

MS2a: Bioinformatics and Computational Biology - 16MT

Recommended Prerequisites

None. In particular, no previous knowledge of Genetics will be necessary.

Aims & Objectives

Modern molecular biology generates large amounts of data, such as sequences, structures and expression data, that needs different forms of statistical analysis and modelling to be properly interpreted. The fields of Bioinformatics and Computational Biology have this as their subject matter and there is no sharp boundary between them. Bioinformatics has an applied flavour while Computational Biology is viewed as the study of the models, statistical methodology and algorithms needed to do bioinformatics analysis. This course aims to present core topics of these fields with an emphasis on modelling and computation.

Synopsis

- Fundamental Data Structures in Biology: Sequences, Genes, Networks and RNA secondary structure.
- Stochastic Models of Sequence and Genome Evolution including models of single nucleotide/amino acid/codon evolution.
- Phylogenies: enumerating phylogenies, the probability of sequences related by a specified phylogeny, the minimal number of events needed to explain a data set (Parsimony).
- Likelihood and algorithms (Markov Chain Monte Carlo) for inference based on the likelihood.
- Software packages for sample-based inference.
- Alignment Algorithms. Comparing 2 strings, an arbitrary number of strings, find segments of high similarity in 2 strings.
- Common Patterns in a set of Sequences.
- Network Inference and Network Evolution
- Detection of Recombinations in Sequences

First lecture is Wednesday 15th 11AM. Weekly lectures (60 minutes) Wednesday 11AM and Thursday 11AM. 1 weekly exercise session (60 minutes) at time to be decided.

Method of Assessment

It is proposed to assess this course by mini-project.

Further Reading

C. Semple and M. Steel, *Phylogenetics*, Oxford University Press (2003).

Durbin et al., *Biological Sequence Analysis*, Cambridge University Press (1998).

T. Jiang et al., (editors) *Current Topics in Computational Biology*, MIT Press, (2003).

M. S. Waterman et al., *Computational Genome Analysis: An Introduction*, Springer (2004).

More information about bioinformatics can be found at <http://www.stats.ox.ac.uk/research/genome/>