Bioinformatics

Introduction

David Gilbert

Bioinformatics Research Centre

www.brc.dcs.gla.ac.uk

Department of Computing Science, University of Glasgow

Admin

- Timetable
  - Lectures:
    - Tuesday, 15.00-16.00, University Gardens 7:101
    - Thursday 15.00-16.00, Boyd Orr 513(D)
  - Lab:
    - Tuesday, 10.00-11.00, computer lab, Boyd Orr 618 (level 4)
- Assessment:
  - 1 Coursework (30%)
  - Exam (70%)
- Summer project - optional

- Module information, resources & reading list:

  www.brc.dcs.gla.ac.uk/~drg/courses/bioinformaticsHM
- Course staff
  - Course lecturer: Professor David Gilbert
  - Guest lecturers: Ms Tamara Polajnar, Dr Susan Rosser
  - Course demonstrator: Ms Xu Gu
  - Computing systems support officer: Dr Allan Beveridge

Introductory material

- What is Bioinformatics & why study it
- Brief overview of Bioinformatics
- Current hot topics
- Resources

Why study this module?

- The course is about the application of techniques from computer science to solve problems in molecular biology.
- An exciting area & relatively young field
- Pace of research driven by the large & rapidly increasing amount of data being produced from the life sciences.
- Bioinformatics is **not** number-crunching for molecular biologists, but is about the application of techniques from computer science such as data abstraction, modelling, simulation, machine learning and text mining to analyse biological data.
- The applications include sequence analysis, the prediction and analysis of protein structure and function, and the simulation and analysis of biochemical systems.

Why study this module?

- Includes the latest hot topics in the field, the focus of very exciting research programmes in the University of Glasgow
- **Systems Biology** studies the relationships and interactions between various parts of a biological system in order to understand how the whole system functions.
- **Synthetic Biology** - the structured engineering of biological systems for useful purposes. We will look at the work of Glasgow's undergraduate team which won the Environment and Sensor prize in last summer's iGEM competition (head-to-head with MIT and Brown University).

Prior knowledge

- The course will focus on computing techniques - design of algorithms and use of programmes and databases used to analyse, organise and display biological data, rather than on biology.

- Specially designed for students who do not have training in the Life Sciences: You do not need to have a biological background to do the module - the course will give you the specific knowledge required.

- It will be supported by members of the Bioinformatics Research Centre, who have backgrounds in biology, bioinformatics and computer science.

Fun!

- A very practical course, where you will be able to put your programming skills to practical use!

- Exploitation of parallelism: students will have access to the Bioinformatics Computing Cluster which comprises 45 Sun X2200 servers each with 2 dual core processors giving 180 CPU cores

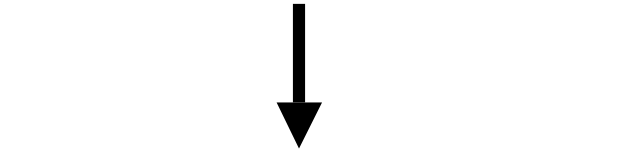
Module contents

- An introduction to the biological background to molecular biology.
- Sequence analysis: algorithms, tools, techniques and databases
- Basic probability theory for bioinformatics.
- Basic concepts of evolutionary theory and phylogenetic analysis.
- Text mining and Information Retrieval for bioinformatics
- Some techniques for the representation and modelling biochemical networks.
- Databases for bioinformatics
- Systems biology & Synthetic biology

Bioinformatics & Systems Biology

DNA

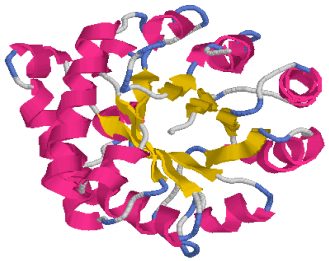
"gene"

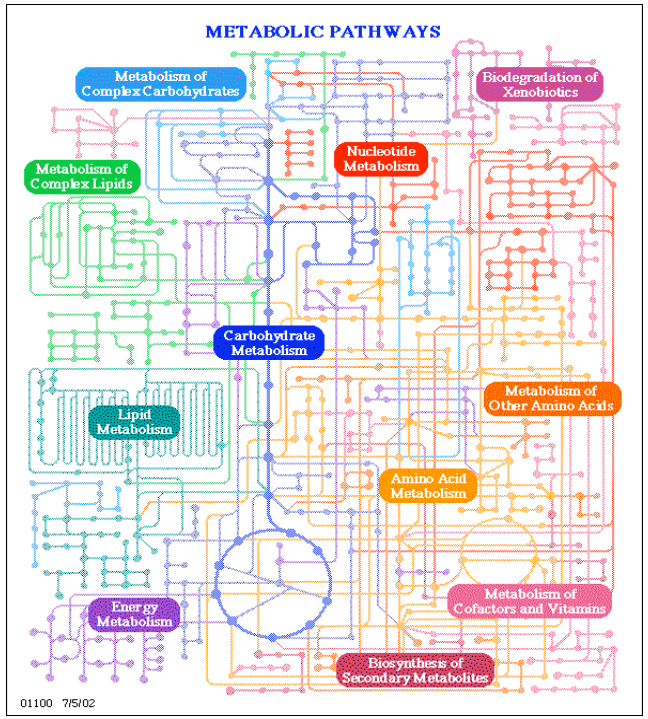


mRNA

Protein sequence

Folded Protein



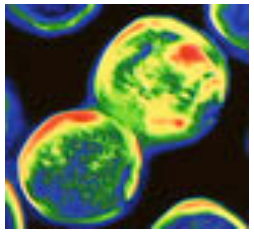


Yersinia pestis

Arabidopsis thaliana

Buchnerasp. APS

Aquifex aeolicus



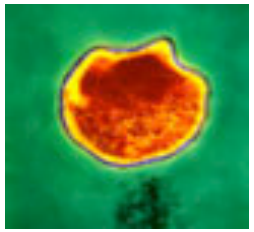
Archaeoglobus fulgidus

Borrelia burgorferi

Mycobacterium tuberculosis

Caenorhabitis elegans

Campylobacter jejuni

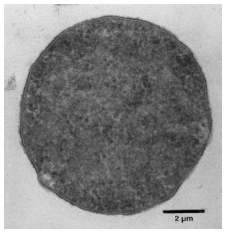


Chlamydia pneumoniae

Vibrio cholerae

Drosophila melanogaster

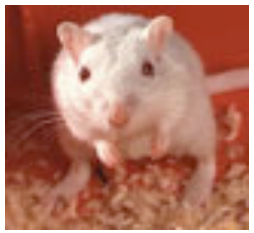
Escherichia coli



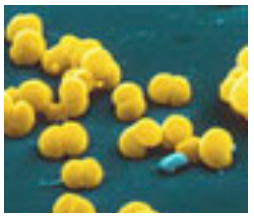
Thermoplasma acidophilum

Helicobacter pylori

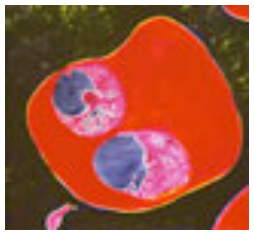
Mycobacterium leprae



mouse

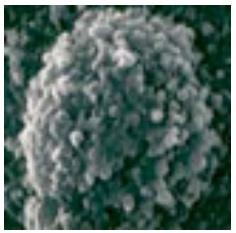


Neisseria meningitidis Z2491



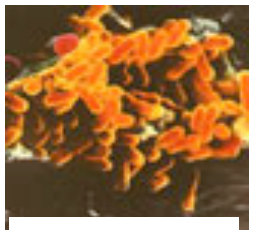
Plasmodium falciparum

Pseudomonas aeruginosa

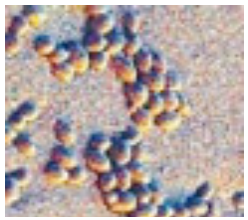


Ureaplasma urealyticum

rat



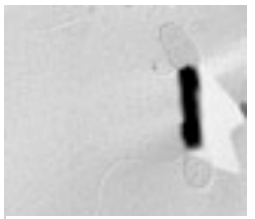
Rickettsia prowazekii



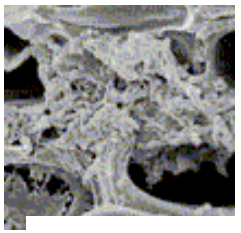
Saccharomyces cerevisiae

Salmonella enterica

Bacillus subtilis

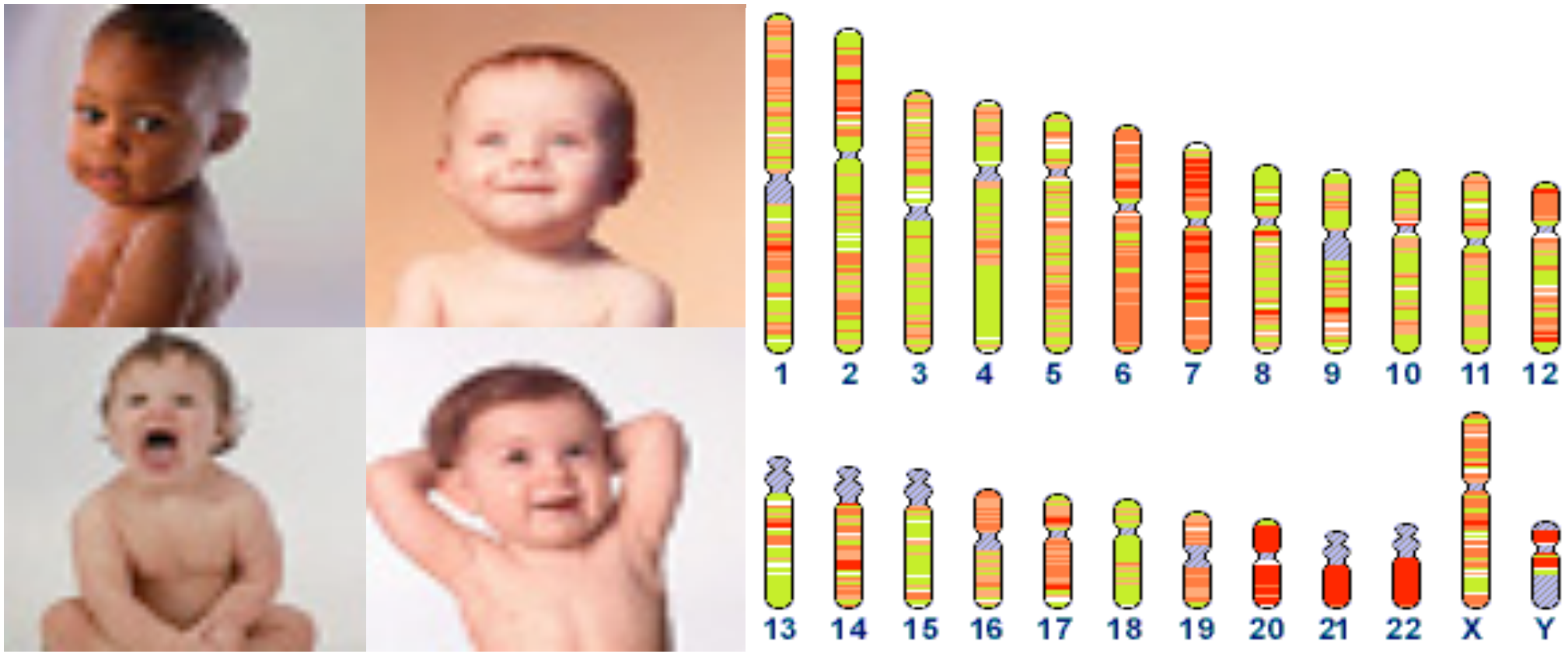


Thermotoga maritima



Xylella fastidiosa

Human Genome



Database Growth

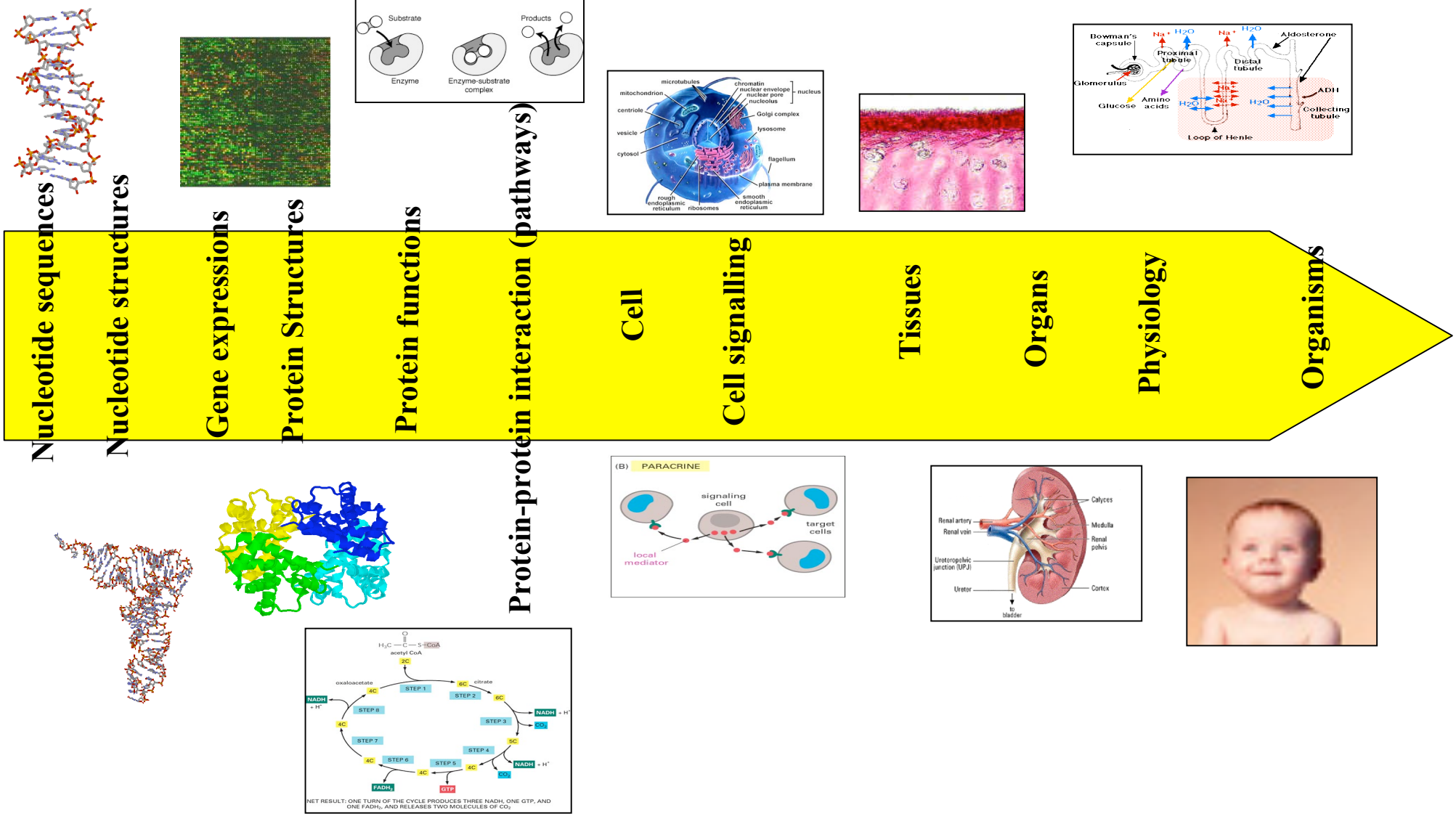
EMBL - sequences

Data deluge is an URBAN MYTH???

D

- Nucleotide Seq (GenBank, EMBL, …)
- Biochemical Pathways (KEGG, WIT…)
- Molecular Classifications (SCOP, CATH,…)
- Motif Libraries (PROSITE, Blocks, …)

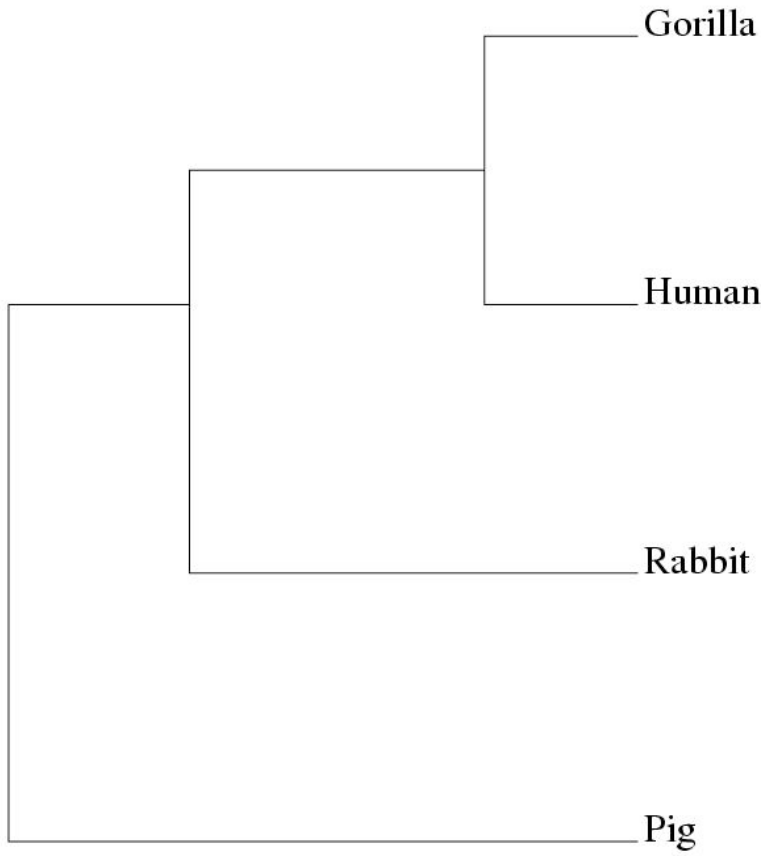
The Complexity of Biological Data



How can we ***analyse*** the flood of data ?

Data: don't just store it, analyse it ! By comparing sequences, one can find out about things like

- How organisms are related & evolution
- How proteins function
- Population variability
- How diseases occur



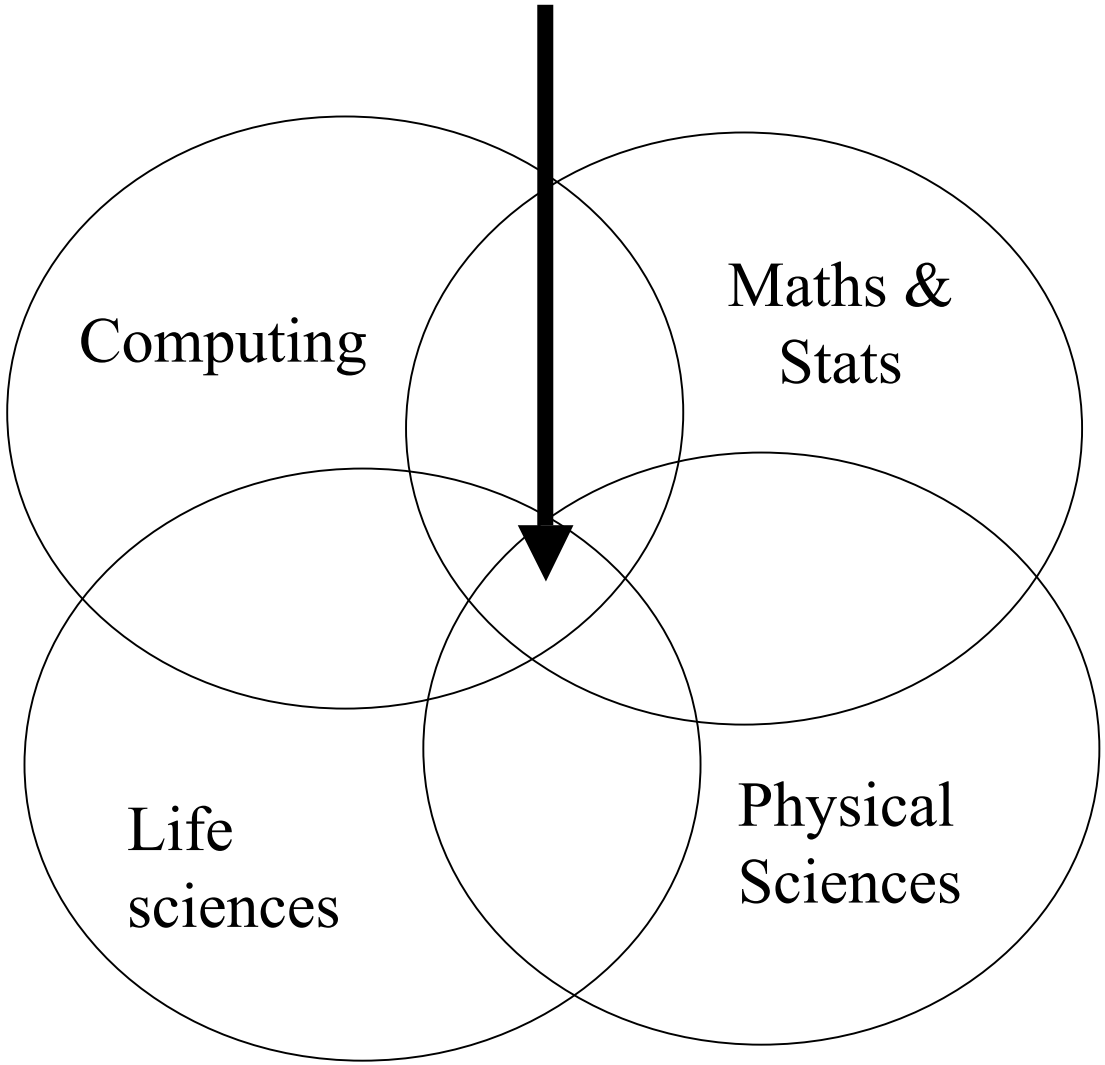
Bioinformatics (Computational Biology - USA)

- Bio - Molecular Biology
- Informatics - Computer Science

- Bioinformatics - the study of the application of
  - molecular biology, computer science, artificial intelligence, statistics and mathematics
  - to model, organise, understand and discover interesting information associated with the large scale molecular biology databases,
  - to guide assays for biological experiments.

- Systems Biology: modelling & analysis of biological *systems* ("putting it all together…")

Bioinformatics in context - a new discipline?

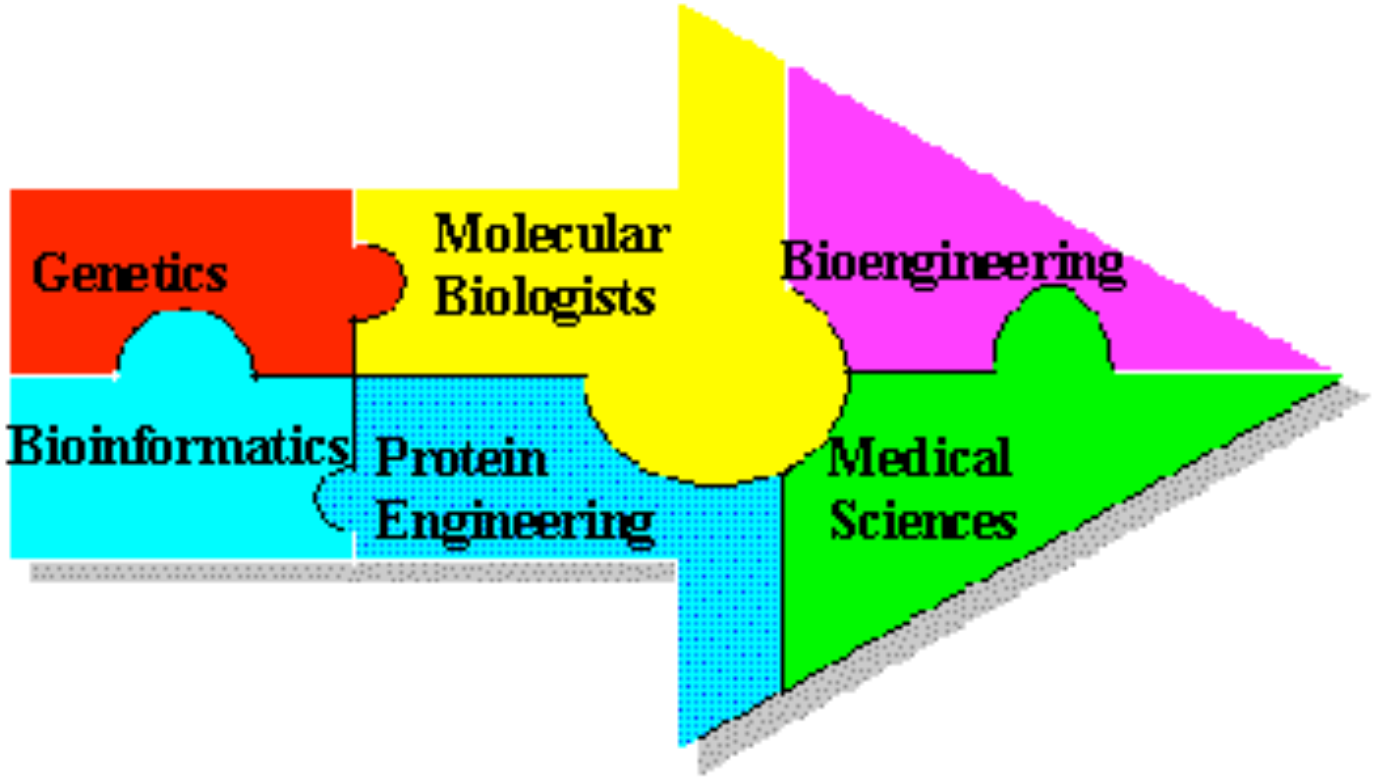


Aim of research in Bioinformatics

Understand the functioning of living things - to "improve the quality of life".

- drug design
- identification of genetic risk factors
- gene therapy
- genetic modification of food crops and animals, etc.
- application to e.g. biotechnology

Bioinformatics in context (applications)

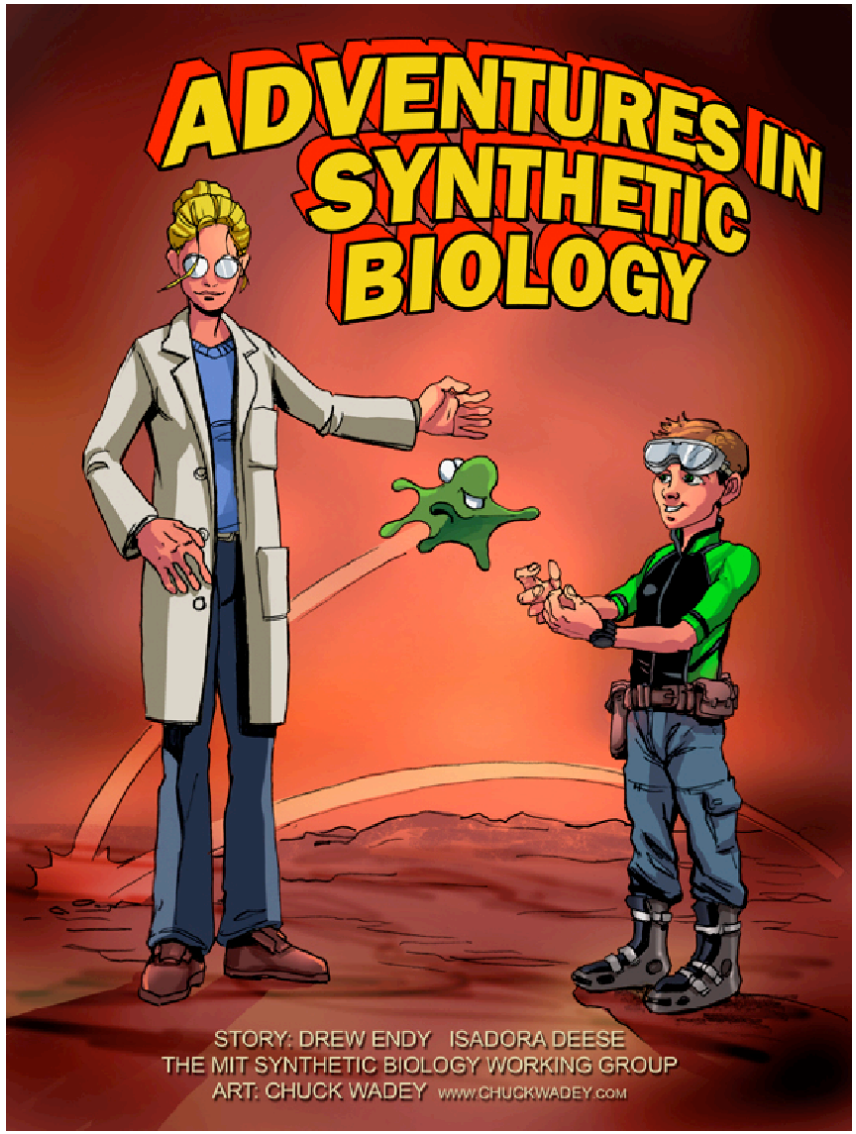


Related but different...

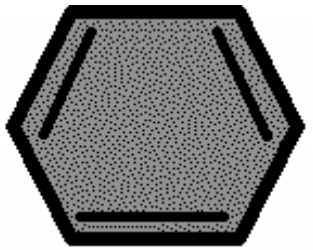
Apply principles from biology to derive novel approaches in computer science:

- biocomputing
- neural computing
- genetic algorithms
- evolutionary computing

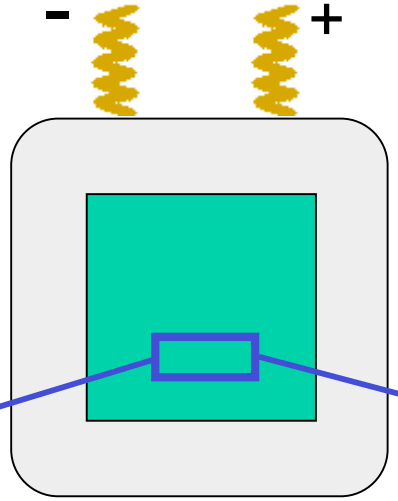
What is synthetic biology?

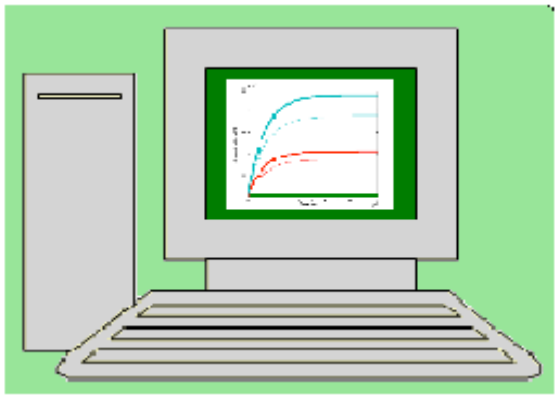


- Design & construction of new biological parts, devices, and systems
- Re-design of existing, natural biological systems for useful purposes
- Involves
  - *Standardisation*
  - *Decoupling*
  - *Abstraction*



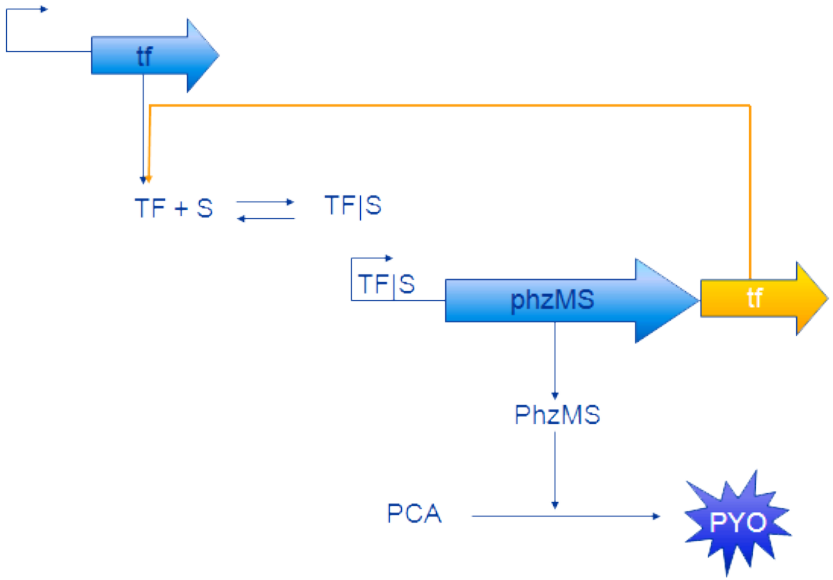
Pollutant





Electrical Output

Microbial Fuel Cell



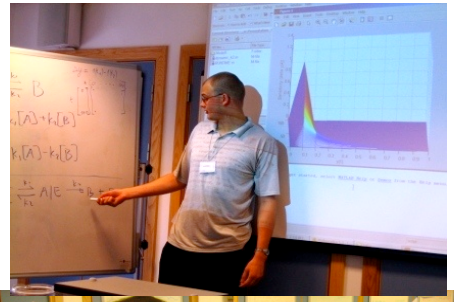
$$\dot{TF} = \alpha_{TF} - \delta_{TF}TF - \beta_{TFS}sTF + k_dTFS + \beta_{TF}\frac{TFS}{\gamma_{TF} + TFS} \quad (1)$$

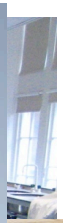
$$\dot{TFS} = \beta_{TFS}sTF - k_dTFS - \delta_{TFS}TFS \quad (2)$$

$$\dot{PhzMS} = \beta_{PhzMS}\frac{TFS}{\gamma_{PhzMS} + TFS} - \delta_{PhzMS}PhzMS \quad (3)$$

$$\dot{PYO} = \alpha_{PYO}PhzMS - \delta_{PYO}PYO \quad (4)$$

PYOCYANIN






Best Enviro & Sensing
Glasgow


iGEM
International Genetically
Engineered Machines
Competition

Support material

Course texts and required reading:

- Introduction to bioinformatics - Arthur Lesk. Publisher: Oxford University Press. Year 2002. ISBN: 0199251967. Category recommended
- Bioinformatics: Sequence and Genome Analysis by David W. Mount. Pubisher Cold Spring Harbor Laboratory Press,U.S. Year 2004. Isbn 0879696877. Category recommended
- Fundamental Concepts of Bioinformatics Krane & Raymer. publisher: Benjamin Cummings. year: 2002. isbn: 032119022X. category: recommended but note difficulties in obtaining it
- Post Genome Informatics Kanehisa. Publisher OUP. Year 2000. Isbn 0198503261. Category background
- An Introduction to Bioinformatics (Attwood & Parry-Smith). Recommendation: Useful

Other texts of interest:

- Bioinformatics: An Introduction for Computer Scientists, J. Cohen, ACM Computing Surveys, 36(2), 122-158, 2004.
- Protein Bioinformatics, I. Eidhammer, I. Jonassen and W. Taylor, Wiley 2004
- Developing Bioinformatics computer skills, C. Gibas and P Jambeck, O'Reilly, 2001
- Algorithms on strings, trees and sequences, Dan Gusfield, CUP (1997+).
- Biological sequence analysis, R. Durbin, S. Eddy, A.Krough and G. Mitcheson, CUP, (1998+).
- Introduction to biological computing, J Setubal and J Meidanis, PWS publishing company, 1997.
- The computational linguistics of biological sequences. David B. Searls. In Larry Hunter, editor, Artificial Intelligence and Molecular Biology, chapter 2, pages 47- 120, AAAI Press, 1993.
- Introduction to Computational Biology, Michael Waterman. Chapman & Hall, 1995
- Bioinformatics for Dummies, Jean-Michel Claverie, Cedric Notredame, Jean-Michel Claverie, Cedric Notredame, 2003

Resources

- www.brc.dcs.gla.ac.uk/~drg/bioinformatics/resources.html
- www.ebi.ac.uk/microarray/biology_intro.html
  - Very good introduction to molecular biology for computer scientists.
- www.ebi.ac.uk/2can
  - Bioinformatics educational resource at the EBI
- International Society for Computational Biology: www.iscb.org
  - very good rates for students, and you get on-line access to the Journal of Bioinformatics.