

Identifying candidate pH-resistant methanogenic Archaea

Methanogenesis is an anaerobic microbial metabolism that results in the production of methane, a potent greenhouse gas. The microbes responsible for this process are methanogenic Archaea (methanogens), which vary across different ecosystems and are significantly influenced by the pH of the environment. A graduate student in Stuart's lab is attempting to culture pH-resistant methanogenic Archaea from lake sediments to use in her experiments. From one enrichment, she has produced 50 different isolated colonies and wants to identify which ones are the best candidates for her growth experiments.

Methanogens can be identified by the presence of the methyl-coenzyme M reductase (McrA) gene, which catalyzes the last step of methanogenesis and is highly conserved across all methanogens. pH resistance can be identified by the number of copies of the HSP70 gene, which is involved in protein biogenesis and refolding for stress resistance (Macario et al. 1999). You have been provided the proteomes (the translated genomes) of all 50 isolates in the proteomes directory, as well as reference protein sequences for the McrA gene and the HSP70 gene.

Use Unix code and bioinformatics tools to determine which isolates Stuart's graduate student should move forward with. Provide all of the code in a single bash script. **Your script should include all of the necessary steps to search each genome for the genes of interest and should produce a summary table collating the results of all searches.** Your script should be efficient, flexible to inputs and outputs, and well commented. Additionally, provide a text file with the names (i.e. proteome_01) of the candidate pH-resistant methanogens based on your results.