
CO4200 Algorithms for Bioinformatics

Credits: 15 **Convenor:** Dr S Fung **Semester:** 2nd

Prerequisites: *Desirable: Java programming*

Assessment: *Coursework: 40%*

1.5 hour exam in May/June: 60%

Lectures: 24 hours

Problem Classes: 4 hours

Laboratories: 12 hours

Private Study: 72.5 hours

Subject Knowledge

Aims This module introduces students to the algorithmic solution of computational problems in bioinformatics, including aspects of probabilistic modelling.

Learning Outcomes Students should be able to: describe a number of computational problems arising in bioinformatics; state and discuss algorithmic approaches to the solution of such problems; discuss probabilistic models underlying computational tasks in bioinformatics; design and implement efficient algorithms; apply modelling and algorithm design to the solution of bioinformatics problems.

Methods Class sessions together with course notes, recommended textbooks, worksheets, and some additional hand-outs and web support.

Assessment Marked problem-based worksheets and programming assignments, traditional written problem-based examination.

Skills

Aims To teach students scientific writing, modelling and problem solving skills.

Learning Outcomes Students will be able to: write short, clear, note based, summaries of technical knowledge; solve abstract and concrete problems (both routine seen, and simple unseen).

Methods Class sessions together with worksheets.

Assessment Marked problem-based worksheets, traditional written examination.

Explanation of Prerequisites A basic understanding of discrete mathematics and probability will be helpful.

Course Description Processing biological data requires complex computations on large volumes of data. To ensure that these computations complete within a reasonable amount of time, one must design the algorithms (computer procedures) after a careful study of the characteristics of underlying data and making use of existing algorithm design principles. This module aims to introduce students to the algorithmic solution of computational problems in bioinformatics. Students will learn a number of probabilistic models that underlie the formulation of biological data processing tasks as computational problems, and will be introduced to efficient computer algorithms for solving, and some key principles for designing efficient algorithms for solving, these problems.

Detailed Syllabus

- Introduction to algorithms
- String matching
- Pairwise sequence alignment
- Hidden Markov models
- Restriction site mapping

- Multiple sequence alignment
- Phylogenetic trees
- Genome rearrangement

Reading List

- [B] R. Durbin, S. Eddy, A. Krogh, G. Mitchison, *Biological sequence analysis – Probabilistic models of proteins and nucleic acids*, Cambridge University Press, 1998.
- [B] N.C. Jones, P.A. Pevzner, *An introduction to bioinformatics algorithms*, MIT Press, 2004.

Resources Web page, study guide, worksheets, handouts, lecture rooms with OHP and data projector.

Module Evaluation Course questionnaires, course review.