NORBIS

The national doctoral school in bioinformatics, biostatistics and systems biology

NORBIS

An introduction to sequence comparison and database search

Time and place: November 16-20 2015, University of Bergen, Norway
Taught by: Inge Jonassen, Cedric Notredame, Des Higgins
ECTS: 5

Course description

This course gives an introduction to methods for aligning biological sequences. Its goal is to present an overview of the basic concepts of sequence alignments and some of their applications with a strong emphasis on homology based multiple sequence alignment modelling, one of the most widely used method in biology.

The first part is dedicated to **molecular evolution**. We will focus on the implications of molecular evolution on sequence variation. We will use these concepts to define homology. We will then see how specific mathematical models (the substitution matrices) have been derived in order to quantify the evolutionary relationship between sequences. In the next part we introduce the Needleman and Wunsch algorithm (Dynamic programming), a very basic algorithm that makes it possible to derive pairwise alignments from the sequences while using the substitution matrices. Next, we will see how these pairwise alignment methods can be applied to database searches and we will develop the main concepts behind the BLAST algorithm. We will finally introduce the notion of multiple sequence alignment and show how a group of related sequences can be compared in order to infer common properties. We will then see the main principles behind two multiple sequence alignment package: the Clustal programs and TCoffee and the current challenges when modelling sets of homologous sequences (RNA and proteins).

Course program

The course will be given over a week (5 days - Monday-Friday) with lectures (3-4 hours including discussions) in the mornings and practical hands on sessions in the afternoons. On the last day there will be a written exam (two hours) and a summing up section where the students can provide feedback. The students will receive a reading list before the course and are expected prepare well for the course. In addition to the exam, the students will do a project after the course and deliver a report within a month.

Prerequisites

The practicals will involve using command line programs and some familiarity with Linux is expected. We expect students to be familiar with basic concepts in programming and algorithms. Basic knowledge in linear algebra and statistics is expected.

Learning outcomes and competence

Students will learn the principles underlying sequence comparison including an evolutionary understanding of sequence alignments, development and use of substitution matrices, and the heuristic methods for database sequence homology search in the Blast programs. Students will also gain an understanding of the concept of multiple sequence alignment, and their applications. An important focus of the course will be the detailed understanding of evolutionary based multiple sequence comparison methods, including the Clustal and TCoffee software series. This will make the students able to integrate various sources of data (protein sequences, protein 3D structures, RNA sequences, RNA 2D/3D models) into high quality multiple sequence models thus offering them an entry point into homology modelling for evolutionary, structural and functional analysis.

Evaluation

The students will go through a two hours written exam. In addition the report will need to be approved. Grades: pass / no pass.

Note: Traveling and accommodation expenses shall be covered by NORBIS for the PhD students affiliated to our school.