

## **Graded Assignment Comparative Genomics**

Compute sets of orthologs for two species for which a whole genome sequence is available. Every student has their private set of species (Table 1).

- 1. Compute orthologs using Best Bidirectional Hits**
  - a. Download protein sequences for your two species.
  - b. Run BLAST for protein sequences of your two species (locally on your PC or on the HPC [https://blast.ncbi.nlm.nih.gov/Blast.cgi?CMD=Web&PAGE\\_TYPE=BlastDocs&DOC\\_TYPE=Download](https://blast.ncbi.nlm.nih.gov/Blast.cgi?CMD=Web&PAGE_TYPE=BlastDocs&DOC_TYPE=Download) )
  - c. Create a python script that generates BBH for your two species. Pay attention to the pitfalls that have been mentioned in class.
  - d. How many protein coding genes do your genomes contain? How many orthologs do you identify?
- 2. Check orthology with a phylogenetic tree.** Pick an example of orthologs from [1] of co-orthology.
  - a. Retrieve homologs to sequences from at least 25 different species (including your 2).
  - b. Create Multiple Sequence Alignment and a phylogeny for these homologs. Include bootstrapping.
  - c. Visualize the tree including branch lengths and bootstrap values.
  - d. Create a species tree for the 25 species using one of the methods discussed in class.
  - e. Identify Speciation and Duplication events in the phylogeny of the homologs.
  - f. Discuss the conclusions about the orthology case.
- 3. Identify functional regions.**
  - a. Write a python script to identify conserved regions in the Multiple Sequence Alignment from [2] with one of the sequence conservation methods discussed in class.
  - b. Discuss the conserved regions.
- 4. BONUS\* Conservation of promotor regions.**
  - a. Identify regulatory regions of the orthologs studied in [2] and [3].
  - b. Design a strategy to identify conservation in these regions.

Write a report about your task including Introduction (1 paragraph) Methods, Results, Discussion, References. Submit the report together with scripts and result files on Toledo.

Table 1 Students and species

<b>Name</b>	<b>Species1</b>	<b>Species2</b>
Darina Abaffyova	Prevotella buccalis	Chlorobium limicola
Jose Ignacio Alvira Larizgoitia	Arabidopsis thaliana	Solanum tuberosum
Nnamdi Joseph Asouzu	Hydra vulgaris	Caenorhabditis elegans
Aditya Badola	Apis mellifera	Latimeria chalumnae
Sara Becelaere	Streptococcus pneumoniae R6	Agrobacterium fabrum
Tim Blokker	Thermotoga maritima	Thermus thermophilus HB27
Emery Alexander Bosten	Verrucospora maris	Moorella thermoacetica
Danie Daaboul	Vibrio cholerae	Oceanicola batsensis
Ben De Maesschalck	Pneumocystis murina	Thalassobacter stenotrophicus
Arthur Dogot	Methanosarcina barkeri	Aeropyrum pernix
Mehmet Erim Erdal	Listeria innocua	Ureaplasma parvum
Peiyan Feng	Bartonella henselae	Photobacterium profundum
Biancamaria Florenzi	Robiginitalea biformata	Blastopirellula marina
Kevin Francis Menezes	Loktanella vestfoldensis	Maritimibacter alkaliphilus
Jaldert Francois	Deinococcus geothermalis	Nitrobacter hamburgensis
Sairam Ganapathy	Rickettsia tamurae	Mycoplasma hominis
Emma Gheysen	Pseudomonas putida	Bartonella bacilliformis
Seppe Goovaerts	Sphingomonas wittichii	Kordia algicida
Dominique Holtappels	Staphylothermus marinus	Lactococcus lactis cremoris
Jonas Jonker	Clostridium spiroforme	Neisseria mucosa
Prabhat Juyal	Blautia hydrogenotrophica	Nakamurella multipartita
Shemy Khaled Mohamed Ezzat Ahmed	Sphaerobacter thermophilus	Thermosiphon africanus
Joon Klaps	Escherichia coli	Collinsella intestinalis
Tine Logghe	Lactobacillus reuteri	Geoglobus acetivorans
Hannelore Longin	Acetohalobium arabaticum	Hirschia baltica
Blanca Lorente	Thermococcus	Shuttleworthia

Echeverria	gammatolerans	satelles
Lucy Anna Mee	Xenophilus azovorans	Zavarzinella formosa
Joren Mommaerts	Agrobacterium rhizogenes	Chlamydia trachomatis
Kristen Michelle Nader	Staphylococcus aureus	Streptococcus pyogenes
Inne Nauwelaers	Bacillus cereus	Lactobacillus animalis
Michael Shawn Neilson	Pimelobacter simplex	Emiliana huxleyi
James O'Reilly	Malus domestica	Cucumis melo
Pinar Onat	Mycoplasma californicum	Mycoplasma agalactiae
Leena Putzeys	Oryza sativa	Zea mays
Nina Rank	Phytophthora infestans	Saccharomyces cerevisiae
Diego Sainz Garcia	Yarrowia lipolytica	Aspergillus niger
Adriana Samareanu	Neurospora crassa	Puccinia graminis
Marta Sousa Santos	Candida albicans	Fusarium graminearum
Maria-Alexandra Stanciu	Caenorhabditis elegans	Daphnia pulex
Alexander Ian Taylor	Anopheles gambiae	Drosophila melanogaster
Maria Tsontaki	Nasonia vitripennis	Apis mellifera
Casper van Bavel	Latimeria chalumnae	Danio rerio
Hannelore Vermoesen	Oryzias latipes	Xenopus tropicalis
Stefaan Verwimp	Ornithorhynchus anatinus	Homo sapiens
David Wouters	Loxodonta africana	Ovis aries
Magdalena Zielonka	Trichoplax adhaerens	Cavia porcellus