

# Running blast from command line

June 25, 2015

# BLAST ?

[Home](#) [Recent Results](#) [Saved Strategies](#) [Help](#) [\[Sign In\]](#) [\[Register\]](#)

NCBI/ BLAST/ blastn suite

Standard Nucleotide BLAST

blastn blastp blastx tblastn tblastx

Enter Query Sequence

BLASTN programs search nucleotide databases using a nucleotide query. [more...](#) [Reset page](#) [Bookmark](#)

Enter accession number(s), gi(s), or FASTA sequence(s) [?](#) [Clear](#)

Query subrange [?](#)

From

To

Or, upload file

Choose File

no file selected [?](#)

Job Title

Enter a descriptive title for your BLAST search [?](#)

☐ Align two or more sequences [?](#)

Choose Search Set

Database

☐ Human genomic + transcript ☐ Mouse genomic + transcript ☒ Others (nr etc.):  

Nucleotide collection (nr/nt) [?](#) [?](#)

Organism

Optional

Enter organism name or id—completions will be suggested

☐ Exclude [+](#)

Enter organism common name, binomial, or tax id. Only 20 top taxa will be shown [?](#)

Exclude

Optional

☐ Models (XM/XP) ☐ Uncultured/environmental sample sequences

Limit to

Optional

☐ Sequences from type material

Entrez Query

Optional

[YouTube](#) [Create custom database](#)

Enter an Entrez query to limit search [?](#)

Program Selection

Optimize for

☒ Highly similar sequences (megablast)  
☐ More dissimilar sequences (discontiguous megablast)  
☐ Somewhat similar sequences (blastn)  
Choose a BLAST algorithm [?](#)

# Why I want to run BLAST from command line?

- I can choose any database
- Automation :
  - Use output file in downstream program eg. R
  - BLAST can be a part of your code
- Speed

# Web interface

BLASTN programs search nucleotide databases using a nucleotide query. [more...](#)

## Enter Query Sequence

Enter accession number(s), gi(s), or FASTA sequence(s) [?](#) [Clear](#)

Query subrange [?](#)

From

To

Or, upload file  no file selected [?](#)

Job Title

Enter a descriptive title for your BLAST search [?](#)

☐ Align two or more sequences [?](#)

## Choose Search Set

Database


☐ Human genomic + transcript ☐ Mouse genomic + transcript ☒ Others (nr etc.):

Nucleotide collection (nr/nt)  [?](#)

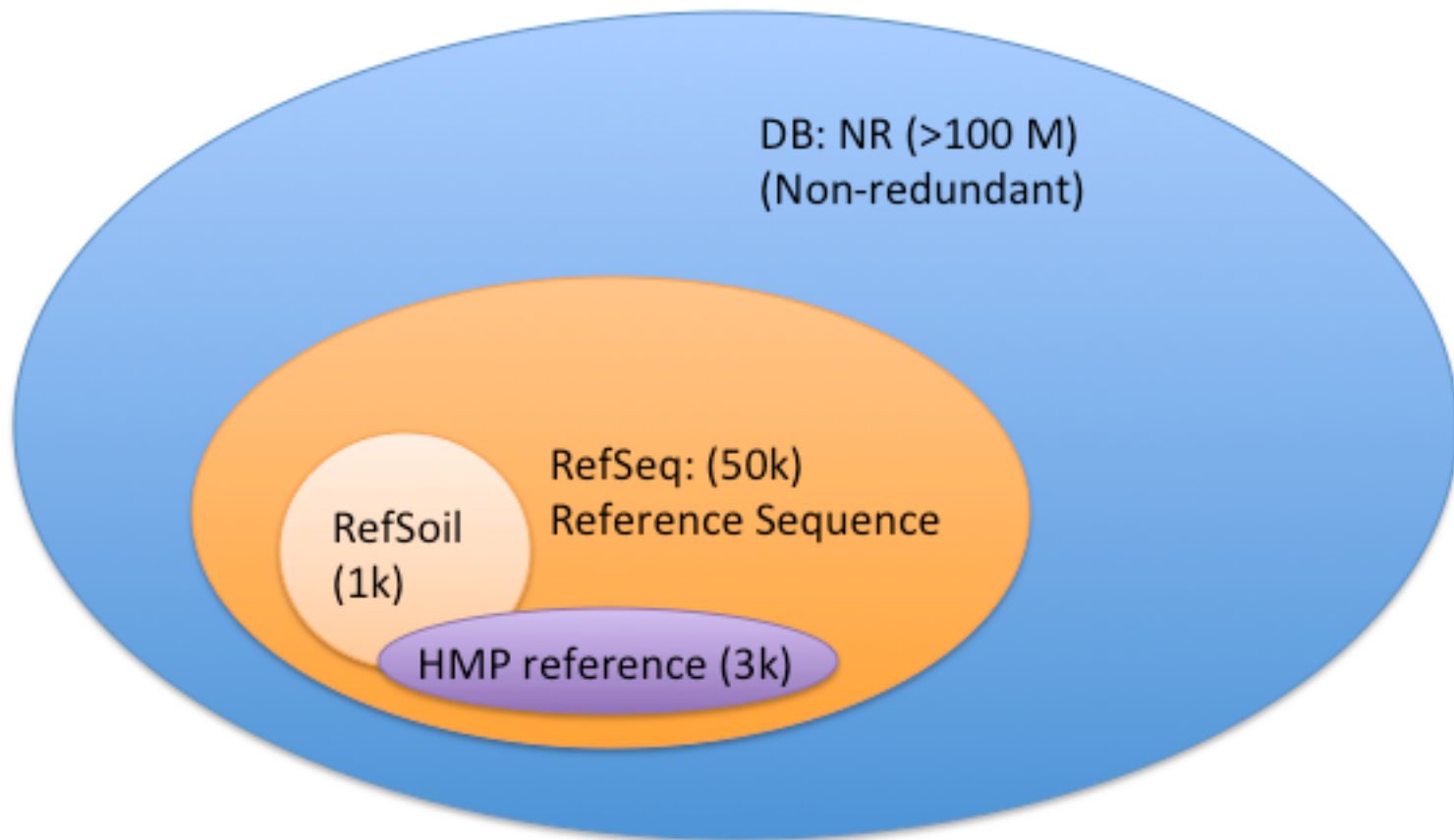
Organism [Optional](#)

Enter organism name or id--completions will be suggested ☐ Exclude

Enter organism common name, binomial, or tax id. Only 20 top taxa will be shown [?](#)



# Compare Database (DB)



# HMP: A good example of curated DB

**HMP**  
NIH HUMAN  
MICROBIOME  
PROJECT

**Current News**

- January 2015  
Metagenome Analysis Workshop  
March 3-6
- September 2014  
IHMC 2015 from Mar. 31 to Apr. 2
- May 2014  
Poster and Booth at ASM 2014

[More News Items](#)

**Publications**

- Fungal Signature in the Gut Microbiota of Pediatric Patients With Infl...
- Identification of a gene in Mycoplasma hominis associated with preterm...
- Diet in the pathogenesis and treatment of inflammatory bowel diseases....

[More Publications](#)

**Partner Resources**

- NIH Common Fund
- NCBI HMP Data Repository

**OVERVIEW** REFERENCE GENOMES MICROBIOME ANALYSIS HEALTH & ETHICS RESOURCES OUTREACH DATA BROWSER

Welcome to the Data Analysis and Coordination Center (DACC) for the National Institutes of Health (NIH) Common Fund supported Human Microbiome Project (HMP). This site is the central repository for all HMP data. The aim of the HMP is to characterize the human microbiome and its role in health and disease. The HMP is a collaborative effort between the NIH and the Common Fund. The HMP is a central repository for all HMP data. The HMP is a central repository for all HMP data.

**GET DATA**

home > blast

To begin, select a Reference Genome BLAST Database from the menu below, or select an individual organism.

**Skin** Select an individual organism

- ☒ BLAST - nucleotide sequence against nucleotide sequence of predicted genes in this genome
- ☐ BLASTP - protein sequence against amino acid sequence of predicted genes in the genome
- ☐ TBLASTN - protein sequence against the entire genome sequence
- ☐ BLASTX - translated nucleotide query against amino acid sequence of predicted genes in the genome
- ☐ TBLASTX - translated nucleotide query against the entire genome sequence

Paste nucleotide or protein sequence below:

Upload File: Choose File no file selected (Max Size: 20M)

Job Title:

**Submit**

\* Run against NCBI databases: **NCBI Blast**

**Areas of Interest**

**Ethical Implications**  
The ethical, legal and social implications of the HMP.

**DACC Member Organizations**

**Related Sites**

# Use meta-seq data as a DB

DataBase

Your Isolates or gene of interest

> My Baby

```
GGAGAGTTTGTATCTGGCTCAGGACGAACGCTGGCGGCGTGCTTAACACATGCAAGTCGAACGGTCCAGT
GTCTAACACCGAGTGTTAAGGAGCAGATAAGTGGATGACAGCGAGAGCGCGAAGAATGAGCGCTCCACG
CATTTCAATAAGTGTCAACACAGCTTCTGGAATGCTGGGTGTTAGATGCTGGATAGTGGCGGACGGGTGAG
TAACGCGTGGATAACCTACCTGCTAGACCGGACAACTTGGAAACGAGGGCTAATACCGGATGAGCTT
AATTAGTGGCATCACTGATTAAGGAAAGATGGCTCTGAAATGCTATCGTTAGTAGATGGATCCGCGTC
TGATTAGCTAGTTGGTGGGTAAAGGCTTACCAAGGCGACGATCAGTAGCCGCGCTGAGAGGGTGACCG
CCACACTGGGACTGAGACACGGCCAGACTCTTACGGGAGGCGAGCTGGGGAATCTTCCGCAATGGACG
AAAGCTCGACGGAGCAACGCCGCGTGTACGACGAAGGCCCTTCGGGTTGTAAGTACTGCTTCAGGGACG
AGCGTAAGTATGTAATAATGTACTTACATGACGGTACCTGAGGAGGAAGCCCGGCTAACTACGTGCC
AGCAGCCGCGGTAACTACGTAGGGGGCAAGCGTTGTCGGAATCATTGGGCGTAAAGGGCGCGTAGCGGA
TACTTAAGTCTGGTGTAAACCTAGGGCTCAACCTGGGACTGCATCGGAACTGGGTATCTTGAGGAC
AGGAGAGGAAAGTGAATTCACGCTGTAGCGGTGAATGCGTAGATATGTTGAGGAAACACAGTGGCGAA
GGCGACTTTCTGGACTGTAACCTGACGCTGAGGCGCAAGGCTGGGGAGCAACAGGATTTACTCAATCCG
GTAGTCCACGCCGTAACGATGAGTGCTAGGTGTAGAGGGTATCGACCCCTCTGTGCCGAGTTAACAC
ACTAAGCACTCCGCTGGGAGTACGGCCGAAGGTTGAACTCAAAGGAATGACGGGGGCCGCGACAA
GCGGTGGAGCATGTGGTTTAATTCGACGCAACGCGGAAGAACCTTACCAAGGCTTGACATCCATAGAATCT
TTGGGAAACCAGAGAGTGCCCTTCGGGAGCTATGAGACAGGTGGTGATGGTTGTCTGCTAGCTCGTGTC
GTGAGATGTTGGGTTAAGTCCCGCAACGAGCGCAACCCCTATGTTAGTTGCTAACCGGTAATGGTGAGC
ACTCTAGACAGACTGCCGCTGACAAACCGGAGGAAGGTGGGGATGACGTCAATCATCATGCCCTTATG
TCTTGGGCTACACACGCTGCTACAATGGCCAGTACAGACGGAAGCGAAGCTGTGAGGTGAAGCCAATCCGA
GAAAGCTGGTCTCAGTTGCGGATTGTTCTCTGCAACTCGAGAACATGAAGTCGGAATCGCTAGTAAATCGCA
GGTCAGCACTATGCGGTGAATACGTTCCCGGCTTGTACACACCGCCGCTCACACCAGAAAGTCTGCA
ACACCCGAAGCCGCTGAGGTAACCCGAAGGGAGCTAGCCGTCGAAGGTGGGCGGATGATTGGGTGAA
GTCTGAACAAGGTAGCCGTATCGGAAGGTGCGGCTGGATCACCT
```

Presence / Absence  
Abundance (OTU table)

OTUs of Soil sample

```
>HMPREF9498_00001 hypothetical protein [Enterococcus faecalis, TX4248]
TTGCAAGCGGAGCAACATGGTACCATTGTCGATATTTGAACGACTCTACCAAAATA
CTGGTAGGAGTCGTTCAAATTTTGGGGTCCCCCAATCCCAGCTGCTGTTATTTTGGCT
TTTACTGTGAAATTCGCTTAATTTCTCCTGTTTTCGGAAGGACATAAGGGGAACGGG
GGGGCACACA
>HMPREF9498_00002 hypothetical protein [Enterococcus faecalis, TX4248]
ATGGGATTAATGACAGGATACACATGACAAAAGGAGTACGAAGTCACACAGGCTGTA
CTGGAGCGCACTAGATAAGTTTGAGAGTGAACTAGAAGCGCTATGCTTGATTGGAAGCCA
CTTGACTTATCAGATAGTACTGATGTCGAGACGAACGCTAAGAATGTTTAAATTAT
CTAAGTCGGTCATACATTCTACTTATTGATACGAGCTACACAAAGTACGCTAGAAGAA
GATATTTCTGTTGGGATTTGAAGATTGACGGTCAATCTAGACGAAGAAGAATAA
>HMPREF9498_00003 initiator RepB protein [Enterococcus faecalis, TX4248]
ATGAGAGATAAATCGGAATAGCTGTTAAATATCAAAATGAGTTAAATCTTGTGCCACTT
AAAACTTTAATGCAAAACAAATGGATCTGCTTTTGGCTTATGCGCAAGTAAGAAGAT
AAAGGAGTAGAGAATATTCTGTTCTTCTCGAGGAATTAAGGAATGAGTGAATACAAA
ATGACTGCTACAAAAGCTTTGTAGACAGACTAGAAAGTTGATAAAGATGTCGTAAC
CTTTCTTATCGTACAGAAAAGTACGCAAAATAGAGTATTTGTTTATCAATGGATT
AAAATCGACAAAACAAAAGTTTGTGAGGTTGAGGTAACTCAAGATTAGACTATATA
ATAAATGGTTTAAACACAGAGTTTAGTAGGTTTGAAGCTGCGCTTCTTACTTCAATCGC
TCAACATATGCTAAACATTGTTGTTTAAATGCAATATGCTCAACAGGTTATTAT
GTAGTCAGTATTGAAGACTCTCAGAGAATTTGTAGACATACCGCAATATTATCAATGGGT
AACATTGACCAAAAGTTTAAACACGCAATGAAGGAATCCATCAATTTTGAGAAT
TTAGAAGTAAACAAAATCAAGCCAAAAGGCAATAGGTTGCCAACTAGAAATTTACG
TTTACTGGTCTGAAACCAACAAACCAAGCTAACCATGACGATGGTAAACGGAGAA
TAA
>HMPREF9498_00004 CHAP domain protein [Enterococcus faecalis, TX4248]
ATGAGAAAGATAAAAAATATGGCGCTTTTATAGGATTTTATAGTTGTATGAGCTTT
TCTTTTAAATCAGGGGTAGTAAATGCAAGAGATAAAGTAAAGTGATGATTAATCTGCT
AAATGGAAAAATCTGCTTTAGGTGCAACTACTGATGATTGGGGTATGGCTACAAGATCA
TGTAATCTTCTTGTGGCAATCGATTGAGTGGGAAATAGTTGATATAGACGTGAT
GGTTAGACTGGAATGCTAACCAATGGGGAATATGCTAGAAGGCAAGGCTATAGGTA
GATATGATACCCAGAGGAGTCAAGCTTATGGAACAGTATGATCATGTAGCTTGG
GTAGCGCGAGTTGATGGGATAGAGTTTGTGTTGAAAGATAAATGATGATATACAGA
AGATATAATTCTAGATGATTAGTAAAAACTCTGTTGATGGGTATATACATTTTAAAGAT
ATATCTCTCTCTCTGAAAAACAGCTATTAAATAATATATTACTATAACAAAGATGGA
CAAGATTTTTTTGAAATATTGAGTGTACTACAAAAGCAAAACAGTAGAGGAAGA
ATTTATTATGTAAGATAGGCTACAAATTTGCTAACGGACAAATTTGATCTCTGTATAT
GAAGATGAAAAGGAACAGCTGGATTGGATATTAAAGCAAGGATCTTCAGAAATTTACA
CAGTTAGAAACTTTTAAACGAAAAAAGTTTCTTAAACGAATATGCGCAACACTTTGG
AAAGATTTTAAATATTCAAGGGAATTTAGGAGTTCAGTGGGAAATGATACACGGTA
AACCGAGTATTCTATACTCTACCAATTAATGCTATCTATCATGATTTGATCGAGGGAAC
ACGTTCTGCGGTTGTGAATAGTGGGCGATAAAGATATAACACGCAACACGTAAT
AAATATAGCACTATCAGAGGCTAATCAGGATATTTTGGTAATATGAAATGCAAGAT
ACTAAAGGAAAAACAGAGAGGAAGAATTTATTACTAGTAGGCTATAATTTAGGC
GACGGTAGTGTTCTGTATTCATTGTACAAAGTAGAGCTGACGAGTAAGGTGAGTTAT
GTAAGAGAGATGTG
>HMPREF9498_00005 hypothetical protein [Enterococcus faecalis, TX4248]
GTGCTTGAGCAAGCGAAAAATGGGACAAACGATCCGCTGATCGTTGTTCCACGAC
GAAGAAAAGTTAGAAGAGAACAGCAAGTGTATAGAAGAAAAATGAATACATATGAAGAA
GAGTGGCAGGGTTACGACGAAGGAGAACTAGTAA
>HMPREF9498_00006 hypothetical protein [Enterococcus faecalis, TX4248]
TTGTCACTGAAAAACAATTTGCTAACCAATAGCGACAAATCAAGAATTAGATCAATTA
GCACGGCAAGCACAAGAAATGCTAAAAAGCAGAAAGTTGAGGAGGCTTAGAGACTATC
AAGCTTTTATCGAGTCTTGA
>HMPREF9498_00007 hypothetical protein [Enterococcus faecalis, TX4248]
ATGAGTAGTGTGTTTGTCAAAGTTTGGGGCGAATGCGGTTAGATGGTACGATCCGCT
GATCAAGTGGTGAATGAAGTATTGTCTTCAAGCAAGCTTTTATAAAGGATGATGGCG
GTCGAGTCAGCCATAAAGTTTGTGGGGAAAAAGGAAGCGGAGACGGATTTCTGAATAC
CAAGAAGCGATATTATCATCAATTTGGCTTGAAGATCAAGTATTAAGTGAAGAATTAT
TCATTATCGATATTCGTTAACTATTATTGTTGGTTGATTTGGATATTTTATAGGA
GCATTGTTATCTGTGGTGGTTTCTTCTTCAAAAACATCGAGCTTGGTACGCCAATTTA
```

OTUs of Water sample

```
>HMPREF9498_00001 hypothetical protein [Enterococcus faecalis, TX4248]
TTGCAAGCGGAGCAACATGGTACCATTGTCGATATTTGAACGACTCTACCAAAATA
CTGGTAGGAGTCGTTCAAATTTTGGGGTCCCCCAATCCCAGCTGCTGTTATTTTGGCT
TTTACTGTGAAATTCGCTTAATTTCTCCTGTTTTCGGAAGGACATAAGGGGAACGGG
GGGGCACACA
>HMPREF9498_00002 hypothetical protein [Enterococcus faecalis, TX4248]
ATGGGATTAATGACAGGATACACATGACAAAAGGAGTACGAAGTCACACAGGCTGTA
CTGGAGCGCACTAGATAAGTTTGAGAGTGAACTAGAAGCGCTATTGCTTGATTGGAAGCCA
CTTGACTTATCAGATAGTACTGATGTCGAGACGAACGCTAAGAATGTTTAAATTAT
CTAAGTCGGTCATACATTCTACTTATTGATACGAGCTACCAAGATACGCTAGAAGAA
GATATTTCTGTTGGGATTTGAAGATTGACGGTCAATCTAGACGAAGAAGAATAA
>HMPREF9498_00003 initiator RepB protein [Enterococcus faecalis, TX4248]
ATGAGAGATAAATCGGAATAGCTGTTAAATATCAAAATGAGTTAAATCTTGTGCCACTT
AAAACTTTAATGCAAAACAAATGGATCTGCTTTTGGCTTATGCGCAAGTAAGAAGAT
AAAGGAGTAGAGAATATTCTGTTCTTCTCGAGGAATTAAGGAATGAGTGAATACAAA
ATGACTGCTACAAAAGCTTTGTAGACAGACTAGAAAGTTGATAAAGATGTCGTAAC
CTTTCTTATCGTACAGAAAAGTACGCAAAATAGAGTATTTGTTTATCAATGGATT
AAAATCGACAAAACAAAAGTTTGTGAGGTTGAGGTAACTCAAGATTAGACTATATA
ATAAATGGTTTAAACACAGAGTTTAGTAGGTTTGAAGCTGCGCTTCTTACTTCAATCGC
TCAACATATGCTAAACATTGTTGTTTAAATGCAATATGCTCAACAGGTTATTAT
GTAGTCAGTATTGAAGACTCTCAGAGAATTTGTAGACATACCGCAATATTATCAATGGGT
AACATTGACCAAAAGTTTAAACACGCAATGAAGGAATCCATCAATTTTGAGAAT
TTAGAAGTAAACAAAATCAAGCCAAAAGGCAATAGGTTGCCAACTAGAAATTTACG
TTTACTGGTCTGAAACCAACAAACCAAGCTAACCATGACGATGGTAAACGGAGAA
TAA
>HMPREF9498_00004 CHAP domain protein [Enterococcus faecalis, TX4248]
ATGAGAAAGATAAAAAATATGGCGCTTTTATAGGATTTTATAGTTGTATGAGCTTT
TCTTTTAAATCAGGGGTAGTAAATGCAAGAGATAAAGTAAAGTGATGATTAATCTGCT
AAATGGAAAAATCTGCTTTAGGTGCAACTACTGATGATTGGGGTATGGCTACAAGATCA
TGTAATCTTCTTGTGGCAATCGATTGAGTGGGAAATAGTTTATATAGACGTGAT
GGTTAGACTGGAATGCTAACCAATGGGGAATATGCTAGAAGGCAAGGCTATAGGTA
GATATGATACCCAGAGGAGTCAAGCTTATGGAACAGTATGATCATGTAGCTTGG
GTAGCGCGAGTTGATGGGATAGAGTTTGTGTTGAAAGATAAATGATGATATACAGA
AGATATAATTCTAGATGATTAGTAAAAACTCTGTTGATGGGTATATACATTTTAAAGAT
ATATCTCTCTCTCTGAAAAACAGCTATTAAATAATATATTACTATAACAAAGATGGA
CAAGATTTTTTTGAAATATTGAGTGTACTACAAAAGCAAAACAGTAGAGGAAGA
ATTTATTATGTAAGATAGCTACAAATTTGCTAACGGACAAATTTGATCTCTGTATAT
GAAGATGAAAAGGAACAGCTGGATTGGATTTGATTTAAGCAAGGATCTTCAGAAATTTACA
CAGTTAGAAACTTTTAAACGAAAAAAGTTTCTTAAACGAATATGCGCAACACTTTGG
AAAGATTTTAAATATTCAAGGGAATTTAGGAGTTCAGTGGGAAATGATACACGGTA
AACCGAGTATTCTATACTCTACCAATTAATGCTATCTATCAATTTGATCGAGGGAAC
ACGTTCTGCGGTTGTGAATAGTGGGCGATAAAGATATAACACGCAACACGTAAT
AAATATAGCACTATCAGAGGCTAATCAGGATATTTTGGTAATATGAAATGCAAGAT
ACTAAAGGAAAAACAGAGAGGAAGAATTTATTACTAGTAGGCTATAATTTAGGC
GACGGTAGTGTTCTGTATTCATTGTACAAAGTAGAGCTGACGAGTAAGGTGAGTTAT
GTAAGAGAGATGTG
>HMPREF9498_00005 hypothetical protein [Enterococcus faecalis, TX4248]
GTGCTTGAGCAAGCGAAAAATGGGACAAACGATCCGCTGATCGTTGTTCCACGAC
GAAGAAAAGTTAGAAGAGAACAGCAAGTGTATAGAAGAAAAATGAATACATATGAAGAA
GAGTGGCAGGGTTACGACGAAGGAGAACTAGTAA
>HMPREF9498_00006 hypothetical protein [Enterococcus faecalis, TX4248]
TTGTCACTGAAAAACAATTTGCTAACCAATAGCGACAAATCAAGAATTAGATCAATTA
GCACGGCAAGCACAAGAAATGCTAAAAAGCAGAAAGTTGAGGAGGCTTAGAGACTATC
AAGCTTTTATCGAGTCTTGA
>HMPREF9498_00007 hypothetical protein [Enterococcus faecalis, TX4248]
ATGAGTAGTGTGTTTGTCAAAGTTTGGGGCGAATGCGGTTAGATGGTACGATCCGCT
GATCAAGTGGTGAATGAAGTATTGTCTTCAAGCAAGCTTTTATAAAGGATGATGGCG
GTCGAGTCAGCCATAAAGTTTGTGGGGAAAAAGGAAGCGGAGACGGATTTCTGAATAC
CAAGAAGCGATATTATCATCAATTTGGCTTGAAGATCAAGTATTAAGTGAAGAATTAT
TCATTATCGATATTCGTTAACTATTATTGTTGGTTGATTTGGATATTTTATAGGA
GCATTGTTATCTGTGGTGGTTTCTTCTTCAAAAACATCGAGCTTGGTACGCCAATTTA
```