Overview

Sequencing of highly conserved genes (also known as “targeted loci”), such as 16S ribosomal RNA and cytochrome C oxidase subunit I, found in samples collected from environmental and clinical specimens is crucial for identification of organisms and pathologic agents present in these samples, and important in phylogeny, population genetics, as well as microbial ecology studies. The BLAST web services [1] from NCBI offer a convenient tool to help you make sense of your 16S RNA sequence data. These include nucleotide BLAST (blastn) against the curated 16S Microbial ribosomal RNA database, or more broadly against the bacterial and archaeal subset in the NT database, and MOLE-BLAST [2], which performs a more comprehensive analysis and generates multiple alignment based phylogenetic tree(s) to provide a visual presentation of the relationships among the input query sequences in the context of matches database entries. MOLE-BLAST often provides good taxonomic placement for sequences that do not have perfect matches in the database.

Setting Up the Search

From the NCBI BLAST homepage, click “nucleotide blast” link (A) under the “Basic BLAST” section to open the search page. Paste in the query sequences (B, provided as accessions), change the database from default “Nucleotide collection (nr/nt)” to “16S ribosomal RNA sequences (Bacteria and Archaea)” (C), and click BLAST button (D) to initiate the search.

Such a BLAST search compares the input 16S RNA sequences from uncultured bacteria, already deposited in the Nucleotide database, against those 16S entries from the Targeted Loci projects curated by NCBI staff. The top identical or near identical match covering the whole length of the input query helps you identify the source organism of the input query sequences.
16S Ribosomal RNA BLAST Results

From the Description table, you can see that the second query matches to the 16S ribosomal RNA gene from *Synechococcus rubescens* strain with 100% length coverage and 99.70% identity (A, note the increased precision, note the increased precision). With this result, you can identify the query sequence to be from that organism. However, the result for the sixth query is less certain. Even though the query coverage is nearly complete at 99%, the identity is only 91.08% (B). Because of this lack of identical/near identical match, the source organism for this sequence cannot be ascertained. For this sequence, you can re-run the search using “Edit and Resubmit” link (C) to search against “(nr/nt)” database with uncultured/environmental sample sequences excluded, where a better match to an *Acinobacterium acl-B2* entry can be found.

The “16S Ribosomal RNA Sequences” Database

Sequences in this database are curated 16S ribosomal RNA sequences from two Targeted Loci projects, PRJNA33175 for bacteria (D) and PRJNA33317 for archaea (E). You can get the complete list of curation efforts from the umbrella project PRJNA224725 (https://www.ncbi.nlm.nih.gov/bioproject/224725).

**Related Factsheets**

1. Guide to BLAST Web Site
2. MOLE-BLAST