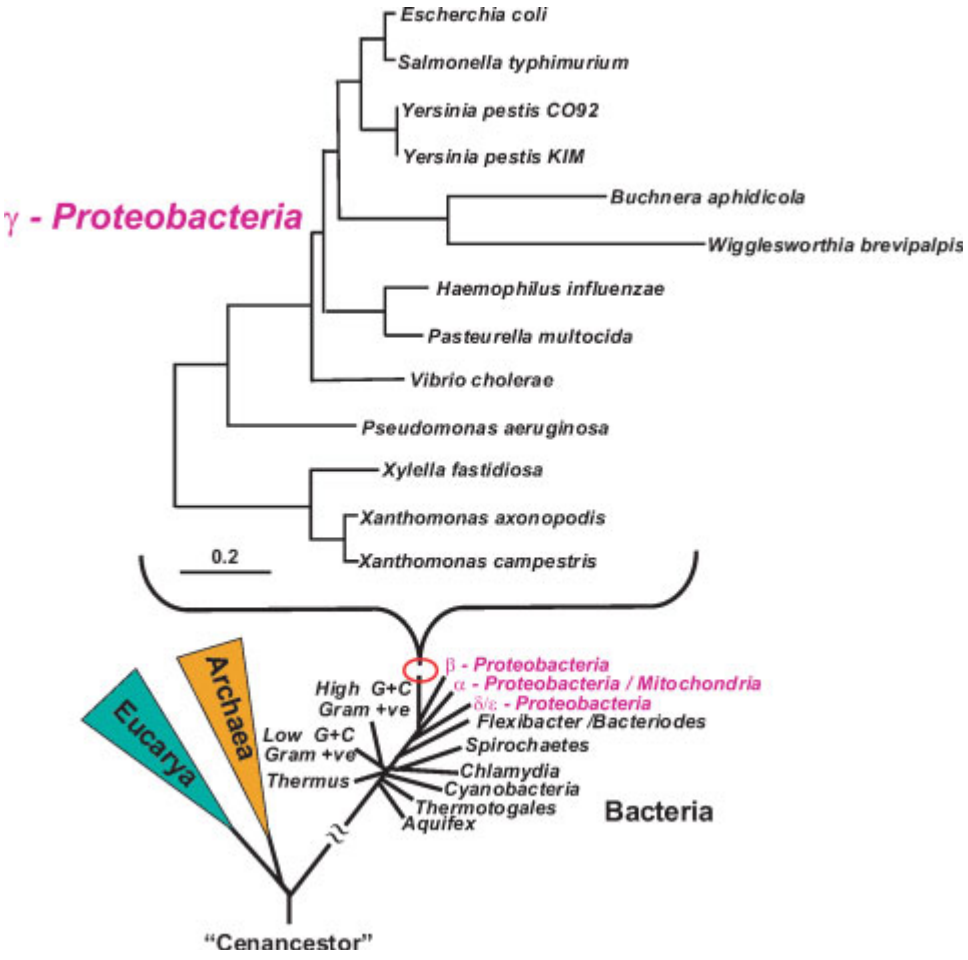
Bio-2304 Phylogeny project Fall-2007

Introduction

Prokaryotes are the most abundant life-form on our planet. Even though it is clear that the diversity of Bacteria and Archaea is enormous the evolutionary relationships among these microorganisms remains unclear. Before DNA sequences were available microbiologists tried to organize bacteria into a taxonomic system based on their shape, size and other morphological features, but they soon became frustrated over the lack of characters which could distinguish them. However, DNA sequencing turned the tide and microbial systematics was revolutionized. The 16S ribosomal (r) RNA (16S rRNA) gene has been a very important molecular marker to dissect evolutionary relationships. Also single or multiple protein sequences and protein-coding genes, and even whole genomes are being used for the same purpose.

Problem

You have been given the task of resolving the evolutionary relationships of a subset of gamma-proteobacteria by using the classical phylogenetic marker, that is the 16S rRNA gene. Below is a phylogeny of gamma-proteobacteria based on 205 protein sequences, as published by Brown and Volker 2004 (BioEssays 26:463-468).



You will use NCBI (http://www.ncbi.nlm.nih.gov/) or a similar sequence databank to identify and acquire the 16S rRNA sequences from the same bacterial species as in the figure shown above. You may use the 16S rRNA sequence from a closely related species if one or more species are missing from the databanks.

Next, the 16S rRNA sequences need to be imported into BioEdit. Here, the sequences can be aligned automatically (ClustalW is implemented in BioEdit) and/or manually.

Once, you have your final dataset you will use different phylogenetic software packages to reconstruct the evolutionary relationships between the bacteria.

Methods

- 16S rRNA genes can be found at the NCBI web site, or similar.

- Import sequences into BioEdit and align them automatically. Make sure to inspect your alignment manually for any obvious mistakes. Adjust alignment if necessary.

- Next challenge is to export your alignment from BioEdit. (1) If you want to use Phylip (PhyML) them you can simply save your dataset in Phylip format. (2) If you want to use MEGA, then you first have to export your dataset in msf (.msf) format, and then use then MEGA software to convert your msf formatted dataset to MEGA format (.meg). (3) PAUP and MrBayes require nexus formatted files. I recommend that you use MEGA or PhyML first. Then you may try Phylip.

- Do at least the following analyses:

    - Generate a Neighbor Joining (NJ) tree using MEGA (choose evolutionary model)
    - Run NJ Bootstrap analysis, with 2000 replicates.
    - Generate a Maximum Likelihood (ML) tree using PhyML (choose evolutionary model)
    - Run ML Bootstrap analysis, with 500 replicates.

Use the “Phylogeny” lecture notes and “Phylogeny” practicals for resources and protocols.

Consider these thought and questions:

(1) Does your 16S rRNA trees exactly match that of Brown and Volker 2004?
(2) Discuss similarities and differences in NJ and ML trees. E.g., what do you think is the most correct.
(3) Are your phylogenies supported by bootstrap analyses.
(4) Explain your choice of evolutionary model.

(5) Use the SSU rRNA secondary structure model below to see if you can identify highly conserved and less conserved regions. Try to explain the result.

