

## Exercise session 5

[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. The most widely used program package for easy phylogeny constructions is the MEGA-software. Read the paper which describes its latest version: Tamura, K. et al., 2007. Be prepared to give a short (~5 minutes) presentation about the paper. An answer to this, when sending notes to Laura, is that you have read the paper: yes, I have read the paper.

2. Download the MEGA4-software to your computer from <http://www.megasoftware.net/mega.html> and find out what kind of phylogenies can be constructed with this program package. The general rough classification of different methods is: parsimony methods, distance matrix methods, maximum likelihood methods and Bayesian methods. Phylogenetic trees have topologies, and branching patterns, but uncertainty is present. What is MEGA's method for statistical confidence or credibility inference (i.e. how is the uncertainty noticed).

3. When you inspect the contents of MEGA4, you notice that it includes a set of example datasets. Select one of them and perform neighbor-joining, UPGMA and parsimony phylogeny analyses without bootstrapping and including 1000 x bootstrapping. Please, read the MEGA-tutorial ("help"). Skip most of the options, e.g. nucleotide substitution models, patterns etc., as they are out of the scope and topics of this Introduction to bioinformatics course; just use the default settings. Describe what kind of differences you find between parsimony, neighbor-joining and UPGMA phylogenies, as well as between no-bootstrapping and 1000 x bootstrapping.

4. M27325 refers to a gene sequence in a cow. Collect a dataset from ~15-20 animal species, perform phylogeny analyses by MEGA, and describe the results. What animals (on the basis of this gene) form clear and statistically supported clusters, what animals are clustered together, but not in a statistically supported way.

Note that when you have collected the data, you must align the sequences and construct a file in a format acceptable by MEGA. You'll find all instructions from MEGA-tutorial. Note also that MEGA4 includes the aligning facility. You can also do the alignment by other means (Clustal by webservice or by Clustal in your own computer, if you have downloaded it for previous purposes. Aligning by Clustal included in MEGA is sometimes difficult, in practice, although it should be easy.)