

Assignment phylogenetic trees - Comparative genomics

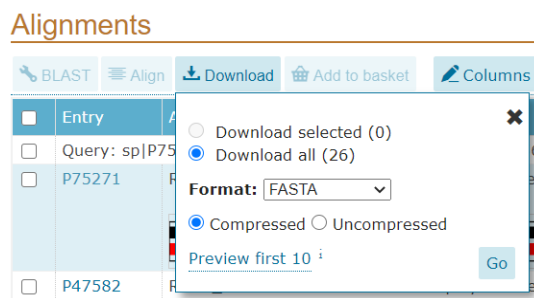
Biancamaria Florenzi, r0777934

Step 1 : search Mpn515 in UniProt (<http://uniprot.org>)

- From what species is this protein? It is from *Mycoplasma pneumoniae* strain ATCC 29342 / M129.
- What is the function of this protein? DNA-dependent RNA polymerase, it catalyzes the transcription of DNA into RNA using the four ribonucleoside triphosphates as substrates.

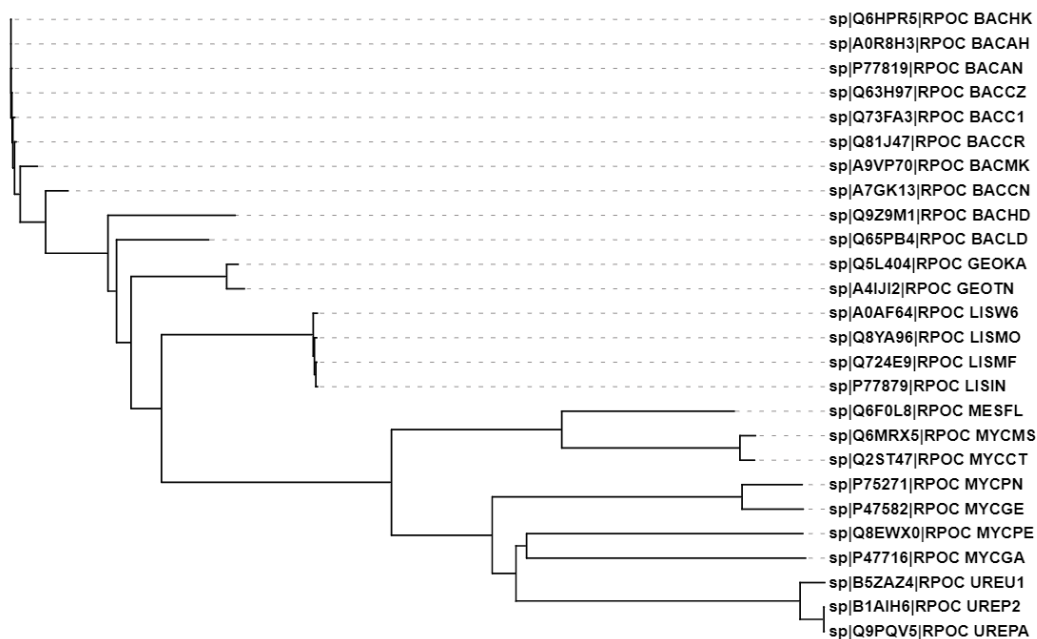
Step 2: find homologs in at least 25 distinct species.

I ran a BLAST query for the protein looking for 250 matches (default) and then downloaded by filtering the reviewed sequences, because they were 26 with the initial sequence and from different organisms. By downloading in format fasta I already obtain a single file that contains all the sequences.

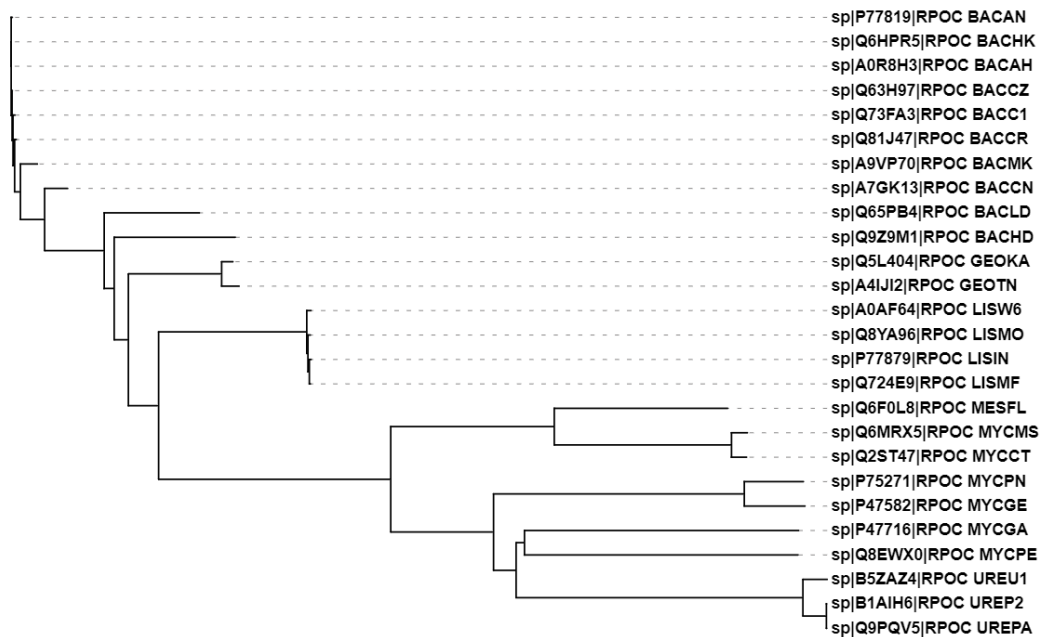


Step 3: make a multiple alignment or (<http://www.ebi.ac.uk/Tools/msa/>), use at least 2 alignment programs. Create a phylogenetic tree for each of your alignments, compare them. Visualize your trees in iTol (<http://itol.embl.de>).

I first use muscle and then clustal omega. I obtain a file containing the alignment and one containing the tree, for each of the programs. I upload the tree file (text) to iTol for visualisation of the trees.



Above the tree from the muscle alignment.



Tree from the clustal omega alignment

I carry out a similar exercise but using the terminal and the hpc. I copy the 2 files with the sequences using the scp; after extracting clustalw and setting the path, I can use the command clustalw2 to run the program. But first I explore the two files, both contain 18 sequences (in 360 lines).

```
✓ [Nov/09 20:50] r0777934@r06i00n03 ~/comparative_genomics $ grep '>' 16S_Myco.fsa.txt | wc -l
18
```

I perform a multiple alignment in clustalw2, with the slow option and resetting the gaps before alignment. I obtain again two output files for each run, one for the alignment and one for the tree

```
✓ [Nov/09 20:56] r0777934@r06i00n03 ~/comparative_genomics $ ls 16S*
16S_alignment 16S_Myco.fsa.txt 16S_tree
```

```
✓ [Nov/09 20:57] r0777934@r06i00n03 ~/comparative_genomics $ head 16S_alignment
CLUSTAL 2.1 multiple sequence alignment
```

```
16S_Mycoplasma_pneumoniae      -----ATTTTCTGAGAGTTTGATCCTGGCTCAGGATTAACGCTG
16S_Mycoplasma_gallisepticum   -----AGAGTTTGATCCTGGCTCAGGATTAACGCTG
16S_Ureaplasma_parvum         -----AATTTAAAGAGTTTGATCCTGGCTCAGGATTAACGCTG
16S_Ureaplasma_urealyticum     -----ATTAACGCTG
16S_Ureaplasma_urealyticum_1   TAGAATCCGTC AATTTTAAAGAGTTTGATCCTGGCTCAGGATTAACGCTG
16S_Ureaplasma_gallorale      -----ATTAACGCTG
16S_Ureaplasma_felinum        -----GAGTTTGATCCTGGCTCAGGATTAACGCTG
```

```
✓ [Nov/09 20:57] r0777934@r06i00n03 ~/comparative_genomics $ head 16S_tree
```

```
(
(
(
(
(
(
(
16S_Mycoplasma_pneumoniae:0.05518,
16S_Mycoplasma_gallisepticum:0.04116)
:0.02535,
```