

# Biometry and bioinformatics II, fall 2014

## Biometry and bioinformatics II

---

### Lecturer

Sirkka-Liisa Varvio

### Scope, type and prerequisites

- 5-10 cr. Assignments given at the beginning of the course is the 5 cr part of the course, and extension 5 cr -> 10 cr is by doing more assignments.
- In statistics curriculum: intermediate / advanced studies.
- Prerequisites: At least intermediate level of probability and statistical inference courses and basic familiarity to working with sequence data (such as provided by the course Biometry and bioinformatics I, for example). The spirit of the course is learning by doing assignments.

### Short description of contents

The ultimate goal is to provide understanding of life through molecular information, via statistical and computational analyses. Models of molecular evolution and phylogeny inference, going further from the basics which were in the course Biometry and bioinformatics I: maximum likelihood, bayesian and network inference. Past, or traditional, practice on phylogenetics has been to describe relationships among species within the disciplines systematics and taxonomy. Today the scope is much wider. Phylogenetics is used as a necessary tool in all kind of bioscience research questions: relationships among genes, gene families, alleles, cell lines, various genomic components, histories of populations, evolutionary and epidemiological dynamics of bacteria and viruses, etc. Phylomics is the natural extension of genomics and metagenomics, phylotyping the current way of understanding ecosystems within and outside organisms (environments) and evolutionary thinking is a complementary framework to medical genetics. Statistical and computational approaches, based on predictions derived from population genetics theory, aim at dissecting neutral (stochastic) and selected (darwinian natural selection) components of genetic diversity.

### Time schedule

The first session will be Tuesday 4.11, 14-17, computer class C128 (not 28.10 as the exam of Biometry and bioinformatics I is still going on). The assignments for 5-10 cr part of the course will be launched during the sessions 4.11 and 11.11. Working after these two common sessions can be accomplished according to personal, and negotiable, time schedules. There will be no exam. Grading is based on the quality of assignments and pass/fail is also possible.

### Lecture material and scientific publications

- [Modelling molecular evolution.pdf](#), Literature:
- [Phylogeny inference methods.pdf](#), Literature: [Phylogenetics - Principles and practice.pdf](#), [Short tutorial article.pdf](#), [Original NJ-algorithm paper.pdf](#), [Original MrBayes paper.pdf](#), [Traditional and Bayesian phylogeny estimation.pdf](#), [MCMCMC.pdf](#), [Felsenstein\\_Confidence limits on phylogenies.pdf](#)

## Assignments

You can choose which assignments you perform. Assignments 1, 2, 3 and 4 are 1 cr, each, assignment 5 will be 1-3 cr (depending on the amount of work you want to do). In addition, there will be essay topics (1 cr = 5 page essay about a topic, Assignments 6). You can also tailor a molecular evolutionary project for yourself about an interesting gene, or some other genomic compartment.

- Assignment\_1\_BB\_II.pdf
- Assignment\_2\_BB\_II.pdf
- Assignment\_3\_BB\_II.pdf
- Assignment\_4\_BB\_II.pdf
- Assignment 5 will be about EBOLA. Read these: [EBOLA phylogenetics.pdf](#) , [EBOLA\\_Phylogenetics\\_2.pdf](#), and start getting familiar with the existing data in NCBI nucleotide database! [Recent news\\_ Delays hinder EBOLA genomics.pdf](#)
- **Assignment 6, topics:**
  - Phylo\_history\_1.pdf, Phylo\_history\_2.pdf
  - Protein\_evolution\_1.pdf, Protein\_evolution\_2.pdf, Protein\_evolution\_3.pdf, Protein\_evolution\_4.pdf, Protein\_evolution\_5.pdf
  - Phylo\_cancer\_1.pdf, Phylo\_cancer\_2.pdf, Phylo\_cancer\_3.pdf, Phylo\_cancer\_4.pdf, Phylo\_celltypes.pdf, Phylo\_languages.pdf
  - Phylo\_virus\_1.pdf, Phylo\_virus\_2.pdf, Phylo\_bacteria\_1.pdf, Phylo\_bacteria\_2.pdf, Phylo\_lateral\_gene\_transfer\_1.pdf, Phylo\_lateral\_gene\_transfer\_2.pdf
  - Phylo\_mammals\_1.pdf, Phylo\_mammals\_2.pdf
  - Phylo\_metagenomics\_methods.pdf, Phylo\_human\_metagenomics\_1.pdf, Phylo\_human\_metagenomics\_2.pdf, Phylo\_env\_metagenomics\_1.pdf, Phylo\_env\_metagenomics\_2.pdf
  - Phylo\_statistics\_1.pdf, Phylo\_statistics\_2.pdf, Phylo\_HMM\_2.pdf

## Datafiles

- data\_1.txt
- data\_2.txt
- data\_3.txt, data\_3\_aligned.txt, data\_3\_aligned\_IUPACedited.txt
- data\_4.fasta
- data\_HLA\_gene.fasta

## Software, servers, links to databases

- <http://www.ncbi.nlm.nih.gov/>, Literature: [NCBI BLAST.pdf](#),
- ClustalX2 (in C128), <http://www.ebi.ac.uk/Tools/msa/clustalw2/>, Literature: [Multiple sequence alignment.pdf](#),
- <http://mafft.cbrc.jp/alignment/server/>
- MEGA5 (in C128), Literature: [MEGA5 original paper.pdf](#),
- DnaSP5-software (in C128),
- MrBayes-software (in C128), Move the file (the training file primates.nex) to Z, and when strating MrBayes: execute z:\primates.nex, [MrBayes Manual 3.1.pdf](#), [MrBayes Commands.pdf](#), [This is useful clarification.pdf](#),
- [Link to data-conversion](#) (use this, or use Clustal, for obtaining a nexus-format from the data so that it can be submitted to MrBayes)
- [Link to choosing a model](#) (note also model choice option in MEGA5) [Modeltest.pdf](#) , [This is useful, AIC.pdf](#)
- SIFT (Sorting Intolerable From Tolerable) and PolyPhen (prediction of functional effects of human nsSNPs)

## BB\_II\_MOODLE

## Background literature, e-books in HU-library

- Z Yang, Computational molecular evolution
- P Lemey et al, The phylogenetic handbook
- M Nei and S Kumar, Molecular evolution and phylogenetics

---

## Registration

Did you forget to register? [What to do?](#)

---

