Abstract

The increasing throughput of sequencing raises growing needs for methods of sequence analysis. Biological macro-molecules such as rDNA, rRNA and proteins play an essential role in all living organism. Structurally, they are all chains of residues belonging to a small set of basic molecules and the functional characteristics of each macro-molecule are determined by the order and composition of its component. It is therefore not surprising that comparison and alignment of biological sequence is one of the most important contributions from computational biology to modern bio-sciences.

Bacterial identification becomes a challenge particularly in case they are either involved in an industrial process with heavy investments at risk or are a serious threat to human beings. Sequencing of the rrs (16S rDNA) of bacteria is vigorously pursued for correct identification and classification. It has led to a large database of 1,237,963 entries http://rdp.cme.msu.edu/. The key questions which we are addressing:

1. Whether there are certain latent and as yet un-explored features in the nucleotide sequences of the rDNA, which can be used to define the phylogenetic limits of a particular genus or species?