

Comparative Genomics course, HT07, 1: Basic genome analysis

Lab instructions here. Questions can be directed to Kristoffer Forslund, krifo at-sign sbc dot su dot se.

Part 1 - where to go?

- Look over the various tools mentioned in the toolbox and try them out.
- Begin work on the first programming assignment.
- Try (optional, but you need to learn how to do this for later exercises) to write a script to automatically download, say, all genes from a given species from a database.
- Try (optional, but you need to learn how to do this for later exercises) to write a script that can translate the sequence names in a FASTA file according to a map file (on the form [oldname], [newname]) and back again.

Part 2 – classify your genomes!

- Use online BLASTing, for instance at NCBI, or similar tools, to classify your genomes. Take, say, a few thousand bases and see if you can find out where they are from.
- That is, try to find out which genome comes from which species, or at least, investigate what type of organism they are.
- How many bacteria are there? How many archaea or eukaryotes?
- Also, one or two genome files may be strange somehow. Can you find which, and in what sense?
- Try to determine whether what you have is the whole genome or only a part of it.