

Exercise session 2

[Introduction to bioinformatics](#), Autumn 2009

Group 1: Thursday 24.9 12-14 Exactum BK106

Group 2: Thursday 24.9 16-18 Exactum BK106.

General instructions:

Problems for each exercise session will be distributed approximately one week before the session. You are expected to be prepared to present your solutions in the exercise session.

In addition, you need to send notes of the assignments you are going to mark to [Laura Langohr](#) by email before exercises (Thursdays 12.15).

These exercise notes should contain a brief description of the steps you took to solve the assignment, as well as the results. Important: When sending email, use subject of form "ITB exercise X, where X is the exercise session number (1/2). Send your notes in email text body. If you need to include a figure, send it as an attachment.

Assignments

1. In [PubMed](#), search for articles published during the last three months about whole genome sequencing.
 - How many articles discuss about human genome?
 - You can notice that in articles concerning the human genome, the interest is in human diseases. On what kind of diseases?
 - List other (than human) organisms on which whole genome sequencing has been published during the last three months. (You don't have to give a complete list, just show that you understand the question.)
2. By using NCBI facilities, find out:
 - from how many Eukaryotic species genome sequences have been completed.
 - In addition, there are Eukaryote genome sequences under "assembly" or in "progress". How many?
 - Explain briefly, what is the difference between completed, under assembly, in progress.

3. By using NCBI, find out and explain briefly what are the statuses of cattle (*Bos taurus*) and horse (*Equus caballus*) genome projects. Are they published?

4. Accession number to dog insulin gene sequence is NM_001130093. On the basis of this information, construct a FASTA-file consisting of dog, cat, swine, human, chimpanzee and rabbit insulin sequences. Align the sequences by using Clustal. One possibility is to use one of the free servers, for example <http://www.ebi.ac.uk/Tools/clustalw2/index.html>, another is to download Clustal (X or W) to your own computer from <http://www.clustal.org/>. Your answer to this should only include a brief explanation about what you have done and some general description about the alignment, for example how many nucleotide (base) differences between human, chimpanzee, rabbit etc. Please, do not send the alignment by e-mail, but be prepared to show it in the class. When you perform the alignment, use the default settings and do not pay attention to various scoring, gap penalty etc. options. This kind of things will be taught during the lectures. In the present exercise you just familiarize, a little bit, to alignment as a practical concept.