

Bioinformatics 2

Introduction

Armstrong, 2010

Bio2 lecture 1

Lecture 1

- Course Overview & Assessment
- Introduction to Bioinformatics Research
- Careers and PhD options
- Core topics in Bioinformatics
 - the central dogma of molecular biology

Armstrong, 2010

Bio2 lecture 1

About us...

- Started in Biology (behaviour genetics)
- Got interested in databases (anatomy)
- Commercial and Academic Experience
- ‘wet lab’ and bioinformatics projects
- Office in forum, Lab in HRB
- Lysimachos Zografos – labs and tutorials

Armstrong, 2010

Bio2 lecture 1

The class (2010)

- M.Sc. Classes:
 - Quantitative Genetics and Genome Analysis (assignment 1 and term paper)
 - Bioinformatics 2 (assignment 1 and exam)

Armstrong, 2010

Bio2 lecture 1

What do I think you know?

- Variety of backgrounds and experience:
 - Biological Sciences
 - Computing Sciences
 - Mathematics, Statistics and Physics

Armstrong, 2010

Bio2 lecture 1

Course Outcomes

- Know the core algorithms in bioinformatics
- Experience in using and/or implementing simple solutions
- Appreciate the current ‘state of the art’
 - what has been solved?
 - what are the key limitations?
- Be familiar with the available resources

Armstrong, 2010

Bio2 lecture 1

Course Design

- Lectures cover essential background
- Guest lectures present research level
- Self-study and assignments designed to cover practical implementation
- Tutorial / Lab support

Armstrong, 2010

Bio2 lecture 1

Assessment (Bio2)

- Written assignment
 - Experimental design and data analysis mini project
 - Plagiarism will be refereed externally
 - Cite all sources!!!
 - Late submissions get 0 marks!

Armstrong, 2010

Bio2 lecture 1

Bioinformatics?

- Introduce yourselves to each other.
 - What is Bioinformatics?
 - What does Bioinformatics do for CS?
 - What does Bioinformatics do for Biology?
 - What guest Bioinformatics lecture would you like?
-
- Discuss in groups for 10 min.

Armstrong, 2010

Bio2 lecture 1

Armstrong, 2010

Bio2 lecture 1

What is Bioinformatics?

- Sequence analysis and genome building
- Molecular Structure prediction
- Evolution, phylogeny and linkage
- Automated data collection and analysis
- Simulations
- Biological databases and resources

Armstrong, 2010

Bio2 lecture 1

BioInf and CS

- Provides CS with new challenges with clear medical significance.
- Complex and large datasets sometimes very noisy with hidden structures.
- Can biological solutions be used to inspire new computational tools and methods?

Armstrong, 2010

Bio2 lecture 1

BioInf and Biology

- High-throughput biology:
 - around 1989, the sequence of a 1.8kb gene would be a PhD project
 - by 1993, the same project was an undergraduate project
 - in 2000 we generated 40kb sequence per week in a non-genomics lab
 - Illumina/Solexa systems Gigabases per expt.

Armstrong, 2010

Bio2 lecture 1

BioInf and Biology

- High-throughput biology
- Data management and mining
- Modeling of Biological theories
- Analysis of complex systems
- Design and re-engineering of new biological entities

Armstrong, 2010

Bio2 lecture 1

bioinformatics

- <http://www.bbsrc.ac.uk/science/grants/index.html>
 - Awarded grants database
- Bioinformatics.oxfordjournals.org
- www.biomedcentral.org/bmcbioinformatics
- www.nature.com/msb

Armstrong, 2010

Bio2 lecture 1

Bioinformatics@ed

- Database integration
- Data provenance
- Evolutionary and genetic computation
- Gene expression databases
- High performance data structures for semi-structured data (Vectorised XML)

1/2

Armstrong, 2010

Bio2 lecture 1

Bioinformatics@ed

- Machine learning
- Microarray data analysis
- Natural language and bio-text mining
- Neural computation, visualisation and simulation
- Protein complex modeling
- Systems Biology
- Synthetic Biology

Armstrong, 2010

Bio2 lecture 1

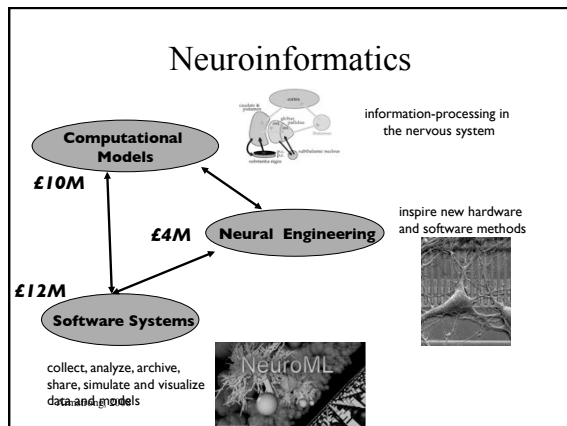
2/2

bioInf activities @ ed

- Self organised reading groups
- <http://www.bioinformatics.ed.ac.uk>
- <http://www.bioinformatics.ed.ac.uk/wiki/SysBioClub/>
 - Regular seminars at 1200 thursdays
- Network science journal club (Lysimachos)
- Edinburgh Proteomics Group (Juri)

Armstrong, 2008

Bio2 lecture 1



Career Options

- Academic Routes
 - Get Ph.D, do Postdoctoral Research - lectureship and independent group
 - M.Sc. RA - becomes semi independent usually linked to one or more academic groups. Career structure is less defined but improving. RAs can do Ph.D. part-time.

Armstrong, 2008 Bio2 lecture 1

Career Options

- Commercial Sector
 - Big Pharma - Accept PhD and MSc entry. Normally assigned to projects and work within defined teams. Defined career structure (group leaders, project managers etc)
 - Spin-out/Small biotech - Accept PhD and MSc entry. More freedom and variety. A degree of ‘maintenance’ work is to be expected.

Armstrong, 2008 Bio2 lecture 1

Career Options

- Hybrid Approaches
 - Commercial and Academic research groups are becoming much closer linked.
 - University academics encouraged to exploit their IPR (intellectual property rights).
 - Companies can get government support to collaborate with academic research groups.

Armstrong, 2008 Bio2 lecture 1

Ph.D.

- Assuming a start date of September 2010
- ‘prize’ studentships advertised on jobs.ac.uk, Nature, Science etc starting NOW!
 - Many linked to nationality/residency (Check details carefully).
- UK ‘quota’ studentships vary with department but contact/apply early.

Armstrong, 2008 Bio2 lecture 1

Ph.D.

- US studentships take longer but are better paid and have extra training/coursework
 - require an entry exam
 - deadlines around summer for ‘11

Armstrong, 2008 Bio2 lecture 1

Bioinformatics 2

Introduction

Armstrong, 2008 Bio2 lecture 1

Slide from <http://www.nd.edu/~networks/>

GENOME
protein-gene interactions
PROTEOME
protein-protein interactions
METABOLISM
Bio-chemical reactions

Armstrong, 2008 Bio2 lecture 1

Slide from <http://www.nd.edu/~networks/>

Lecture 2

PROTEOME

Armstrong, 2008 Bio2 lecture 1

Slide from <http://www.nd.edu/~networks/>

Lecture 2 and later (interaction networks)

PROTEOME
protein-protein interactions

Slide from <http://www.nd.edu/~networks/>

GENOME
protein-gene interactions
PROTEOME
protein-protein interactions

Lectures 3,4
Functional genomics, microarrays and protein-gene interactions

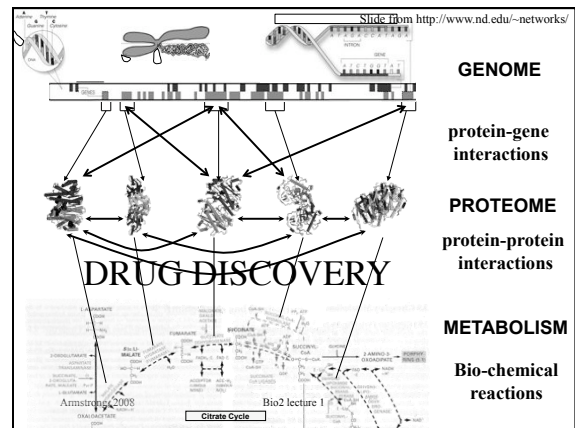
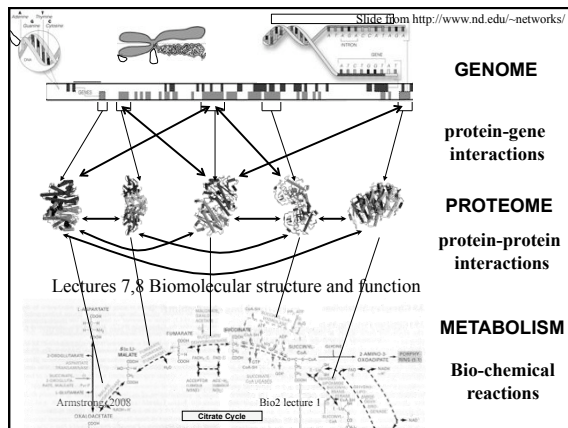
Armstrong, 2008 Bio2 lecture 1

Slide from <http://www.nd.edu/~networks/>

GENOME
protein-gene interactions
PROTEOME
protein-protein interactions

Lectures 5,6 genetics and sequence level bioinformatics

Armstrong, 2008 Bio2 lecture 1



Guest lectures and topics (subject to change)

Yuri Rappsilber (Biological Sciences) Proteomics

Donald Dunbar (Centre for Inflammation Research)
Microarray technologies

Malcolm Walkinshaw (Biological Sciences)
Structure based in silico drug design

Chris Larminie (GalaxoSmithKline)
Bioinformatics in the Pharmaceutical Sector

Armstrong, 2010

Bio2 lecture 1

Preparation work for lecture 2

Understand the 'central dogma of molecular biology'

What is a protein?

What is an amino acid?

How are amino acids joined together to make proteins?

What are the key characteristics of amino acids and proteins
(pH, charge, acidity, hydrophobicity)

What are peptidases? (enzymes that cut protein bonds)

Given a protein sequence how would you predict the weight
of a protein?

Understand that protein sequences vary between species and
are stored/queried from databases.

Armstrong, 2010

Bio2 lecture 1