



RefSeq: 16S rRNA Targeted Loci Project

Marker sequences for phylogenetic analyses

<http://www.ncbi.nlm.nih.gov/genomes/static/refseqtarget.html>

National Center for Biotechnology Information • National Library of Medicine • National Institutes of Health • Department of Health and Human Services

Scope of the RefSeq Targeted Loci project

The RefSeq Targeted Loci (TL) project collects and curates representative sequences from phylogenetically informative loci for use in phylogenetic analyses. Currently the set includes 16S ribosomal RNAs from type strains for prokaryotes, and 18S and 28S ribosomal RNAs from fungi. Future inclusion of other informative marker sequences is also being considered. The input to the TL project consists of third party curated datasets that use sequences in the International Database Sequence Collaboration (INSDC, i.e. GenBank/DDBJ/ENA) as starting materials. The 16S TL consists of prokaryotic type strain sequences identified in sequence databases and literature. Curation efforts of each contributing database may include updates to taxonomy, strain and/or isolate information, sequence alterations, or feature annotations.



Data access

Existing Targeted Loci projects can be accessed through the BioProject homepage (www.ncbi.nlm.nih.gov/bioproject/) by searching with “targeted loci” as the query term. The rRNA sequences from each TL can be retrieved from the nucleotide database (www.ncbi.nlm.nih.gov/nuccore/) using a fielded query. For example, the bacterial 16S rRNA TL sequences can be retrieved using [33175\[BioProject\]](#). These sequences are available as a component of the regular RefSeq release (<ftp.ncbi.nlm.nih.gov/refseq/>) and directly from the TL FTP site (<ftp.ncbi.nlm.nih.gov/genomes/TARGET/>).

Summary from BioProject

Descriptions of novel eubacteria, especially type strains linked to particular species, often include deposition of complete or near-complete 16S rRNA sequences in the public databases. Thus, these 16S sequences from the bacterial TL project can serve as proxies for phylogenetic placement of novel sequences. The dataset containing these sequences is described in the BioProject database record [33175](#). The record, shown below, provides project name, title, description, accession and project ID (A). The total number of sequences currently available is shown in the “Project Data” section (B) and linked to their corresponding records in the nucleotide database. Additional links to the Targeted Loci Project homepage (C) and related resources, such as the German culture collection repository (DSMZ, D), the TL FTP site (E), third party curated resources (F), and the curated strain information database (StrainInfo, G) are also provided.

BioProject [BioProject] Search Limits Advanced Help

Display Settings: [v] **A** Send to: [v]

Name: Bacteria Accession: PRJNA33175 ID: 33175

Title: Bacterial 16S Ribosomal RNA RefSeq Targeted Loci Project

The 16S ribosomal RNA targeted loci project is the result of an international collaboration between a number of ribosomal RNA databases and NCBI to provide a curated and comprehensive set of complete and near full length Reference Sequence records for phylogenetic and evolutionary analyses. Sequences that represent the consensus of all contributing databases in both sequence content and taxonomic assignment are promoted to RefSeqs. All sequences will have the same project ID and can be found as such.

[NCBI RefSeq Targeted Loci Project](#) Less...

Project Data Type: Targeted Locus (Loci)

Attributes: Scope: Multiisolate; Material: Genome; Capture: TargetedLocusLoci; Method type: Sequencing;

Project Data:

| Resource Name | Number of Links |
|---------------|-----------------|
| SEQUENCE DATA | |
| Nucleotide | 7270 B |

Submission:

Registration date: 10-Dec-2008
NCBI

Related information

- Genome
- SRA
- Taxonomy

NCBI Links

- Targeted Loci Project **C**

Related Resources

- DSMZ **D**
- FTP **E**
- Silva
- RDP
- GreenGenes
- StrainInfo **F**
- CRW

G

Recent activity

The 16S rRNA comparison tool

The 16S comparison tool (www.ncbi.nlm.nih.gov/genomes/ssucomp.cgi) allows comparison of the various curation efforts overall, as well as for each particular 16S sequence. Entering an accession (**A**) followed by clicking the “Refresh” button filters the display. The summary table (**B**) provides overall statistics of each contributing database including total number of sequences, multi-copy and unique sequences, and those with vector screen hits (indicating possible vector contamination).

The 16S rRNA comparison tool displays sequence and annotation differences between databases that curate publicly available 16S ribosomal RNA sequences. More information: [RefSeq Targeted Loci Project](#), [16S rRNA Database Annotation Comparison Help Document](#) **J**

Selected accession: **A**

SSU16S rRNA databases: Refresh **B**

| Category | <input checked="" type="checkbox"/> INSD | <input checked="" type="checkbox"/> Silva | <input checked="" type="checkbox"/> CRW | <input checked="" type="checkbox"/> RDP | <input checked="" type="checkbox"/> Green_Genes | <input checked="" type="checkbox"/> Curated_GenBank | <input type="checkbox"/> RefSeq_NR |
|-------------------|--|---|---|---|---|---|------------------------------------|
| Total | 86633 | 7710 | 36783 | 5359 | 64342 | 5788 | 2950 |
| Multicopy | 2677 | 0 | 712 | 0 | 2576 | 0 | 0 |
| Unique | 116 | 4 | 257 | 0 | 473 | 0 | - |
| No INSD Accession | 0 | 3 | 198 | 0 | 0 | 0 | 0 |
| VecScreen Hits | 636 | 0 | 435 | 15 | 149 | 0 | 0 |

The “Results of Comparisons” table (**C**) shown below contains a detailed comparison across different databases. The content of this result table changes upon filtering by accession or deselecting a source database. The numbers in blue (**D**) link to a separate display (**E**) that provides more granular details on the comparison between the two target databases, such as organism, strain, sequence length and links to GenBank display (**F**) and alignment view (**G**). The difference in the RefSeq_NR row (**H**) indicates corrections made by the NCBI staff during the curation process. The comparison of the FEATURE sections (**I**) shows the removal of a primer sequence. More information on this tool is given in the help document (**J**) linked in the summary section above.

Results of comparisons: **C**

| Sources | Same Accession | Identical | Different | Diff. Version | Diff. Organism | Diff. Organism Parent/Child | Diff. Strain | Diff. Strain Absent | Diff. Length | Diff. N Count | Diff. Ambig Count | Diff. Sequence | Absent Accession |
|--------------|----------------|-----------|-----------|---------------|----------------|-----------------------------|--------------|---------------------|--------------|---------------|-------------------|----------------|------------------|
| All selected | 953 | 466 | 487 | 25 | 124 | 80 | 38 | 17 | 385 | 168 | 14 | 396 | 5644 |
| Silva | 7695 | 6679 | 1016 | 0 | 807 | 568 | 166 | 80 | 61 | 12 | 11 | 90 | 78950 |

Records with different sequences: **E**

| accession | gi | accession | organism | strain | start | stop | length | N count | Ambig count | align | VecScreen |
|--|-----------------|-----------|--|----------------------------------|----------|------|--------|---------|-------------|-------|-----------|
| X71851 F View in GenBank View in main page | INSD | 431247 | X71851 | Clostridium lentocellum DSM 5427 | DSM 5427 | 1 | 1508 | 1508 | 11 | 0 | - |
| | Silva | - | X71851 | Clostridium lentocellum DSM 5427 | DSM 5427 | 1 | 1508 | 1508 | 11 | 0 | - |
| | CRW | 431247 | X71851 | Clostridium lentocellum DSM 5427 | DSM 5427 | - | - | 1508 | 11 | 0 | - |
| | RDP | - | X71851 | Clostridium lentocellum DSM 5427 | DSM 5427 | - | - | 1508 | 11 | 0 | - |
| | Green_Genes | 431247 | X71851 | Clostridium lentocellum DSM 5427 | DSM 5427 | - | - | 1508 | 11 | 0 | - |
| | Curated_GenBank | 431247 | X71851 | Clostridium lentocellum DSM 5427 | DSM 5427 | - | - | 1508 | 11 | 0 | - |
| RefSeq_NR | 219846509 | NR_026101 | Clostridium lentocellum RHM5; DSM 5427 | DSM 5427 | 1 | 1485 | 1485 | 11 | 0 | - | H |

FEATURES Location/Qualifiers

X71851

```

source
  1..1508
  /organism="Clostridium lentocellum DSM 5427"
  /mol_type="genomic DNA"
  /strain="DSM 5427"
  /db_xref="taxon:642492"
gene
  1..1508
  /gene="16S rRNA"
rRNA
  1..1508
  /gene="16S rRNA"
  /product="16S ribosomal RNA"
primer_bind
  4..23
  /note="The region between 4-23 is the 5' PCR primer site and has not been sequenced"
  
```

NR_026101

```

source
  1..1485
  /organism="Clostridium lentocellum DSM 5427"
  /mol_type="rRNA"
  /strain="RHM5; DSM 5427"
  /culture_collection="DSM:5427"
  /db_xref="taxon:642492"
  /note="type strain of Clostridium lentocellum"
rRNA
  1..1485
  /product="16S ribosomal RNA"
  
```

I

Specialized BLAST interface for the 16S rRNA Targeted Loci

A specialized BLAST interface to the 16S rRNA Targeted Loci (www.ncbi.nlm.nih.gov/genomes/16S/search16S.html) is available. Here, a sequence can be searched against these existing 16S rRNA sequences and be placed within a pre-computed phylogenetic tree (constructed using maximum likelihood method) for possible identification purposes. The result page (shown using JF501301 as an example input query) lists the best ten hits in a table (A) with indication of the tentative organism placement of the input sequence on the phylogenetic tree (B). Actual placement on the phylogenetic tree is shown below the table (C), where branches containing the top ten matches are expanded with their ranking indicated by a grid to the right (D). Since this display is shown without a percent identity cutoff, caution should be used to prevent over-interpretation of the results.

Home Query: >Uncultured bacterium clone EAS0806_H5 16S ribosomal RNA

16S Ribosomal RNA Reference Sequence A ▶ Top 10 hits to Bacterial set/tree B

Similarity Search

Beta release

This tool visualizes BLAST results of your search by mapping them on a phylogenetic tree

| rank | identity | coverage | organism |
|-----------------------------------|----------|----------|-------------------------------|
| 1 299% in <i>Pantoea ananatis</i> | | | |
| 1 | 99.4 | 100 | <i>Pantoea ananatis</i> |
| 8 <99% ≥95% in Enterobacteriaceae | | | |
| 2 | 96.6 | 100 | <i>Enterobacter cowanii</i> |
| 3 | 95.8 | 100 | <i>Enterobacter asburiae</i> |
| 4 | 95.8 | 99.4 | <i>Serratia rubidaea</i> |
| 5 | 95.3 | 98 | <i>Cedecea davisae</i> |
| 6 | 95.2 | 100 | <i>Erwinia persicina</i> |
| 7 | 95.2 | 100 | <i>Citrobacter werkmanii</i> |
| 8 | 95.0 | 100 | <i>Enterobacter amnigenus</i> |

Click the root node of a subtree to select and change resolution (Max. 500 leaves)
Leaves: 71. resolved: 9

Select subtree Table Search Reset Unselect Align

1 [min] [max] 71

Viewing options, search & markup

Organism: e.g. Escherichia coli or Gammaproteobacteria

Strain:

Culture collection: e.g. ACAM:644

Min. length:

Max. length: e.g. 1600

Accession: e.g. NR_027532 or AF073456

Search & tag Clean form Remove tags

429 seq. from Proteobacteria(1112)

- 33 seq. from Enterobacteriaceae(76)
- 12 seq. from Enterobacteriaceae(76)
- 14 seq. from Enterobacteriaceae(76)
- 17 seq. from Pasteurellales(17)
- 43 seq. from Vibrionaceae(43)

429 seq. from Proteobacteria(1112)

- 12 seq. from Enterobacteriaceae(76)
- 33 seq. from Enterobacteriaceae(76)
- 5 seq. from Buttiauxella(5)
- 14 seq. from Enterobacteriaceae(76)
- 3 seq. from Enterobacteriaceae(76)
- Serratia entomophila NR_025338

| score | %id | %coverage | gi | accession | genbank | length | e-value | define |
|-------------------------|------|-----------|---------------------------|---------------------------|----------|--------|---------|---|
| 786.141 | 99.4 | 100.0 | 219846434 | NR_026045 | U80196 | 501 | 0 | <i>Pantoea ananatis</i> strain 1846 16S ribosomal RNA, partial sequence |
| 732.253 | 96.6 | 100.0 | 219878427 | NR_025566 | AJ508303 | 501 | 0 | <i>Enterobacter cowanii</i> strain 888-76 16S ribosomal RNA, partial sequence |
| 729.083 | 95.8 | 100.0 | 219856821 | NR_024640 | AB004744 | 501 | 0 | <i>Enterobacter asburiae</i> strain JCM6051 16S ribosomal RNA, partial sequence |
| 724.328 | 95.8 | 99.4 | 219856825 | NR_024644 | AB004751 | 498 | 0 | <i>Serratia rubidaea</i> strain JCM1240 16S ribosomal RNA, partial sequence |
| 719.573 | 95.2 | 100.0 | 219857234 | NR_024862 | AF025373 | 501 | 0 | <i>Citrobacter werkmanii</i> strain CDC 0876-58 16S ribosomal RNA, partial sequence |
| 719.573 | 94.8 | 100.0 | 219846438 | NR_026049 | U80205 | 501 | 0 | <i>Erwinia persicina</i> strain HK 204 16S ribosomal RNA, partial sequence |
| 716.403 | 95.0 | 100.0 | 219857252 | NR_024883 | AF047186 | 501 | 0 | <i>Kluyvera georgiana</i> strain ATCC 51603 16S ribosomal RNA, partial sequence |
| 716.403 | 95.0 | 100.0 | 219856823 | NR_024642 | AB004749 | 501 | 0 | <i>Enterobacter amnigenus</i> strain JCM1237 16S ribosomal RNA, partial sequence |
| 713.233 | 94.8 | 100.0 | 219857233 | NR_024861 | AF025371 | 501 | 0 | <i>Citrobacter farmeri</i> strain CDC 2991-81 16S ribosomal RNA, partial sequence |
| 713.233 | 94.8 | 100.0 | 219846405 | NR_025635 | AB004750 | 501 | 0 | <i>Enterobacter aerogenes</i> strain JCM1235 16S ribosomal RNA, partial sequence |
| 710.063 | 95.3 | 98.0 | 219857654 | NR_025243 | AF493976 | 491 | 0 | <i>Cedecea davisae</i> strain DSM 4568 16S ribosomal RNA, partial sequence |
| 710.063 | 94.6 | 100.0 | 219857408 | NR_024996 | AF129443 | 501 | 0 | <i>Raoultella planticola</i> strain ATCC 33531 16S ribosomal RNA, partial sequence |
| 710.063 | 94.6 | 100.0 | 219856822 | NR_024641 | AB004748 | 501 | 0 | <i>Enterobacter gergoviae</i> strain JCM1234 16S ribosomal RNA, partial sequence |
| 706.893 | 94.4 | 100.0 | 228716249 | NR_027549 | AF530475 | 501 | 0 | <i>Escherichia fergusonii</i> ATCC 35469 16S ribosomal RNA, partial sequence |
| 706.893 | 94.4 | 100.0 | 219878203 | NR_025342 | AJ233437 | 501 | 0 | <i>Tatumella ptyseos</i> strain DSM 5000 16S ribosomal RNA, partial sequence |
| 706.893 | 94.4 | 100.0 | 219878192 | NR_025331 | AJ233404 | 501 | 0 | <i>Buttiauxella izardii</i> strain S3/2-161 16S ribosomal RNA, partial sequence |

The result display can be decorated using the “Search” dialog box (E). The example highlights Enterobacter-containing nodes (F). Clicking the “Table” button (G) changes the tree view display to a table view.

16S rRNA sequence database for BLASTn

The 16S rRNA database from TL can also be selected when running standard BLASTn (blast.ncbi.nlm.nih.gov), through the “nucleotide blast” (A) link under the “Basic BLAST” section, by selecting “16S microbial” as the database (B). An example search result page (shown below) lists the 16S rRNA TL matches sorted by the significance of their alignment to the input query (C). Sorting order can be changed by clicking different column headers (D). The Taxonomy reports (E) and Distance tree of results (F) links provide alternative displays, highlighting the hit distribution among different taxa and the distances among the query and matched 16S rRNA TL entries, respectively.

The image illustrates the BLASTn workflow for searching the 16S rRNA database. It is divided into several sections:

- BLAST Home:** Shows the "Basic BLAST" section with a "nucleotide blast" link (A).
- Choose Search Set:** Shows the "Database" dropdown set to "16S microbial" (B).
- Search Results:** Shows the "NCBI/BLAST/blastn suite/Formatting Results" page for query "gb|JF501301| (501 letters)". It includes a table of sequences producing significant alignments (C) and a "Taxonomy Report" (E).
- Table of Sequences:**

| Description | Max score | Total score | Query cover | E value | Ident | Accession |
|--|-----------|-------------|-------------|---------|-------|-----------------------------|
| Pantoea vaqans C9-1 strain C9-1 16S ribosomal RNA, complete sequence | 920 | 920 | 100% | 0.0 | 99% | NR_102966.1 |
| Pantoea ananatis AJ13355 strain AJ13355 16S ribosomal RNA, complete sequence | 909 | 909 | 100% | 0.0 | 99% | NR_074740.1 |
- Distance Tree:** Shows a "Tree method" of "Fast Minimum Evolution" with a "Max Seq Difference" of 0.75. It displays a phylogenetic tree with collapsed branches (green triangles) for clarity. A text box notes: "Certain branches (green triangles) are collapsed for clarity."