**Practical 4: Phylogenomics**

Group no. 11

Fuqi Xu, Shuhan Xu, Milda Valiukonytė

**Exercise 1.**

1. Found the multi FASTA files with protein sequences for each genome.

2. Configured BLAST.

3. Created BLAST databases for each of the proteomes.

4. Chose 09.fa.txt.pfa as the reference proteome and performed BLAST search against other proteomes, generating XML output files (blast\_17, blast\_49, blast\_51).

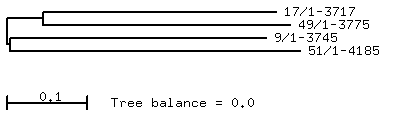
5. Created XML parser (script: blastResultParser\_modified\_best.py) and parsed the four output files (output: best\_orthologs\_17, best\_orthologs\_49, best\_orthologs\_51)

6. Combined the best hits into one cluster file (script: cluster\_maker.py, output: clusters)

7. Wrote a script (10\_clusters.py) which outputs multi FASTA files with sequences for each ortholog cluster (output: cluster\_1, cluster\_2, etc).

**Exercise 2.**

1. Made a multiple sequence alignment for each of the ortholog clusters using KALIGN (1\_aligned, 2\_aligned, etc).
2. Concatenated alignments into alignment of one long metagene (script metagene.py, output:finalfile)



4.