

# Graded Assignment Comparative Genomics

This is an individual assignment, **do not copy scripts, data or method descriptions** from fellow students, publications or the internet. Hand in your files through the Turnitin assignment. Turnitin automatically checks for plagiarism with files from the internet, scientific publications as well as other students' assignments.

Write a report about your task including Introduction (1 paragraph) Methods, Results, Discussion, References. Include the scripts and first ten lines of result files as appendices to the report as text in the pdf (NOT as screenshots). Make one pdf with your report and your scripts included as appendices to upload.

1. Create a multiple sequence alignment and phylogenetic tree for a gene family with an interesting phylogenetic pattern (duplications, horizontal gene transfer, domain fusion, etc).
  - a. Download the protein sequence
  - b. Run BLAST to find homologs in a diverse set of species (minimally 25)
  - c. Create a multiple sequence alignment and phylogenetic tree with bootstrap values.
2. For the same set of species as in (1.) create a species tree. Think about appropriate data (16S, ITS regions, concatenated alignments)
3. Annotate the tree from (1.) with duplication, speciation and HGT events (where applicable). Discuss which genes/proteins are orthologs, paralogs etc.
4. From the Multiple sequence alignment in (1c) study conserved regions, discuss the functions of these regions.

Suggestions for “interesting” proteins:

Gcn5-Related *N*-Acetyltransferases

Frizzled receptors

Proteasome catalytic subunits

Exosome

RNA polymerase B-C

tRNA synthetases

reverse gyrase (bacteria/archaea)

Hox genes