Kellen Tyrrell

Dr Brandon Briggs

BioA455

Identifying unknown sequences with phylogenetic analysis

Methods

To identify the genus of the unknown sequences, they were compared to known 16S rRNA sequences. BLAST (Altschul, 1990) was used to gather 21 other known sequences from the NCBI website. To create a multi sequence alignment using the known sequences, unknown sequences, and the 16S rRNA sequence of *Mathanocaldococcus jannaschii*, the program Muscle (Edgar, 2004) was used with a maximum of 100 iterations. Then, a phylogenetic tree based on the generalized time-reversible evolutionary model was created using FastTree (Price, 2009) and the aligned sequences. The tree was visualized using Dendroscope (Huson, 2007) and was rooted using the *Methanocaldococcus jannaschii* 16S rRNA sequence.

Results

The genera of the unknown sequences were discovered to be, respectively, *Hydrogenobacter*, *Hydrogenobacter*, *Hydrogenobaculum*, *Sulfurhydrogenibium* and *Vibrio*.
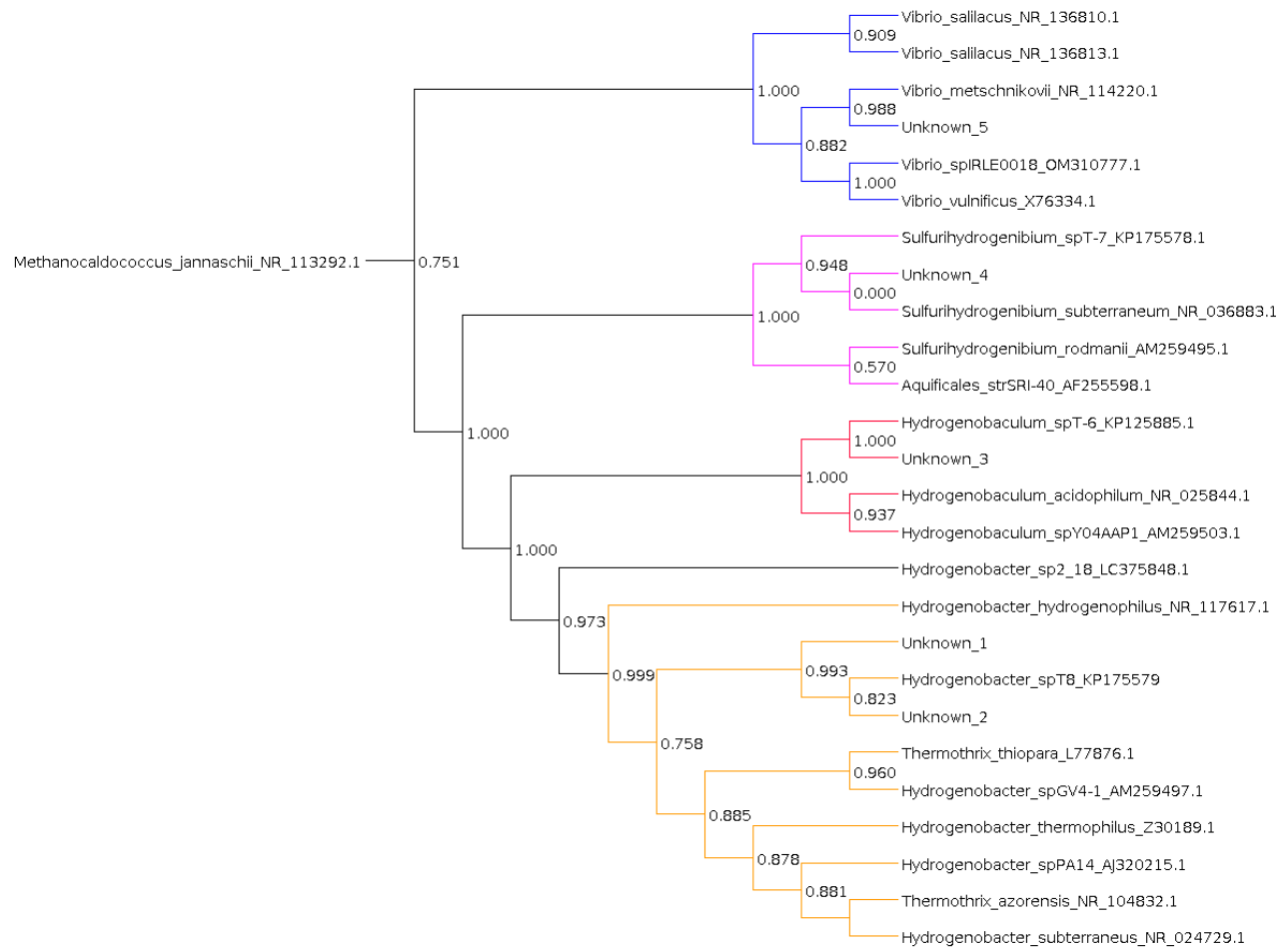
Figure 1. Dendroscope visualization of phylogenetic tree created from fasttree program. The genus Vibrio, Sulfurihydrogenibium, Hydogenobaculum and Hydrogenobacter are colored in blue, magenta, red and yellow respectively.

References

Altschul, S., Gish W., Miller W., Myers E., Lipman., D. (1990). Basic local alignment search tool. *Journal of Molecular Biology.* 215(3), 403-410

Edgar, R.C. MUSCLE: a multiple sequence alignment method with reduced time and space complexity. *BMC Bioinformatics* 5, 113 (2004).

Huson D., Richter D., Rausch C., Dezulian T., Franz M., and Rupp R. (2007). Dendroscope: An interactive viewer for large phylogenetic trees. *BMC Bioinformatics.* 8, 460 (2007).

Price M., Dehal P., Arkin A. (2009). FastTree: Computing Large Minimum Evolution Trees with Profiles instead of a Distance Matrix. Molecular *Biology and Evolution*. 26(7), 1641-1650.