

BSC5936 Introduction to Bioinformatics

Introduction!

1/16/2003

Course BSC5936: Topic (c) M.S.Chapman

1

Bioinformatics is the study of...

- Structure of Biological information
- & use of the information to infer:
 - Evolutionary systematics
 - Gene structure;
 - How many; distribution; coding vs. non-coding...
 - control;
 - Regulation
 - Molecular structure
 - Function of (hypothetical) gene products
 - Annotation

1/16/2003

Course BSC5936: Topic (c) M.S.Chapman

2

Bioinformatics brings together

- Large databases of biological information
- Computational Techniques of analysis
- Databases
 - Sequence (1D)
 - Genome projects
 - Human, mouse, fruit-fly, slime-mould, pea, rice...
 - - many others
 - 120,000 protein sequences
 - 2003; minimally-redundant; Swiss-Prot
 - Molecular Structure (3D)
 - 20,000 structures, PDB

1/16/2003

Course BSC5936: Topic (c) M.S.Chapman

3

What is it good for? Academic Examples...

- How many genes are there?
 - Initiation / termination motifs
 - Comparative analyses
 - Mouse genome revealed 1,200 more human genes
- How many copies of each gene?
 - What is their diversity?
 - When expressed? (micro-array+sequence data)
- What does the gene do?
 - Homology;
 - Searching for sequence motifs;
 - Structural motifs following 3D prediction

1/16/2003

Course BSC5936: Topic (c) M.S.Chapman

4

Scientific/Technical Applications

- Molecular Biology:
 - Primers; restriction sites; splicing sites...
- Sequencing:
 - Coding regions; open-reading frames...
- Localization; Assays
 - Hybridization markers...
- Expression; localization
 - Localization signal sequences...

1/16/2003

Course BSC5936: Topic (c) M.S.Chapman

5

Future Applications - Medical

- Genetic diseases: Cystic fibrosis; Breast Cancer...
 - Screening
 - Gene Therapy?
- Designer medicine - individualized treatments
- Drugs - new targets

1/16/2003

Course BSC5936: Topic (c) M.S.Chapman

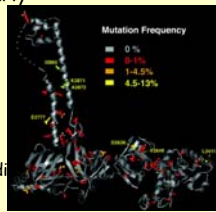
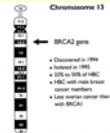
6

Example BRCA-2

- ~1/3 all cancers arise when this tumor suppressor fails
- 1994... BRCA-1 & BRCA-2 Locus from correlating patterns of gene markers w/ familial early onset breast cancer.

➤ Function:

- Biology → DNA repair.
- Alone or in concert?
- Structure → Domain homology:
 - Motifs for nucleotide-binding
 - Ss and ds DNA repair



1/16/2003

Course BSC5936: Topic (c) M.S.Chapman

7

Types of Computation:

- Pattern recognition
 - What things are similar?
 - Motifs; sequences; folds...
- Modeling
 - Extrapolating from existing data
 - Sequence to structure or functional class
 - Structure to function or mechanism
 - Structure to predicting
 - Interaction sites
 - Effects of mutations

1/16/2003

Course BSC5936: Topic (c) M.S.Chapman

8

State of the Art Varies

- Some problems are well characterized
 - Restriction sites, ORFs, pairwise alignment...
 - Established algorithms; commercial software
- Some are for hopeful researchers...
 - Simulating molecular motion; docking...
 - Improvements needed before generally applicable
- Many are in grey area
 - Multi-sequence alignment; Gene annotation; Structure prediction...
 - Useful applications
 - Limits in robustness, predictive power
 - Improvements being developed

1/16/2003

Course BSC5936: Topic (c) M.S.Chapman

9

Computational Challenges - Pattern Recognition

- Inexact matches
 - Substitutions - 1D mostly solved - matrices
 - Insertions / Deletions - 1D pair-wise partially solved - penalties
 - Exchanges, Inversion...
- Order of Complexity
 - N^N to test all gap configurations.
 - $N \log N$ to find best pairwise for given penalties
 - Heuristic kluges for multiple sequences
- Metric for pattern similarity?
- ...and we still have not spoken about 3-D...

1/16/2003

Course BSC5936: Topic (c) M.S.Chapman

10

Why am I here?

- Develop a research interest in improving methods
- Know which program I need to run
- Understand the grey-area methods
 - Which algorithm is most suitable?
 - Often competing programs have very different approaches
 - What is its potential?
 - What parameters are critical and will affect the outcome?
 - How robust are the answers?
 - What other limitations are there?

1/16/2003

Course BSC5936: Topic (c) M.S.Chapman

11