability, and the Box Model.
- B. Comparing nucleotide sequences
 - 1. Simple scoring schemes (fixed alignment and no gaps)
 - a. Examples of scores—Excel sheets that compute the scores.
 - b. Box model
 - c. Adjusting the weights.
 - 2. Alignment problem
 - a. Global search strategy and its drawbacks
 - b. Examples of global searches using Excel
 - c. Alignment searches with minimal word starters
 - d. Examples of minimal word searches
 - e. Choosing a minimal word size
 - 3. Allowing gaps
 - a. Gap penalty concept
 - b. Default penalties
 - c. Choosing you gap penalties
- C. Comparing amino acid sequences
 - 1. Amino acid similarities and differences
 - 2. Postulated protein evolution
 - 3. The PAM matrix
 - 4. Markov process primer
 - 5. The Blossum matrices
- D. Scoring amino acid comparisons
 - 1. Which Blossum to use and why
 - 2. Understanding the bit score and the box model
 - 3. Expected values and probability
 - 4. The alignment problem in amino acid comparisons
- E. Amino acid to nucleotide sequence comparisons
 - 1. Reading frames
 - 2. The six way comparison of two sequences

Part Three--NCBI Website

- A. Brief tour
 - 1. Mapviewer
 - 2. How to get your sequence
 - 3. Online Medelian Inheritance in Man (OMIM)
 - 4. Tutorials
 - 5. Glossaries
- B. Exercise one
 - ... here we have a series of follow-the-leader investigations similar to the ones already on the NCBI website.

Part Four--Special Topics in Bioinformatics

- A. Orthologs and Homologues
- B. SNPs and Haplotypes
- C. Cladogram generation
- D. Alleles and mutations (a series of case studies)
- E. Gene expression (demo the genepaint website)
- F. Visualizing proteins with 3D models
- G. Electronic PCR

Part Five—Student project/notebook

- A. Targeted examples--at the beginning of the course each student (pair of students?) is assigned the following:
 1. A class chromosome, disease, gene, protein, nucleotide sequence, and amino acid sequence. These will be the same for everyone and the instructor will use these values in the other four parts of the course.
 2. An individual chromosome, disease, gene, protein, nucleotide sequence, and amino acid sequence. These will be different for everyone (every pair?). The student will use her/his individually assigned item to mimic the steps in the other four parts and report the results in her/his notebook.

Additional notes:

Although the bulk of the outline is in parts one and two, the bulk of the course is in the other three parts. I recommend reliance on Camtasia movies. I have recommended an information theoretic restriction because 1) there is already too much material and 2) there is a tendency to let the actual mechanics of the cell confuse information considerations. These are entirely my own ideas as to what we should be doing. I welcome your suggestions, contributions, corrections, website referrals, brainstorming, grant money, etc. I am especially in need of simple molecular biology graphics. Please feel free to take any of the topics above and compose material for our course.