Career Paths for 6/7 Majors

Chris Burge

Dept. of Biology, MIT
# Intersections Between Computer Science and Molecular Biology

## Bioinformatics

Development of tools for mining biological data
- sequences
- expression measurements
- structures
- interactions/networks
- images
- integrated datasets

## Biologically Inspired Computing

- Genetic Algorithms
- DNA Computing
- Synthetic Biology

## Computational Molecular Biology

Application of computational/modeling approaches to address fundamental problems in biology and medicine

## New areas?
Career Paths

- Academic (Research / Teaching)
- Medicine (e.g., medical genetics, medical informatics)
- Government - NIH (NCBI, NHGRI, etc.), national labs, etc.
- Industry - Software, Biotech, Pharma, Instruments/Devices
- Consulting
- Editing / publishing
- Patent law
A Sampling of People and Events in the History of Computational and Systems Biology
The 1970s and Earlier - Sequence Databases, Similarity Matrices and Molecular Evolution

How do protein sequences evolve?
How should similarity between two proteins be scored to most accurately detect homology?

- First protein sequence databases / protein family classification
- PAM matrices for protein sequence comparisons (still used!)

What can molecular sequences tell us about organismal evolution?

- Inferring branching orders and ages
- Use of ribosomal RNA to infer phylogeny
- Discovery of third ‘domain’ of life - Archaea

Margaret Dayhoff

Carl Woese

Russ Doolittle
The 1980s: Sequence Alignment/Search

Which specific residues/positions in a pair of proteins are homologous?
• Smith-Waterman alignment algorithm

What RNA secondary structure has minimum folding free energy?
• Nussinov algorithm
• Zuker algorithm

How to rapidly and reliably find homologs to a query sequence in a sequence database?
• FastA and BLAST algorithms and associated statistics

Temple Smith and Michael Waterman

David Lipman  Bill Pearson  Webb Miller  Stephen Altschul  Samuel Karlin

Ruth Nussinov  Michael Zuker
The ‘90s: Hidden Markov Models, Ab Initio Protein Structure Prediction and the Birth of Comparative Genomics

How to identify domains in a protein?
How to identify genes in a genome?
General framework for such problems?
• Hidden Markov Models

How to study gene expression globally, infer gene function from expression?
• Microarrays and clustering

How to predict protein function by comparing genomes?
• gene fusions, phylogenetic profiling, etc.

How to predict protein structure directly from primary sequence?
• Rosetta algorithm
High-Throughput Biological Experiments

Genomics
Comparative Genomics
Transcriptome Profiling
Genomewide RNAi, synthetic genetic screens
Automated Imaging of Cells, Organelles, Proteins, etc.
High-throughput Protein analysis - mass spec, Y2H, protein arrays, etc.
Microfluidic devices for biological systems
Chemical genetic screening

...
“Next-gen Sequencing”

1. Primer, template, dNTPs and polymerase
2. Sample preparation DNA (5 µg)
3. Template dNTPs and polymerase
4. PCR amplification
5. Break emulsion
6. Template dissociation
7. Cluster growth
8. Fluorescence pulse
9. Time
10. Limit of detection zone

100–200 million molecular clusters

Metzker NRG 2010
Growth of DNA Sequencing Capacity

- NGS (bp/$) Doubling time 5 months
- Hard disk storage (MB/$) Doubling time 14 months
- Pre-NGS (bp/$) Doubling time 19 months

L. Stein 2010
Biologists’ Ability to Generate Data is Surpassing Our Ability to Analyze it Effectively
“The Old Genome Informatics Ecosystem”
“The New Genome Informatics Ecosystem”

“Compute Cloud”

L. Stein 2010
Modeling Gene Networks in the Cell

The Dream of Systems Biology
Educational Paths in Computational Biology

- Bachelors
- M.S. / M.Eng.
- M.D.
- Ph.D. - traditional biology, traditional CS/engineering or interdisciplinary program
- Postdoc / additional training (for academic career)
- Industry careers possible at multiple levels
MIT Computational & Systems Biology PhD Program

Philosophy

Ensure that students have:

• a background in modern molecular/cell biology
• a foundation in quantitative/engineering disciplines to enable them to create new technologies as well as apply existing methods
• exposure to a broad range of subjects emphasizing application of quantitative approaches to biological problems
• emphasizes logic of scientific discovery over mastery of a specific set of facts

Curriculum

• biology
• mathematics/computer science/engineering
• computational and systems biology

http://csbi.mit.edu/education/phd.html
Career Paths

• Academic (Research / Teaching)
• Medicine (e.g., medical genetics, medical informatics)
• Government - NIH (NCBI, NHGRI, etc.), national labs, etc.
• Industry - Software, Biotech, Pharma, Instruments/Devices
• Consulting
• Editing / publishing
• Patent law