Today is a little different in that we start with a gene name instead of a protein sequence:

fusA

Q1 Can you find this gene in *E. coli*? What about *B. subtilis*? Save the sequences

Q2 What does the protein product do? What does it have in common with the protein from Day 1?

Q3 Use a couple of tools to predict functional associations with other proteins. What are most of those proteins?

Q4 Take either the *B. subtilis* or *E. coli* protein sequence and Blastp against RefSeq, limiting to eukaryotes. In what organelles do the hits function? Why are proteins from these organelles similar to bacterial proteins?

Q5 Find *Ostreococcus tauri* in the results list and use a tool to find its protein's targeting peptide. Research the organism a little to decide if it should be classified as 'plant' or 'non-plant'. Can the subcellular location be predicted? If so, where is the cleavage site? What is the thylakoid lumen anyway?

Q6 Find a homologue of this protein in humans. Use the name to search OMIM. Are there any mutations of the encoding gene associated with diseases?

Q7 Make a multiple sequence alignment of the two bacterial proteins along with the top BlastP 100 hits. For our purposes today you can delete any sequences that have so many insertions that they make the alignment hard to read (*Vollenhovia emeryi* seems to be a trouble maker in this respect). Are the transit peptides apparent in the sequence alignment? Do you think there are paralogues in the dataset? How would you address that question more carefully if you were to plan a full bioinformatic analysis of this protein family in the future?

Continue analysing your pet protein using the tools you've come across in the course