

TAXONOMIC AND FUNCTIONAL METAGENOMIC
ANALYSIS OF OCEAN BACTERIA

DEPARTMENT OF COMPUTER SCIENCE AND ENGINEERING
JADAVPUR UNIVERSITY

SUBMITTED BY

DIPESH PAL

REGISTRATION NUMBER: 133682

CLASS ROLL NUMBER: 001510503020

EXAMINATION ROLL NUMBER: MCA186019

UNDER THE SUPERVISION OF

PROF. ANASUA SARKAR

DEPT OF COMPUTER SCIENCE AND ENGINEERING

OF

JADAVPUR UNIVERSITY

MAY 2018

DECLARATION OF ORIGINALITY AND COMPLIANCE AND **ACADEMIC THESIS**

I hereby declare that this project entitled "TAXONOMIC AND FUNCTIONAL METAGENOMIC ANALYSIS OF OCEAN BACTERIA" contains original research work done by the undersigned candidate, as part of his Master of Computer Application (MCA).

All information in this document has been obtained and presented in accordance with academic rules and ethical conduct.

I also declare that, as required by these rules and conduct, I have fully cited and referenced all materials and results that are not original to this work.

Name: Dipesh Pal

University Registration Number: 133682

Examination Roll Number: MCA186019

Thesis Title: TAXONOMIC AND FUNCTIONAL METAGENOMIC ANALYSIS OF OCEAN BACTERIA

(Signature with date)

CERTIFICATE OF RECOMMENDATION

This is to certify that the thesis entitled “TAXONOMIC AND FUNCTIONAL METAGENOMIC ANALYSIS OF OCEAN BACTERIA” has been satisfactorily completed by Dipesh Pal (University Registration Number: 133682 Examination Roll Number: MCA186019). It is a bonafide piece of work carried out under my guidance and supervision and be accepted in partial fulfilment of the requirement for the Degree of Master of Computer Application, Department of Computer Science and Engineering, Faculty of Engineering and Technology, Jadavpur University, Kolkata.

Prof. Anasua Sarkar (Thesis Supervisor)
Department of Computer Science and Engineering
Jadavpur University, Kolkata-700032

Countersigned

Prof. Ujjwal Maulik
Head, Department of Computer Science and Engineering,
Jadavpur University, Kolkata-700032

Prof. Chiranjib Bhattacharjee
Dean, Faculty of Engineering and Technology,
Jadavpur University, Kolkata-700032

CERTIFICATE OF APPROVAL

This is to certify that the thesis entitled "TAXONOMIC AND FUNCTIONAL METAGENOMIC ANALYSIS OF OCEAN BACTERIA" is a bonafide record of work carried out by Dipesh Pal in partial fulfilment of the requirements for the award of the Degree of Master of Computer Application in the Department of Computer Science and Engineering, Jadavpur University during the period of January 2018 to May 2018. It is understood that by this approval the undersigned do not necessarily endorse or approve any statement made, opinion expressed or conclusion drawn there in but approve the thesis only for the purpose for which it has been submitted.

Signature of External Examiner

Date:

Signature of Supervisor

Date:

ACKNOWLEDGEMENT

First and foremost, I would like to start by thanking God Almighty for showering me with the strength, knowledge and potential to embark on this wonderful journey and to persevere and complete the embodied research work satisfactorily.

I am pleased to express my deepest gratitude to my thesis guide, Prof. Anasua Sarkar, Department of Computer Science and Engineering, Jadavpur University, Kolkata for her invaluable guidance, constant encouragement and inspiration during the period of my dissertation.

I am highly indebted to Jadavpur University for providing me the opportunity and the required infrastructure to carry on my thesis.

I am thankful to all the teaching and non-teaching staff whose helping hands have smoothed my journey through the period of my research.

Last but not the least; I would like to thank my family members, classmates, seniors and friends for giving me constant encouragement and mental support throughout my work.

Dipesh Pal

University Registration Number: 133682 of 2015-2016

Examination Roll Number: MCA186019

Master of Computer Application

Department of Computer Science and Engineering

Jadavpur University

TABLE OF CONTENTS

CHAPTER 1: INTRODUCTION	8
1.1 INTRODUCTION TO BIOINFORMATICS AND METAGENOMICS.....	8
1.2 IMPLEMENTATION FRAMEWORK.....	9
1.3 SEQUENCES.....	10
CHAPTER 2: LITERATURE REVIEW	12
2.1 Introduction to Literature Reviews.....	12
2.2 Survey of Related Work.....	12
CHAPTER 3: METHODOLOGIES	13
3.1 GETTING STARTED.....	13
3.2 ORF FINDERS SMS & NCBI.....	13
3.3 ORF CLASSIFICATION.....	14
3.4 ORF SELECTION:	14
3.5 CHECKING CONSERVED PROTEIN DOMAIN EXISTENCE USING INTERPRO.....	15
3.6 BLAST.....	15
3.7 MULTIPLE SEQUENCE ALIGNMENT.....	16
3.8 PHYLOGENETIC TREE.....	16
CHAPTER 4: EVALUATION	17
4.1 DATA ANALYSYS AND RESULTS.....	17
4.2 DISCUSSION.....	52
CHAPTER 5: RESULT SUMMERY	62
5.1 CONCLUSION.....	62
REFERENCES	63

ABSTRACT

Today it is a big issue to identify some sort of clue or sample in case of forensic research or investigation procedure. In such cases bioinformatics comes to help us with its vast databases and tools. There are also many researches going on the improvements of bioinformatics. Bioinformatics can help us with the information which has been developed using computer technology and hence a lot more efficient and precise.

In our project we have used this bioinformatics to identify some sequences with their functions, activities, belongings, molecular components and many more things. We are first given a nucleotide sequence. Then we analyse that sequence with various bioinformatics tools like INTERPRO, BLAST, CLUSTAL OMEGA etc. The very first input for the very first tool is the given sequence and next we use the output information to some other tool. So, our job is to follow a systematic procedure for approaching the final result of our analysis. And at the last stage of our analysis we reach to the conclusion that what functionalities the sequence have, where it belongs, what kind of sequence it is and some other information. Thus, finally our project gets completed with all the analysis methods and achieved results.

CHAPTER 1: INTRODUCTION

1.1 INTRODUCTION TO BIOINFORMATICS AND METAGENOMICS:

The term “Bioinformatics” means the science of developing computer databases and algorithms to facilitate and expedite biological research. Actually, it is the application of computer technology to the management of biological information. This project is all about to use bioinformatics methods and tools for analysing a nucleotide sequence to identify the properties of it. Here, our aim is to collectively annotate distinct DNA fragments randomly distributed from the available public sequence pool and find out the putative function of the protein product, as well as the most likely taxonomic classification of the host organism. Here we will be using the online bioinformatics tools and data achieved from them. The final result will come after all kind of required analysis of the sequence. In this case, at the time of experimenting with the sequences, the output of one analysis using some tool may be the input of some other tool. Thus, after crossing all levels of analysis we could reach our desired goal. Now, we are about to get into the details by analysing our sequence practically.

Metagenomics is the study of genetic material recovered directly from environmental samples. The broad field may also be referred to as environmental genomics, ecogenomics or community genomics. While traditional microbiology and microbial genome sequencing and genomics rely upon cultivated clonal cultures, early environmental gene sequencing cloned specific genes to produce a profile of diversity in a natural sample.[1] Such work revealed that the vast majority of microbial biodiversity had been missed by cultivation-based methods. Several other works on metagenomics analysis on microbial community exist [9-13].

Here first we find the Open Reading frames(ORFs) using SMS and NCBI tools. Then we classify the ORFs using some rules and finally choose one ORF. Next, we check the conserved protein domain existence with the help of INTERPRO and then find the BLAST hits with BLAST and SMARTBLAST. Then we use the BLAST hits and using them we get some matrices or tables. Thereafter we find the multiple sequence analysis by ClustalW and using the alignments we find the corresponding phylogenetic tree using PhyML. Also, we get the classification details from our analysis. At last, we get our results with the biological processes, Molecular function, cellular component etc.

1.2 IMPLEMENTATION FRAMEWORK:

For our projects we will be using some online bioinformatics tools mainly. But list of all tools and software required for our metagenomic analysis are here:

- ❖ Sequence Manipulation Suite (SMS)

The sequence manipulation suite is a collection of web-based program for analysing and formatting DNA and protein sequences. The output of each program is a set of HTML commands, which is rendered by our web browser as a standard web page.[3] We can print and save the results and we can edit them using an HTML editor or a text editor. It helps us in finding possible ORFs.

- ❖ NCBI ORF Finder

The NCBI houses a series of databases relevant to biotechnology and biomedicine and is an important resource for bioinformatics tools and services. It also helps us in finding possible ORFs.

- ❖ INTERPRO

INTERPRO is a database of protein families, domains and functional sites in which identifiable features found in known proteins can be applied to new protein sequences in order to functionally characterise them. The contents of INTERPRO consist of diagnostic signatures and the proteins that they significantly match.

- ❖ Definition List

It is a tool that creates a table which lists all the distinct functions of the homologs detected by BLAST with their range of E-values.

- ❖ TaxReports

This tool is used to extract the taxonomic information from the sequences producing significant results of BLASTp hits.

- ❖ Taxonomy List

It is applied for building a synthesis table of our observations. A synthesis table is a that allows a researcher to sort and categorize the different arguments e.g. E-values presented on an issue.

- ❖ Clustal Omega and ClustalW

Clustal Omega is a tool to verify that the ORF integrates convincingly in its presumed homolog family[5]. We use both of them to find multiple sequence alignments in FASTA, PhyML or Clustal format. Where the Clustal Omega gives the output based on our provided sequences only, ClustalW generates the output by checking all available data resources residing in their vast databases.

- ❖ PhyML

This tool helps us in finding the actual phylogenetic tree with the name of the predicted classes for our sequence.

1.3 SEQUENCES:

We have analysed about 15 unknown sequences without taxonomy classification and functional annotation from Tara Ocean Expedition and from then we have chosen four major sequences. We didn't get proper biological functions related to the conserved protein domains. That is why we are selecting the four sequences with proper biological functions. Here we are going to use and analyse the sequences below. Besides we have analysed some other sequences also. But still there are four major sequences we have selected for our analysis. These are here:

Sequence 1:

>GOS_7191010 [Global Ocean Sampling expedition: Open Ocean: Tropical South Pacific: International: 600 miles from F. Polynesia]

```
CAATCTTCAGTTAGGATATAGTCATGATATTGATTTTGATATACCTGAAGGTATAAAAATTAACGTGGAAAAACAAACA
ACATTAATAAATTTTCAGGTTTTGATAAACAACAAGTTGGATCAGTTGTATCAAATTAATAAATTTGCGTAAAATTGAACC
ATATAAAGGCCAAAGGTATTAGAGAAAAAGGTCAATATGTTCTTAAAAAAGAGGGAAAGAAAAATAATGAAATTAAC
ACCAGTAAAAGAAAAAGATTTAGAGTAAGTAACAAAGTAAAAAATGTTGCCTCAAATAATAGATTTAGATTAAGTATAT
CTAGATCTTCAAAAAATATTTTCAGCACAAATTATTGACGATAAAAAAAGTGTAACCTTAATTTTCAGCTTCGTCTATTGAA
AAAGATATTAACCTAATACTAAAATAAATAAACAGAAATTATCAAAAATTGTAGCTGAAAGACTTGCTAAAAAAGCTC
AAGAGAAAAAGATAACAAAAATTTTTTCGATAGGGGAATTTATAAATATCATGGAAGAGTAAGAGTTTTTCGCTGATAC
ACTTAGAAAAAATGGAATGGAATTTTAAAATGGAAAATATTAAGACAAAAAATGATGATATCTTAGAAAAGTTAG
TTCATATTAATCGTATTACAAAAGTTGTAAGGTTGGAAGAAGTTTGGTTTCTCAGCTCTTGTAGTGGTTGGTAACCA
AGCAGGAAGAATTGGTATAGCACATGCCAAAGCTAAACAAGTTCCTGATGCTATTAATAAAGCAAATGAAATGGCTAG
AAGAAAATAATTACATACCACTTAGAGAACGTCGAACAATACATCATGATGTGAAAGCAAAGATGGTTCTGGAAA
AATTGTATTAAGGGCAGCCTCTAAAGGAAGTATTATAGCTGGAGGACCTGTGCGAGCAGTTTGTGAAGTTTTAGT
GTAAAAGA
```

Sequence 2:

>GOS_7191020 [Global Ocean Sampling expedition: Open Ocean: Tropical South Pacific: International: 600 miles from F. Polynesia]

```
TACTCTGTCAGTGGTATTTAATAATTTCTTACTTAAAAAGAAGGGCAAATATAAAGATACGGGAAATATTCTTCCAG
GGCATGGCGGTATCTTAGATAGAGTAGACGGAATAATCTTTGCAATTCCAATTGGTATTTAATATATTTTTTTTAGGA
CTTAGTTAATGATCAAAAAAATTGCAATTTTAGGATCAACAGGGTCAATTGGAAAATCTTTATTAATAAATTATCGAAGCA
GACAGAAAAAGTTTTAAAATTTTTGTTAACAGCAAATAAAGATTACACATCTTTAATTAACAAGCTAACAAATTTAA
TGTTAAGAATATAATTATCACTGATCAGAAAAATTACAATATTGCGAAGCTAAAAATTAATAAATAAATAAATAAATTT
TTAATAATTATAATCAATTTAAAATATTTTTAAAATAAATAAATTGATTATGCGATGAGTGCAATTACCGGATTAGATGGT
TTAAAACCAACAATAGATATAATTAATAATACAAAAAATATTGCTATTGCAAATAAAGAGTCGATTATATGTGCGTGGG
ATATTATAAAAAACAGAAATTAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAA
TAAAAGATTTTAACAATAGTTCAATTGAAAAAATTTATCTTACTGCATCAGGTGGACCTTTTTATGCTACGCCAATTAATA
AATTAAGAATATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAA
TACAATGATAAATAAGATTTATGAGTCAATAGAAGCAAAAAATATTTTTAATATTCTTACAAAAAATTGAAATAATTA
TACATCCAAATCATATGTGCACGCTTAATTAATTAATCAATAGTGGACCCATAAAAAATTATAGCTCACGATACGACAATG
ATGATTCCATATTTAATACTTTGTATTTAACTCTGACAAAAAATAAAGTCTAAAGAACTTGATTTTAAA
```

Sequence 3:

>GOS_3889010 [Global Ocean Sampling expedition: Open Ocean: Tropical South Pacific: International: 600 miles from F. Polynesia]

GGTAGTTATATAAAATTTAATATTATCCGCATTATTAGTATCTGTTCTGTATTTGGGAGGTTGGGGATTTCTATACCAGTT
GAAACTATTTCAAGAATTTTAAATATTCCAATAGATACATCAGTAATACAAGTAATTACTGCAACCATAGGTATTGTAAT
GACAATTCCTAAAGCATATTTATTAGTTTTTATAGCAATCTTACTGAGATGGACTACACCTAGAGTGAGGATCGATCAAT
TACTTGATTTAGGCTGGAAGTTTCTTTTACCAATTTTCATTGGCAAACCTTTTATTAAGTGCAGGCTTAAACTAGCATTAC
CAAATTTTTTTGGCGGATAATTTTAAAGTATTTTACTTACCAAATTCAAAAAATGTTGTATAATAATGTTAAGTTAAGT
TTCTGACAATGAAAGATTTTTTTCAAAAAGTCAATAGTTATGTTAAAGAAGCTTTTGGAGCGAGTAAATATTTGTTTGAC
GGAATAACTGTTACCTTTGACCATCTTAAAAGGAGGCCTGTAAGTGTCCAATATCCCTACGAAAAGCTTATTCCATCTGA
AAGATATAGAGGAAGAATTCATTATGAGTTTGATAAATGTATTGCATGTGAAGTTTGTGTGCGAGTTTGTCCAATAAAT
TTACCGGTAGTTGACTGGGTAATGAATAAAGAGACCAAGAAGAAAGAGTTAAGAAATTACTCAATTGATTTTGGTGCA
TGCATTTTTTGCAGTAATTGTGTTGAATATTGCCCTACAAATTGTTTATCAATGACTGAAGAATATGAATTAGCAACTTTT
GATAGGCATAATTTAAATTACGATAATGTTGCTCTTGAAGATTACCAACCAATGTCACCTTCTGATCCAAGTGTCAAATC
ACTTAAAGAGTTAACTTATCTTCCAAAACGAGTAATGGATCCTCATGAAG

Sequence 4:

>GOS_1767030 [Global Ocean Sampling expedition: Open Ocean: Tropical South Pacific: International: 600 miles from F. Polynesia]

TTTATCACGAGGACAAACAGTTAATAATGTTGAAATAGACAAAGCTGTTTTATGTCCTTTAAAGCCTGGAGAGGCTTCAT
TACATCACGGTTTAAACTTTCATGCTTCAAACCAAATAATAGCAAAGATAGAAGAATAGGATTGAATTTTCAATACATT
ACCCCTGATTTAAAGCAATTAAGTAAAGATGATAGTGCTTTGTGCGTTAGAGGAGAAGATAATTATAATTTTTATA
AAAAAGATGTAATTGCTAATGAAAATTTTGATATAAAGGGACATCAAAGGCTTATTGAGCGAACAGAACATTACAAAA
AAACAATAAGAGAAAATTAATTTTATATGAAATACAGAGAGTTTAGTGATTTGGGATGGAAAGTTTCTGAAATTGGAAT
AGGGTGTGGGCTATTGGTTCTGAGTGGGGTGATGTATCAATAGAAGATGCTAAAGAAGCACTTTATACTTCCATTGAC
CATGGAATTAATTTTTTTGATACAGCAGATGTCTATGGAGATGGCAGAAGTAAAAAATAATTAGTGAGGTATTAATA
ATTCTAATGAACAAATATATGTTGCTACAAAAGCAGGGAGAAGATTAACCCACATAATGCAGATGGATACAATCTAAA
AAATATTGAAAGTTTTATTGATAGATCTTAGCAATTTAGGAGTTGATATAATTGATTTAGTTCAATTACATTGTCCTCC
ATCAGAGATATGTGGAAAGCCAGAAACATATGAAATGATGGATGAGATTGTTAAAAAAGGTAAGATAAAATATTATGG
AGTTAGTGTTGAAAAAGTATCTGAGGCTTTAGATGCTATTCAGCACCCAAATGTAAAGAGTATACAAATTATTTAATA
TATTTGCCAAAACCAAATGAAATATTTTTTCAAGAAGCTAAAAAATAA

CHAPTER 2: Literature Reviews

2.1 Introduction to Literature Reviews:

A literature review discusses published information in a particular subject area, and sometimes information in a particular subject area within a certain time period. It can be an analysis of literature or published sources, on a particular topic.

A literature review is a simple summary of the sources, but it usually has an organizational pattern and combines both summary and synthesis. A summary is a recap of the important information of the source, but a synthesis is a re-organization, or a reshuffling, of that information. It might give a new interpretation of old material or combine new with old interpretations. Or it might trace the intellectual progression of the field, including major debates. And depending on the situation, the literature review may evaluate the sources and advise the reader on the most pertinent or relevant.

2.2 Survey of Related Work:

From September 2009 to December 2013, the oceanographic ship **Tara** sampled plankton across all of the planet's oceans (right, bottom). Ocean plankton produce half the oxygen we breathe. If terrestrial forests fill up our first lung, oceans fill up the second. The ocean's microscopic planktonic life is also an important carbon sink, with far reaching implications in climate change scenarios. Hence, the Tara Oceans pan-oceanic international expedition aims to define the current state of plankton biodiversity, from tropical coral reefs to Antarctica, from viruses to fish larvae.

Metagenomic sequencing of the plankton DNA contained in the thousands of Tara Oceans samples has begun at the GENOSCOPE (France). These short pieces of DNA contain precious information on the identity of the plankton species which populate seawater, as well as hints to the metabolic functions at work in these tiny ocean drifters. At this molecular scale, bioinformatics is the essential tool to shed light on the information locked up in the DNA sequences.

CHAPTER 3: METHODOLOGIES

3.1 GETTING STARTED:

Now, we are going to discuss the procedure of how we have analysed the sequences in detail. For the four sequences, mentioned before, the methods are same. So, we shall mainly discuss the process with only one sequence only. We have taken the first DNA sequence as our primary data.

We have got our sequence from the public sequence pool available on the internet. First of all, we need to create an account for that website with the given instructions. Then we can see, by going to the cart option after successful login, that our cart is empty initially. Next, we have to select an ocean sampling location and click on “Add a new sequence to your cart”. We can add a new sequence into our cart when the cart is empty, or when we have annotated all the available sequences in our cart. We can view our sequence with a unique accession number inside the cart anytime.

3.2 ORF FINDERS SMS & NCBI:

Now, our first job with the sequence-1 will involve the putative Open Reading Frames (ORFs). There are two major tools that we have used for our work.

- Sequence Manipulation Suite (SMS)
- NCBI ORF Finder

➤ In case of SMS we shall use the following criteria:

- Do not contain any STOP codons
- Contains at least 75 codons
- Can be on either direct or reverse strands
- Can be in frames 1, 2 and 3 on each strand
- Can be complete or incomplete at the 5' or 3' ends, or both

➤ In case of NCBI we shall use the following criteria:

- Minimal ORF Length: 75
- Genetic Code: Standard
- ORF Start codon to use: ATG and alternative initiation codons

We use the sequence-1 as our input to both NCBI and SMS tools and get the ORFs and their corresponding protein and nucleotide sequences as output.

3.3 ORF CLASSIFICATION:

Next, we have to classify the ORFs in the following categories:

- **False positive** (Some codons without STOP)
- **ORFan** (Protein coding gene without homologs)
- **Novel** (Protein coding gene having homologs with unknown function)
- **Known** (Protein coding gene having homologs with known functions)

In order to classify the ORFs, the following elements should be considered:

- The presence of homologous proteins (E-value $<1E-10$) ==> NOVEL or KNOWN)
- The ORF size (ORF >100 aa without homologs is probably an ORFan)
- An ORF without homologs is certainly a False positive if this ORF mainly overlaps another ORF with homologs.

3.4 ORF SELECTION USING SMS & NCBI:

If homologs clearly exit, we can conclude that the sequence-1 is coding DNA whatever the ORF size. Otherwise, the true or false positive nature of the ORF will essentially depend on the ORF size. There is not any exact threshold for it. If the DNA fragment doesn't appear to contain any ORFs and is too short then the sequence-1 is "non-coding". If the sequence appears to carry a true coding ORF, maybe it is very long ORF, or it contains many homologs, or both, then the sequence should be considered as "coding"[8].

We shall check all these elements and criteria soon in the next step. We choose a random ORF which is not short in length and not a subsequence of some other ORF in the list. In our case it is ORF3 at the top of the list.

Its direct strand extends from base 548 to base 949. The nucleotide sequence of that is:

```
CTGATACACTTAGAAAAAATGGAATGGAATTTTAAAATGGAAAATATTAAGACAAAAAAAATGATGATATCTTAGAAA
AGTTAGTTCATATTAATCGTATTACAAAAGTTGTAAAAGGTGGAAGAAGGTTTGGTTTCTCAGCTCTTGTAGTGGTTGG
TAACCAAGCAGGAAGAATTGGTATAGCACATGCCAAAGCTAAACAAGTTCCTGATGCTATTAATAAAGCAAATGAAAT
GGCTAGAAGAAAATAATTACATAACCACTTAGAGAACGTCGAACAATACATCATGATGTGAAAGCAAAGATGGTTC
TGGAAAAATTGTATTAAGGGCAGCCTCTAAAGGAACTGGTATTATAGCTGGAGGACCTGTGCGAGCAGTTTGTGAAGT
TTTAGTGTA
```

And the protein sequence of it is:

```
MIHLEKMEWNFKMENIKDKKNDILEKLVHINRITKVVKGRRFGFSALVVVGNQAGRIGIAHAKAKQVPDAIKKANEMAR
RKLHIPLRERRTIHHDVKAKDGSBKIVLRAASKGTGIIAGGPVRAVCEVLV
```

3.5. CHECKING CONSERVED PROTEIN DOMAIN EXISTENCE USING INTERPRO:

Now we'll be checking whether our ORF contains any of the known conserved protein domains stored in one of the domain databases using **INTERPRO** tool. In this case the input would be the above protein sequence.

INTERPRO is a federation of all other databases and it contains all known domains[4]. So, it is better to work with it than other. Although it takes time to analyse the protein sequence submitted as input, still it is a good choice to use INTERPRO for getting efficient result.

Here the selected ORF protein sequence is the input and it gives us the protein family membership, homologous superfamily and the biological processes, molecular function, cellular component etc as output.

3.6. BLAST:

In the next step we go for **BLAST** to identify putative sequence homologs of our ORF in public sequence databases. We use here online BLAST services using NCBI. There are two approaches which can be used to identify homologs of our sequences:

- **BLASTp**: ORF protein sequence against protein database
- **BLASTx**: Here, we need to submit our nucleotide sequence that will be translated by BLAST in all six frames before comparison with protein database. Also, it is useful if there is a doubt as to the ORF location on our DNA fragment, or if our ORF search did not yield any convincing ORFs.

Among the above two approaches we have used **BLASTp**.

Our queries were based on the two following databases mentioned below:

- **NR**: The most comprehensive database available (useful for subsequent phylogenetic analysis)
- **SWISSPROT**: Very small database with highly accurate and informative annotations (useful for subsequent functional hypotheses)

Here we show the data input format and the interfaces used for **BLAST**:

The screenshot displays the NCBI BLAST web interface. At the top, there are navigation tabs for 'blastn', 'blastp', 'blastx', 'tblastn', and 'tblastx'. The main heading is 'Enter Query Sequence'. Below this, there is a text area for 'Enter accession number(s), gi(s), or FASTA sequence(s)' containing a FASTA format sequence: >|c1 [ORF3:548:949: unnamed protein product] followed by a multi-line protein sequence. To the right of the text area are 'Clear' and 'Query subrange' options. Below the text area is a 'Choose File' button and a 'Job Title' field. A checkbox for 'Align two or more sequences' is also present. The 'Choose Search Set' section includes a 'Database' dropdown set to 'UniProtKB/Swiss-Prot(swissprot)', an 'Organism' field, and an 'Exclude' checkbox. There are also checkboxes for 'Models (XM/XP)' and 'Uncultured/environmental sample sequences'. An 'Entrez Query' field is available for limiting the search. The 'Program Selection' section shows radio buttons for various algorithms: 'Quick BLASTP (Accelerated protein-protein BLAST) New', 'blastp (protein-protein BLAST)', 'PSI-BLAST (Position-Specific Iterated BLAST)', 'PHI-BLAST (Pattern Hit Initiated BLAST)', and 'DELTA-BLAST (Domain Enhanced Lookup Time Accelerated BLAST)'. At the bottom, there is a 'BLAST' button and a checkbox for 'Search using Blastp (protein-protein BLAST)'. A 'Show results in a new window' checkbox is also visible.

Input format for BLAST

Now, we use **Definition List** tool to include a table which lists all the distinct functions of the homologs detected by BLAST with their range of E-values.

In the next stage we use this lineage report as an input to the **Taxonomy List** tool for building a synthesis table of our observations. A synthesis table is a that allows a researcher to sort and categorize the different arguments e.g. E-values presented on an issue.

Next, we apply the **TaxReports** tool to extract the taxonomic information from our sequences producing significant results of BLASTp hits which was also achieved from the BLAST result and this list of sequences will be the input for TaxReports tool.

3.7. MULTIPLE SEQUENCE ALIGNMENT:

We need to verify that the ORF integrates convincingly in its presumed homolog family: the alignment must hence present clear well conserved regions. For this we use **Multiple Sequence Alignment**. Secondly, it will serve as the basis for the **Phylogenetic Tree** inference: the alignment must therefore suggest a sufficient number of mutations. Now, we'll proceed to the multiple alignment with the sequences having legible labels in FASTA format achieved from the **TaxReports** tool. These sequences are used as input to ClustalW to find the multiple sequence alignment.

3.8 PHYLOGENETIC TREE

We now need to find the **Phylogenetic Tree**. So, here we use **PhyML** tool for that. Here we use the multiple sequence alignments in FASTA format that was achieved from the ClustalW tool. Finally that **PhyML** gives us the phylogenetic tree.

CHAPTER 4: EVALUATION

4.1 DATA ANALYSIS AND RESULTS:

Now, we are about to watch the results achieved from various tools used to analyse our sequence. At first, we discuss the results for sequence-1 and therefore we see the results for the rest of the sequences.

For sequence 1:

The results after submitting the sequence to the ORF finders we got the following results:

In case of SMS:

ORF Finder results

Results for 953 residue sequence "GOS_7191010 [Global Ocean Sampling expedition: Open Ocean: Tropical South Pacific: International: 600 miles from F. Polynesia]" starting "CAATCTTCAG"

>ORF number 1 in reading frame 1 on the direct strand extends from base 226 to base 582.

```
ATGAAATTAACACCCAGTAAAAGAAAAAGATTTAGAGTAAGTAACAAAGTAAAAAATGTTGCCTCAAATAATAGATTTA
GATTAAGTATATCTAGATCTTCAAAAAATATTTTCAGCACAAATTATTGACGATAAAAAAAGTGTAACCTTAATTTTCAGCT
TCGTCTATTGAAAAAGATATTAACCTAATACTAAAATAAATAAACAGAAATTATCAAAAATTGTAGCTGAAAGACTTGC
TAAAAAAGCTCAAGAGAAAAAGATAACAAAAATTTTTTCGATAGGGGAATTTATAAATATCATGGAAGAGTAAGAGT
TTTCGCTGATACACTTAGAAAAAATGGAATGGAATTTTAA
```

>Translation of ORF number 1 in reading frame 1 on the direct strand.

```
MKLNTSKRKRFRVSNKVNKVASNNRFLSISRSSKNISAQIIDDKKSVTLISASSIEKDIPNTKINKTELSKIVAERLAKKAQEKKI
TKIFFDRGIYKYHGRVRFADTLRKNRMEF
```

>ORF number 1 in reading frame 2 on the direct strand extends from base 548 to base 949.

```
CTGATACACTTAGAAAAAATGGAATGGAATTTAAAATGGAAAATATTAAGACAAAAAATGATGATATCTTAGAAA
AGTTAGTTCATATTAATCGTATTACAAAAGTTGTAAGGTGGAAGAAGGTTTGGTTTCTCAGCTCTTGTAGTGGTTGG
TAACCAAGCAGGAAGAATTGGTATAGCACATGCCAAAGCTAAACAAGTTCCTGATGCTATTAAAAAAGCAAATGAAAT
GGCTAGAAGAAAATAATTCACATACCACTTAGAGAACGTCGAACAATACATCATGATGTGAAAGCAAAGATGGTTC
TGGAAAAATTGTATTAAGGGCAGCCTCTAAAGGAACTGGTATTATAGCTGGAGGACCTGTGCGAGCAGTTTGTGAAGT
TTTAGTGTA
```

>Translation of ORF number 1 in reading frame 2 on the direct strand.

```
LIHLEKMEWNFKMENIKDKKNDILEKLVHINRITKVVKGGRRFGFSALVVVGNQAGRIGIAHAKAKQVPDAIKKANEMARR
KLIHIPLRERRTIHHDVKAKDGSKIVLRAASKGTGIIAGGPVRAVCEVLV
```

No ORFs were found in reading frame 3.

In case of NCBI:

In this we can see that there we found total 8 ORFs where four ORF3, ORF1, ORF4 and ORF2 have direct strands and ORF7, ORF8, ORF6 and ORF5 have reverse strands.

NCBI Resources How To Sign in to NCBI

ORFfinder PubMed Search

Open Reading Frame Viewer [Help](#)

Sequence

ORFs found: 8 Genetic code: 1 Start codon: 'ATG' and alternative codons

ORF3 (133 aa) Display ORF as... Mark

```
>1c1|ORF3
MIHLEKMEMNFKMENIKDKKNDIILEKLVHINRITKVVKGRRRFGPSALV
VVGNQAGRIGIAHAKAKVQPDAIKKANEMARRKLIHPLRERRTIIHDVK
AKDGSGLVLRRAASKGTGIAGGPVRAVCEVLV
```

Mark subset... Marked: 0 Download marked set as Protein FASTA

Label	Strand	Frame	Start	Stop	Length (nt aa)
ORF3	+	2	548	949	402 133
ORF1	+	1	226	582	357 118
ORF7	-	3	603	439	165 54
ORF8	-	3	171	22	150 49
ORF6	-	3	741	634	108 35
ORF5	-	1	161	60	102 33
ORF4	+	3	153	245	93 30
ORF2	+	2	143	228	84 27

ORF3 Marked set (0)

SmartBLAST BLAST SmartBLAST best hit titles... BLAST

BLAST Database: UniProtKB/Swiss-Prot (swissprot)

NCBI result after submitting sequence-1

We see that it gives us total eight ORFs with four direct strands and four reverse strands. It also shows the corresponding protein sequence for the ORF we select.

After submission of the below protein sequence which is the ORF3 in this case, we get the result from INTERPRO analysis:

MIHLEKMEWNFKMENIKDKKNDILEKLVHINRITKVVKGGRRFGFSALVVVGNQAGRIGIAHAKAKQVPDAIKKANEMAR
 RKLHIPLRERRTIHHDVKAKDGS GKIVLRAASKGTGIIAGGPVRAVCEVLV

Examples: IPR020405, kinase, P51587, PF02932, GO:0007165

Protein sequence analysis & classification

[Home](#) | [Search](#) | [Release notes](#) | [Download](#) | [About InterPro](#) | [Help](#) | [Contact](#)

Overview

Similar proteins

Structures

Filter view on

Entry type

- Homologous superfamily
- Family
- Domains
- Repeats
- Site

Status

- Unintegrated

Per-residue features

- Residue annotation

Colour by

- domain relationship
- source database

Submitted

Length 133 amino acids

Protein family membership

Ribosomal protein S5 (IPR000851)

Homologous superfamilies

Domains and repeats

Detailed signature matches

IPR014721	Ribosomal protein S5 domain 2-type fold, subgroup	G3DSA:3.30.23...
IPR020568	Ribosomal protein S5 domain 2-type fold	SSF54211 (Ribosomal...)
IPR000851	Ribosomal protein S5	PTHR13718 (RIBOSOMA...)
IPR013810	Ribosomal protein S5, N-terminal	PF00333 (Ribosomal_S5) PS50881 (S5_DSRBD)
IPR005324	Ribosomal protein S5, C-terminal	PF03719 (Ribosomal...)
IPR018192	Ribosomal protein S5, N-terminal, conserved site	PS00585 (RIBOSOMAL_S5)
no IPR	Unintegrated signatures	G3DSA:3.30.16... PTHR13718:SF61 (28S...) SSF54768 (dsRNA-bin...)

GO term prediction

Biological Process

[GO:0006412](#) translation

Molecular Function

[GO:0003723](#) RNA binding

[GO:0003735](#) structural constituent of ribosome

Cellular Component

[GO:0005840](#) ribosome

INTERPRO result for ORF3 of sequence-1

Here we see that protein family membership is **Ribosomal S5 with homologous superfamily**. By GO term prediction we see that it has the cellular component **ribosome(GO:0005840)** with biological process as **translation(GO:0006412)** and molecular functions are **RNA binding(GO:0003723)**, **structural constituent of ribosome(GO:0003735)**[6].

After completing the **INTERPRO** analysis we export the data into the **TSV format** which is as follows:

Submitted	5a182076cf5d5b179b0f3700570ce5cf133	SUPERFAMILY	SSF54768	21	89	3.35E-20	T	05-04-2018
Submitted	5a182076cf5d5b179b0f3700570ce5cf133	Pfam	PF03719	Ribosomal protein S5, C-terminal domain	101	131	2.2E-8	
	T	05-04-2018	IPR005324	Ribosomal protein S5, C-terminal	GO:0003735 GO:0005840 GO:0006412			
Submitted	5a182076cf5d5b179b0f3700570ce5cf133	PANTHER	PTHR13718	18	132	9.0E-39	T	05-04-2018
	IPR000851	Ribosomal protein S5	GO:0003735 GO:0005840 GO:0006412					
Submitted	5a182076cf5d5b179b0f3700570ce5cf133	ProSitePatterns	PS00585	Ribosomal protein S5 signature.	41	73	-	
	T	05-04-2018	IPR018192	Ribosomal protein S5, N-terminal, conserved site				
Submitted	5a182076cf5d5b179b0f3700570ce5cf133	Pfam	PF00333	Ribosomal protein S5, N-terminal domain	26	88	3.7E-	
25	T	05-04-2018	IPR013810	Ribosomal protein S5, N-terminal	GO:0003723 GO:0003735 GO:0005840 GO:0006412			
Submitted	5a182076cf5d5b179b0f3700570ce5cf133	ProSiteProfiles	PS50881	S5 double stranded RNA-binding domain profile.	24	87		
	22.311	T	05-04-2018	IPR013810	Ribosomal protein S5, N-terminal	GO:0003723 GO:0003735 GO:0005840 GO:0006412		
Submitted	5a182076cf5d5b179b0f3700570ce5cf133	SUPERFAMILY	SSF54211	93	131	2.07E-8	T	05-04-2018
	IPR020568	Ribosomal protein S5 domain 2-type fold						
Submitted	5a182076cf5d5b179b0f3700570ce5cf133	PANTHER	PTHR13718:SF61	18	132	9.0E-39	T	05-04-2018
Submitted	5a182076cf5d5b179b0f3700570ce5cf133	Gene3D	G3DSA:3.30.160.20	13	86	3.1E-26	T	05-04-2018
Submitted	5a182076cf5d5b179b0f3700570ce5cf133	Gene3D	G3DSA:3.30.230.10	87	132	8.0E-13	T	05-04-2018
	IPR014721	Ribosomal protein S5 domain 2-type fold, subgroup						

From the above table we see that the lowest E-value corresponds to the Ribosomal S-5.

The BLAST result for SWISSPROT database is below:

NIH U.S. National Library of Medicine NCBI National Center for Biotechnology Information Sign in to NCBI

BLAST® » blastp suite » RID-FSP3KDW014 Home Recent Results Saved Strategies Help

BLAST Results

Edit and Resubmit Save Search Strategies Formatting options Download YouTube How to read this page Blast report description

Job title: Icl|ORF3:548:949 unnamed protein product (133 letters)

RID FSP3KDW014 (Expires on 05-18 00:12 am)

Query ID Icl|Query_301682 Database Name swissprot
 Description Icl|ORF3:548:949 unnamed protein product Description Non-redundant UniProtKB/SwissProt sequences
 Molecule type amino acid Program BLASTP 2.8.0+ Citation
 Query Length 133

Other reports: Search Summary Taxonomy reports Distance tree of results Multiple alignment MSA viewer

New Analyze your query with SmartBLAST

Graphic Summary

Show Conserved Domains

Putative conserved domains have been detected, click on the image below for detailed results.

Query seq. 1 11HLEKMEINFRHENIKDKKNDLLEKLVHNRITKYVHGGRPFQFSALVYVGNDRIGTAFHKAKQIPDRIKAVNEHAKKLIH1PLREARTIHQDWAHKDSGKI VLRAHAKSGTGIJAGGPIRAHVCVELV 133
 Superfamilies Ribosomal_S5 superfamily

Distribution of the top 100 Blast Hits on 100 subject sequences

Color key for alignment scores

■ <40 ■ 40-50 ■ 50-80 ■ 80-200 ■ >=200

1 20 40 60 80 100 120

Questions/comments

Descriptions

Sequences producing significant alignments:

Select: All None Selected: 100

Alignments Download GenPept Graphics Distance tree of results Multiple alignment

Description	Max score	Total score	Query cover	E value	Ident	Accession
RecName: Full=30S ribosomal protein S5	205	205	88%	4e-68	86%	Q4FLN5.1
RecName: Full=30S ribosomal protein S5	159	159	87%	1e-49	61%	B8LE6.1
RecName: Full=30S ribosomal protein S5	159	159	87%	2e-49	62%	A7HW58.1
RecName: Full=30S ribosomal protein S5	156	156	87%	1e-48	58%	Q8FZD9.1
RecName: Full=30S ribosomal protein S5	156	156	87%	2e-48	62%	A8IAP6.1
RecName: Full=30S ribosomal protein S5	156	156	84%	2e-48	64%	B2IK79.1
RecName: Full=30S ribosomal protein S5	155	155	90%	2e-48	56%	Q6Q2Y2.1
RecName: Full=30S ribosomal protein S5 (Mesorhizobium jaecanicum MAFF 303099)	155	155	84%	3e-48	63%	Q8N40.2
RecName: Full=30S ribosomal protein S5	155	155	87%	3e-48	57%	A9IW04.1
RecName: Full=30S ribosomal protein S5	155	155	84%	4e-48	65%	A7IPI3.1
RecName: Full=30S ribosomal protein S5	155	155	99%	6e-48	55%	A1USR1.1
RecName: Full=30S ribosomal protein S5	154	154	87%	1e-47	60%	A6X0D5.1
RecName: Full=30S ribosomal protein S5	154	154	83%	1e-47	65%	B1LWQ9.1
RecName: Full=30S ribosomal protein S5	154	154	82%	1e-47	65%	B1Z776.1
RecName: Full=30S ribosomal protein S5	154	154	83%	1e-47	65%	A9W4S3.1
RecName: Full=30S ribosomal protein S5	154	154	87%	2e-47	59%	Q21IG5.1
RecName: Full=30S ribosomal protein S5	153	153	87%	2e-47	59%	Q07KNS.1
RecName: Full=30S ribosomal protein S5	153	153	87%	3e-47	60%	Q8UE35.1
RecName: Full=30S ribosomal protein S5	153	153	83%	3e-47	60%	Q8UE35.1
RecName: Full=30S ribosomal protein S5	153	153	87%	3e-47	60%	Q8UE35.1

Questions/comments

Result of BLAST for SWISSPROT database

This blast result for SWISSPROT database shows that it has the Ribosomal S-5 superfamily. And the distribution for the top 100 BLAST hits on 100 subject sequences are also shown. In the description we see that the topmost hit has the lowest E-value among them.

The BLAST result for NR database is below:

NIH U.S. National Library of Medicine NCBI National Center for Biotechnology Information Sign in to NCBI

BLAST® » blastp suite » RID-FSPCY5F1014 Home Recent Results Saved Strategies Help

BLAST Results

Edit and Resubmit Save Search Strategies Formatting options Download You Tube How to read this page Blast report description

Job title: Protein Sequence (133 letters)

RID: FSPCY5F1014 (Expires on 05-18 00:17 am)

Query ID: Icd|Query_340505 Database Name: nr
 Description: None Description: All non-redundant GenBank CDS translations+PDB+SwissProt+PIR+PRF excluding environmental samples from WGS projects
 Molecule type: amino acid Program: BLASTP 2.8.0+ » Citation
 Query Length: 133

Other reports: Search Summary Taxonomy reports Distance tree of results Related Structures Multiple alignment MSA viewer

Graphic Summary

Show Conserved Domains

Putative conserved domains have been detected, click on the image below for detailed results.

Query seq. M I H L E X M E I N F M E N T I K D K N D D L E K L V H I N R I T K Y V N G R R P F S A L V Y V G N Q N G R I G I A H A W K Q P D H I A W N E Y W R K L I H I P L R E R T I H D W H W D S G K I V L R A A S M G T G I I G G G P M R W C E V L V

Superfamilies: Ribosomal_S5 superfamily

Distribution of the top 100 Blast Hits on 100 subject sequences

Color key for alignment scores

- <40
- 40-50
- 50-80
- 80-200
- >=200

Query: 1 20 40 60 80 100 120

Descriptions

Sequences producing significant alignments:

Select: All None Selected: 100

Description	Max score	Total score	Query cover	E value	Ident	Accession
30S ribosomal protein S5 [Candidatus Pelagibacter ubique]	229	229	90%	4e-75	96%	WP_075503615.1
30S ribosomal protein S5 [Candidatus Pelagibacter ubique]	229	229	90%	4e-75	96%	WP_075535964.1
30S ribosomal protein S5 [Candidatus Pelagibacter sp. HIMB1321]	229	229	90%	8e-75	97%	WP_085114363.1
30S ribosomal protein S5 [Candidatus Pelagibacter ubique]	229	229	90%	9e-75	95%	WP_023853717.1
30S ribosomal protein S5 [Candidatus Pelagibacter ubique]	228	228	90%	1e-74	95%	WP_023648442.1
30S ribosomal protein S5 [Candidatus Pelagibacter ubique]	228	228	90%	2e-74	94%	WP_075485245.1
30S ribosomal protein S5 [Candidatus Pelagibacter sp. TMED239]	227	227	90%	3e-74	94%	OUX08051.1
30S ribosomal protein S5 [alpha proteobacterium HIMB5]	227	227	90%	4e-74	96%	WP_014953343.1
30S ribosomal protein S5 [Candidatus Pelagibacter ubique]	226	226	90%	8e-74	95%	WP_075506793.1
30S ribosomal protein S5 [Candidatus Pelagibacter ubique]	226	226	90%	8e-74	95%	WP_075501168.1
30S ribosomal protein S5 [Pelagibacteraceae bacterium TMED267]	225	225	90%	2e-73	93%	OUX65180.1
30S ribosomal protein S5 [Pelagibacteraceae bacterium TMED267]	224	224	94%	4e-73	91%	OUX63645.1
30S ribosomal protein S5 [Candidatus Pelagibacter sp. RS39]	224	224	90%	4e-73	93%	WP_085148133.1
30S ribosomal protein S5 [Candidatus Pelagibacter ubique]	224	224	90%	5e-73	93%	WP_028036902.1
MULTISPECIES: 30S ribosomal protein S5 [Candidatus Pelagibacter]	224	224	90%	8e-73	93%	WP_008545225.1
30S ribosomal protein S5 [Candidatus Pelagibacter ubique]	224	224	90%	8e-73	96%	WP_075522335.1
30S ribosomal protein S5 [Pelagibacteraceae bacterium TMED267]	222	222	90%	5e-72	92%	OUX63806.1
30S ribosomal protein S5 [Pelagibacteriales bacterium MED-041]	221	221	87%	8e-72	96%	PDH18688.1

Result of BLAST for NR database

This blast result for NR database shows that it has the **Ribosomal S-5 superfamily**. And the distribution for the top 100 BLAST hits on 100 subject sequences are also shown. In the description we see that the topmost hit has the lowest E-value among them.

The result for conserved protein domains is here:

Conserved domains on [lcl|ORF3:548:949 unnamed protein product] View **Standard Results**

Local query sequence

Graphical summary Zoom to residue level [show extra options >](#)

Query seq. MIIHLEKIDINIFKPIENIKKGGDDILEKLVHINRITKYVHGRRFGFSALVYVGNDRIGRIGIAHAKKQNPDAIKANENARRKLIHIFLRRERTIHHDDVRAKDGSGKIVLRAASKGTGIAGGPPRAYCELV

Specific hits

- Ribosomal_S5

Non-specific hits

- rpsE
- rpsE_bact

Superfamilies

- Ribosomal_S5 superfamily

List of domain hits

+	Name	Accession	Description	Interval	E-value
[H]	rpsE	PRK00550	30S ribosomal protein S5; Validated	13-132	5.13e-59
[H]	RpsE	COG0098	Ribosomal protein S5 [Translation, ribosomal structure and biogenesis];	16-132	1.24e-49
[H]	rpsE_bact	TIGR01021	ribosomal protein S5, bacterial/organelle type; This model finds chloroplast ribosomal protein ...	26-132	2.20e-42
[H]	Ribosomal_S5	pfam00333	Ribosomal protein S5, N-terminal domain;	26-88	5.15e-24

Blast search parameters

Data Source: Live blast search RID = FSP3JNP1014
 User Options: Database: CDSEARCH/odd v3.16 Low complexity filter: no Composition Based Adjustment: yes E-value threshold: 0.01 Maximum number of hits: 500

Conserved Protein Domain

This clearly shows the superfamily as Ribosomal S-5.

The SMARTBLAST result is below:

SMARTBLAST >> Formatting Results - FSPUT6NF011 Home Help

Summary [Report description](#)

Query: unnamed protein product Query length: 133 aa

DOMAINS: Ribosomal protein S5, N-terminal domain and 1 other(s)

Your query: unnamed protein product

30S ribosomal protein S5 RpsE
 30S ribosomal protein S5
 30S ribosomal protein S5
 30S ribosomal protein S5
 30S ribosomal protein S5
 30S ribosomal protein S5

Best hits

Description	Max score	Total score	Query cover	E value	Ident	Accession
30S ribosomal protein S5 RpsE [Shewanella oneidensis MR-1]	116	116	89%	9e-33	48%	NP_715888.1
30S ribosomal protein S5 [Pseudomonas aeruginosa PAQ1]	115	115	89%	1e-32	49%	NP_252936.1
30S ribosomal protein S5 [Clostridioides difficile 630]	111	111	85%	6e-31	53%	YP_001086554.1
30S ribosomal protein S5 [Neisseria meningitidis MC58]	110	110	86%	1e-30	47%	NP_273217.1
30S ribosomal protein S5 [Mycobacterium tuberculosis H37Rv]	111	111	85%	2e-30	50%	NP_215235.1

Additional BLAST hits

Description	Max score	Total score	Query cover	E value	Ident	Accession
30S ribosomal protein S5 [Candidatus Pelagibacter ubique]	229	229	90%	7e-81	96%	WP_075503615.1
30S ribosomal protein S5 [Candidatus Pelagibacter ubique]	229	229	90%	7e-81	96%	WP_075535964.1
30S ribosomal protein S5 [Candidatus Pelagibacter sp. HIMB1321]	229	229	90%	1e-80	97%	WP_085114363.1
30S ribosomal protein S5 [Candidatus Pelagibacter ubique]	229	229	90%	1e-80	95%	WP_023853717.1
30S ribosomal protein S5 [Candidatus Pelagibacter ubique]	228	228	90%	2e-80	95%	WP_023648442.1

SMARTBLAST result

From this SMARTBLAST result and the above results we can definitely see that the superfamily is Ribosomal S-5. And the SMARTBLAST result we can see that all the top five BLAST hits are on Ribosomal S-5. We are clear about the superfamily now.

The input for the **Definition List** tool was the sequences producing significant alignments, achieved from the BLAST result, which are:

Sequences producing significant alignments:	(Bits)	Value
WP_075503615.1	30S ribosomal protein S5 [Candidatus Pelagibac...	229 5e-75
WP_075535964.1	30S ribosomal protein S5 [Candidatus Pelagibac...	229 6e-75
WP_085114363.1	30S ribosomal protein S5 [Candidatus Pelagibac...	229 9e-75
WP_023853717.1	30S ribosomal protein S5 [Candidatus Pelagibac...	228 1e-74
WP_023648442.1	30S ribosomal protein S5 [Candidatus Pelagibac...	228 1e-74
WP_075485245.1	30S ribosomal protein S5 [Candidatus Pelagibac...	228 2e-74
OUX08051.1	30S ribosomal protein S5 [Candidatus Pelagibacter ...	227 5e-74
WP_014953343.1	30S ribosomal protein S5 [alpha proteobacteriu...	227 5e-74
WP_075506793.1	30S ribosomal protein S5 [Candidatus Pelagibac...	226 1e-73
WP_075501168.1	30S ribosomal protein S5 [Candidatus Pelagibac...	226 1e-73
OUX65160.1	30S ribosomal protein S5 [Pelagibacteraceae bacter...	225 3e-73
WP_085148133.1	30S ribosomal protein S5 [Candidatus Pelagibac...	224 5e-73
OUX63645.1	30S ribosomal protein S5 [Pelagibacteraceae bacter...	224 7e-73
WP_028036902.1	30S ribosomal protein S5 [Candidatus Pelagibac...	224 8e-73
WP_008545225.1	MULTISPECIES: 30S ribosomal protein S5 [Candid...	223 1e-72
WP_075522335.1	30S ribosomal protein S5 [Candidatus Pelagibac...	223 1e-72
OUX63806.1	30S ribosomal protein S5 [Pelagibacteraceae bacter...	222 5e-72
PDH18668.1	30S ribosomal protein S5 [Pelagibacterales bacteri...	221 1e-71
KRO55093.1	30S ribosomal protein S5 [Pelagibacteraceae bacter...	220 2e-71
Ouw94408.1	30S ribosomal protein S5 [Pelagibacteraceae bacter...	216 1e-69
Ouw94547.1	30S ribosomal protein S5 [Pelagibacteraceae bacter...	213 1e-68
Ouv23961.1	30S ribosomal protein S5 [Candidatus Pelagibacter ...	208 1e-66
WP_006996828.1	30S ribosomal protein S5 [Candidatus Pelagibac...	205 2e-65
KRP05306.1	30S ribosomal protein S5 [Pelagibacteraceae bacter...	201 5e-64
OUX36823.1	30S ribosomal protein S5 [Candidatus Pelagibacter ...	200 1e-63
WP_075534950.1	30S ribosomal protein S5 [Candidatus Pelagibac...	197 3e-62
OCW83531.1	30S ribosomal protein S5 [Pelagibacteraceae bacter...	195 1e-61
WP_075483566.1	30S ribosomal protein S5 [Candidatus Pelagibac...	194 5e-61
Ouw52867.1	30S ribosomal protein S5 [Gammaproteobacteria bact...	193 6e-61
WP_075521549.1	MULTISPECIES: 30S ribosomal protein S5 [Candid...	193 7e-61
OCW78361.1	30S ribosomal protein S5 [Pelagibacteraceae bacter...	191 5e-60
Oux46179.1	30S ribosomal protein S5 [Candidatus Pelagibacter ...	191 9e-60
Ouv44679.1	30S ribosomal protein S5 [Candidatus Pelagibacter ...	190 1e-59
OCW82717.1	30S ribosomal protein S5 [Pelagibacteraceae bacter...	189 6e-59
PHQ86133.1	30S ribosomal protein S5 [Candidatus Pelagibacter ...	185 2e-57
Oux26774.1	30S ribosomal protein S5 [Candidatus Pelagibacter ...	172 1e-52
Ouw24766.1	30S ribosomal protein S5 [Pelagibacteraceae bacter...	172 2e-52
Ouw24121.1	30S ribosomal protein S5 [Pelagibacteraceae bacter...	172 3e-52
Ouv61189.1	30S ribosomal protein S5 [Candidatus Pelagibacter ...	172 3e-52

OUW57874.1 30S ribosomal protein S5 [Candidatus Pelagibacter ... 169 3e-51
 WP_075533907.1 30S ribosomal protein S5 [Candidatus Pelagibac... 169 4e-51
 OUX45222.1 30S ribosomal protein S5 [Candidatus Pelagibacter ... 167 1e-50
 OUX53284.1 30S ribosomal protein S5 [Candidatus Pelagibacter ... 167 2e-50
 OUV51621.1 30S ribosomal protein S5 [Pelagibacteraceae bacter... 164 2e-49
 OUW11979.1 30S ribosomal protein S5 [Candidatus Pelagibacter ... 164 2e-49
 OUW21574.1 30S ribosomal protein S5 [Candidatus Pelagibacter ... 164 3e-49
 OUU67093.1 30S ribosomal protein S5 [Candidatus Pelagibacter ... 163 5e-49
 WP_100693123.1 30S ribosomal protein S5 [Bartonella sp. 1-1C] 163 1e-48
 WP_035006058.1 MULTISPECIES: 30S ribosomal protein S5 [Barton... 163 1e-48
 OUW49217.1 30S ribosomal protein S5 [Candidatus Pelagibacter ... 162 1e-48
 WP_069297842.1 30S ribosomal protein S5 [Pelagibacteraceae ba... 162 2e-48
 WP_069304988.1 30S ribosomal protein S5 [Pelagibacteraceae ba... 162 2e-48
 WP_078680090.1 30S ribosomal protein S5 [Bartonella sp. AR 15-3] 162 2e-48
 WP_099340244.1 30S ribosomal protein S5 [Candidatus Fonsibact... 161 4e-48
 WP_013695847.1 30S ribosomal protein S5 [Candidatus Pelagibac... 161 5e-48
 WP_015398153.1 30S ribosomal protein S5 [Bartonella australis] 161 6e-48
 WP_063292674.1 MULTISPECIES: 30S ribosomal protein S5 [Pseudo... 161 7e-48
 WP_069281395.1 30S ribosomal protein S5 [Pelagibacteraceae ba... 160 8e-48
 OUV51017.1 30S ribosomal protein S5 [Pelagibacteraceae bacter... 160 9e-48
 WP_013652500.1 30S ribosomal protein S5 [Polymorphum gilvum] 161 9e-48
 WP_094071087.1 30S ribosomal protein S5 [Labrenzia sp. VG12] 161 1e-47
 AQS41263.1 30S ribosomal protein S5 [Candidatus Tokpelaia hoe... 160 1e-47
 WP_103222481.1 30S ribosomal protein S5 [Labrenzia marina] 161 1e-47
 WP_038365418.1 30S ribosomal protein S5 [Bosea sp. UNC402CLCol] 160 1e-47
 KFX72073.1 30S ribosomal protein S5 [alpha proteobacterium QL1] 160 1e-47
 WP_055678255.1 30S ribosomal protein S5 [Labrenzia alba] 160 1e-47
 WP_006936434.1 MULTISPECIES: 30S ribosomal protein S5 [Labren... 160 1e-47
 WP_035898399.1 MULTISPECIES: 30S ribosomal protein S5 [Labren... 160 1e-47
 WP_068416144.1 30S ribosomal protein S5 [Labrenzia sp. OB1] 160 1e-47
 WP_009360122.1 30S ribosomal protein S5 [alpha proteobacteriu... 160 1e-47
 WP_078719063.1 MULTISPECIES: 30S ribosomal protein S5 [Barton... 160 1e-47
 PCJ93161.1 30S ribosomal protein S5 [Rhizobiales bacterium] 160 1e-47
 WP_069440453.1 30S ribosomal protein S5 [Methyloceanibacter s... 160 2e-47
 WP_009469220.1 30S ribosomal protein S5 [Roseibium sp. TrichS... 160 2e-47
 WP_008039952.1 30S ribosomal protein S5 [Bartonella tamiae] 160 2e-47
 ODT15626.1 30S ribosomal protein S5 [Kaistia sp. SCN 65-12] 159 2e-47
 WP_020185775.1 MULTISPECIES: 30S ribosomal protein S5 [Methyl... 160 3e-47
 WP_019965855.1 MULTISPECIES: 30S ribosomal protein S5 [Pannon... 160 3e-47
 WP_068314049.1 30S ribosomal protein S5 [Pseudovibrio hongkon... 160 3e-47
 WP_004855694.1 30S ribosomal protein S5 [Bartonella doshiae] 159 3e-47
 WP_014130819.1 MULTISPECIES: 30S ribosomal protein S5 [Pelagi... 159 3e-47
 WP_008548041.1 MULTISPECIES: 30S ribosomal protein S5 [Pseudo... 159 3e-47
 WP_093187996.1 30S ribosomal protein S5 [Pseudovibrio sp. Tun... 159 4e-47
 WP_068082347.1 30S ribosomal protein S5 [Pseudovibrio stylochi] 159 4e-47

WP_006589515.1 30S ribosomal protein S5 [Bartonella birtlesii] 159 4e-47
 WP_103717382.1 30S ribosomal protein S5 [Bosea sp. 1131] 159 4e-47
 WP_051089729.1 30S ribosomal protein S5 [Amorphus coralli] 159 4e-47
 WP_069692139.1 MULTISPECIES: 30S ribosomal protein S5 [Bosea] 159 4e-47
 WP_026304633.1 30S ribosomal protein S5 [Kaistia granuli] 159 5e-47
 WP_073011863.1 30S ribosomal protein S5 [Labrenzia suaedae] 159 5e-47
 WP_097152734.1 30S ribosomal protein S5 [Cohaesibacter gelati... 159 5e-47
 WP_102844538.1 30S ribosomal protein S5 [Methylocella silvest... 159 5e-47
 WP_029075456.1 30S ribosomal protein S5 [Kaistia adipata] 159 5e-47
 WP_055671482.1 30S ribosomal protein S5 [Labrenzia alexandrii] 159 5e-47
 WP_091680665.1 30S ribosomal protein S5 [Methylocapsa palsarum] 159 6e-47
 WP_012589605.1 30S ribosomal protein S5 [Methylocella silvest... 159 6e-47
 WP_069624394.1 30S ribosomal protein S5 [Methyloceanibacter m... 159 6e-47
 WP_029057565.1 MULTISPECIES: 30S ribosomal protein S5 [Stappia] 159 6e-47
 WP_006925145.1 30S ribosomal protein S5 [Bartonella washoensis] 159 6e-47
 WP_019221645.1 30S ribosomal protein S5 [Bartonella senegalen... 159 6e-47

From the Definition list tool we get the following result:

BLAST hit list Definitions

Definitions	nb lines	min E-value	max E-value
30S ribosomal protein	100	5e-75	6e-47

[Results end]

The input for **TaxReports** tool list is:

Sequences producing significant alignments:	(Bits)	Value
Q4FLN5.1 RecName: Full=30S ribosomal protein S5	205	6e-68
B8ELE6.1 RecName: Full=30S ribosomal protein S5	159	2e-49
A7HWS8.1 RecName: Full=30S ribosomal protein S5	158	3e-49
Q6FZD9.1 RecName: Full=30S ribosomal protein S5	156	1e-48
A8IAP6.1 RecName: Full=30S ribosomal protein S5	156	2e-48
B2IK79.1 RecName: Full=30S ribosomal protein S5	156	3e-48
Q6G2Y2.1 RecName: Full=30S ribosomal protein S5	155	3e-48
Q98N40.2 RecName: Full=30S ribosomal protein S5 [Mesorhizobiu...	155	4e-48
A9IW04.1 RecName: Full=30S ribosomal protein S5	155	4e-48
A7IPQ3.1 RecName: Full=30S ribosomal protein S5	155	5e-48
A6X0D5.1 RecName: Full=30S ribosomal protein S5	154	1e-47
B1LWQ9.1 RecName: Full=30S ribosomal protein S5	154	1e-47
A1USR1.1 RecName: Full=30S ribosomal protein S5	154	2e-47
A9W4S3.1 RecName: Full=30S ribosomal protein S5	154	2e-47
B1Z776.1 RecName: Full=30S ribosomal protein S5	154	2e-47
Q07KN5.1 RecName: Full=30S ribosomal protein S5	153	3e-47
Q211G5.1 RecName: Full=30S ribosomal protein S5	153	3e-47
Q11HR9.1 RecName: Full=30S ribosomal protein S5	153	3e-47
Q8UE35.1 RecName: Full=30S ribosomal protein S5	153	4e-47
B9JVQ4.1 RecName: Full=30S ribosomal protein S5	152	6e-47
Q2K9J9.1 RecName: Full=30S ribosomal protein S5	152	6e-47
Q1MIC4.1 RecName: Full=30S ribosomal protein S5	152	7e-47
C3MAZ7.1 RecName: Full=30S ribosomal protein S5	152	9e-47
Q92QF3.1 RecName: Full=30S ribosomal protein S5	152	1e-46
B8ISA2.1 RecName: Full=30S ribosomal protein S5	151	2e-46
B0UHV2.1 RecName: Full=30S ribosomal protein S5	151	2e-46
P66571.1 RecName: Full=30S ribosomal protein S5	150	4e-46
A5VQY9.1 RecName: Full=30S ribosomal protein S5	150	6e-46
Q3SSU9.1 RecName: Full=30S ribosomal protein S5	150	8e-46
B4R8N4.1 RecName: Full=30S ribosomal protein S5	149	1e-45
B6JEY2.1 RecName: Full=30S ribosomal protein S5	149	1e-45
Q9A8T6.1 RecName: Full=30S ribosomal protein S5 [Caulobacter ...	149	2e-45
Q1QN13.1 RecName: Full=30S ribosomal protein S5	148	5e-45
A4YSK9.2 RecName: Full=30S ribosomal protein S5	147	6e-45
B0T2D9.1 RecName: Full=30S ribosomal protein S5	148	6e-45
Q0BUN3.1 RecName: Full=30S ribosomal protein S5	147	6e-45
Q6N4V0.3 RecName: Full=30S ribosomal protein S5; AltName: Ful...	146	1e-44
Q89JA1.1 RecName: Full=30S ribosomal protein S5	146	2e-44
Q28UT7.1 RecName: Full=30S ribosomal protein S5	146	2e-44
Q0BYD1.1 RecName: Full=30S ribosomal protein S5	146	2e-44
A8LM74.1 RecName: Full=30S ribosomal protein S5	146	2e-44

Q134U6.1	RecName: Full=30S ribosomal protein S5	146	3e-44
Q2IXP3.1	RecName: Full=30S ribosomal protein S5	145	3e-44
Q0ANR7.1	RecName: Full=30S ribosomal protein S5	145	4e-44
Q1GK12.1	RecName: Full=30S ribosomal protein S5	144	8e-44
Q5LW41.1	RecName: Full=30S ribosomal protein S5	144	1e-43
Q5FU00.1	RecName: Full=30S ribosomal protein S5	144	1e-43
A1B044.1	RecName: Full=30S ribosomal protein S5	144	2e-43
Q16AC7.1	RecName: Full=30S ribosomal protein S5	143	3e-43
Q2RQX7.1	RecName: Full=30S ribosomal protein S5	142	8e-43
Q5NQ48.1	RecName: Full=30S ribosomal protein S5	144	8e-43
A5FZU8.1	RecName: Full=30S ribosomal protein S5	142	8e-43
Q2N9C8.1	RecName: Full=30S ribosomal protein S5	143	3e-42
A4WVJ1.1	RecName: Full=30S ribosomal protein S5	140	3e-42
Q3J5Q5.1	RecName: Full=30S ribosomal protein S5	140	4e-42
Q2W2K8.1	RecName: Full=30S ribosomal protein S5	141	4e-42
A9H3L2.1	RecName: Full=30S ribosomal protein S5	138	2e-41
A8F2D0.2	RecName: Full=30S ribosomal protein S5	133	2e-39
A8GPD3.1	RecName: Full=30S ribosomal protein S5	132	3e-39
C4K2G2.1	RecName: Full=30S ribosomal protein S5	132	5e-39
Q92GY3.1	RecName: Full=30S ribosomal protein S5	131	1e-38
C3PP90.1	RecName: Full=30S ribosomal protein S5	131	1e-38
Q4UMR2.1	RecName: Full=30S ribosomal protein S5	131	2e-38
A8GT52.1	RecName: Full=30S ribosomal protein S5	130	2e-38
Q1GPB6.1	RecName: Full=30S ribosomal protein S5	132	3e-38
Q68W95.1	RecName: Full=30S ribosomal protein S5	130	3e-38
Q9ZCS2.1	RecName: Full=30S ribosomal protein S5	130	3e-38
A5V5Y5.1	RecName: Full=30S ribosomal protein S5	132	3e-38
A8EZJ9.1	RecName: Full=30S ribosomal protein S5	128	1e-37
B7J484.1	RecName: Full=30S ribosomal protein S5	127	3e-37
C6C1A2.1	RecName: Full=30S ribosomal protein S5	127	3e-37
Q2G8W3.1	RecName: Full=30S ribosomal protein S5	125	8e-36
Q8K966.1	RecName: Full=30S ribosomal protein S5	123	1e-35
A5CCJ5.1	RecName: Full=30S ribosomal protein S5	123	1e-35
Q0AUJ7.1	RecName: Full=30S ribosomal protein S5	123	1e-35
Q3A6N0.1	RecName: Full=30S ribosomal protein S5	122	2e-35
B2A4F6.1	RecName: Full=30S ribosomal protein S5	122	4e-35
P57574.1	RecName: Full=30S ribosomal protein S5	121	7e-35
Q8XV29.1	RecName: Full=30S ribosomal protein S5	121	8e-35
B2UEK2.1	RecName: Full=30S ribosomal protein S5	120	2e-34
Q5GSW1.1	RecName: Full=30S ribosomal protein S5	120	3e-34
Q46WG1.1	RecName: Full=30S ribosomal protein S5	119	3e-34
A5IYW9.1	RecName: Full=30S ribosomal protein S5	121	4e-34
Q0K636.1	RecName: Full=30S ribosomal protein S5	119	4e-34
Q73HA3.1	RecName: Full=30S ribosomal protein S5	119	4e-34
Q39KF0.1	RecName: Full=30S ribosomal protein S5	119	5e-34

Q0BJ29.1	RecName: Full=30S ribosomal protein S5	119	5e-34
A9BRW8.1	RecName: Full=30S ribosomal protein S5	119	5e-34
Q13TI7.1	RecName: Full=30S ribosomal protein S5	119	6e-34
A4J128.1	RecName: Full=30S ribosomal protein S5	119	7e-34
B2JI48.1	RecName: Full=30S ribosomal protein S5	119	7e-34
Q2SU44.1	RecName: Full=30S ribosomal protein S5	119	8e-34
Q62GM2.1	RecName: Full=30S ribosomal protein S5	119	8e-34
Q3J8T1.1	RecName: Full=30S ribosomal protein S5	119	8e-34
A2SLE0.1	RecName: Full=30S ribosomal protein S5	119	9e-34
A1VJ32.1	RecName: Full=30S ribosomal protein S5	118	1e-33
A1S235.1	RecName: Full=30S ribosomal protein S5	118	1e-33
Q1LI54.1	RecName: Full=30S ribosomal protein S5	118	1e-33
B2T734.1	RecName: Full=30S ribosomal protein S5	118	1e-33
Q9Z9J7.1	RecName: Full=30S ribosomal protein S5	118	1e-33

The above result shows that the lowest E-value corresponds to the 30S Ribosomal Protein S-5.

The result we get from the TaxReports tool is:

```
.LUCA
. Bacteria
. .Proteobacteria
. . .Alphaproteobacteria
. . .Rickettsiales
. . .Pelagibacterales
. . . .Candidatus Pelagibacter
. . . .Candidatus Pelagibacter ubique..... 205 6e-68 1 hit Bacteria:Proteobacteria:Alphaproteobacteria:Rickettsiales: RecName: Full=30S ribosomal protein S5
. . . spotted fever group..... 133 2e-39 3 hits Bacteria:Proteobacteria:Alphaproteobacteria:Rickettsiales: RecName: Full=30S ribosomal protein S5
. . . Rickettsia akari..... 132 3e-39 1 hit Bacteria:Proteobacteria:Alphaproteobacteria:Rickettsiales: RecName: Full=30S ribosomal protein S5
. . . Rickettsia peacockii..... 132 5e-39 1 hit Bacteria:Proteobacteria:Alphaproteobacteria:Rickettsiales: RecName: Full=30S ribosomal protein S5
. . . Rickettsia conorii..... 131 1e-38 1 hit Bacteria:Proteobacteria:Alphaproteobacteria:Rickettsiales: RecName: Full=30S ribosomal protein S5
. . . Rickettsia africae ESF-5..... 131 1e-38 2 hits Bacteria:Proteobacteria:Alphaproteobacteria:Rickettsiales: RecName: Full=30S ribosomal protein S5
. . . Rickettsia felis URRWXCa2..... 131 2e-38 2 hits Bacteria:Proteobacteria:Alphaproteobacteria:Rickettsiales: RecName: Full=30S ribosomal protein S5
. . . Rickettsia rickettsii..... 130 2e-38 3 hits Bacteria:Proteobacteria:Alphaproteobacteria:Rickettsiales: RecName: Full=30S ribosomal protein S5
. . . Rickettsia typhi..... 130 3e-38 1 hit Bacteria:Proteobacteria:Alphaproteobacteria:Rickettsiales: RecName: Full=30S ribosomal protein S5
. . . Rickettsia prowazekii..... 130 3e-38 2 hits Bacteria:Proteobacteria:Alphaproteobacteria:Rickettsiales: RecName: Full=30S ribosomal protein S5
. . . Rickettsia canadensis..... 128 1e-37 1 hit Bacteria:Proteobacteria:Alphaproteobacteria:Rickettsiales: RecName: Full=30S ribosomal protein S5
. . . Orientia tsutsugamushi..... 123 1e-35 1 hit Bacteria:Proteobacteria:Alphaproteobacteria:Rickettsiales: RecName: Full=30S ribosomal protein S5
. . . Wolbachia endosymbiont of Brugia malayi..... 120 3e-34 1 hit Bacteria:Proteobacteria:Alphaproteobacteria:Rickettsiales: RecName: Full=30S ribosomal protein S5
. . . Wolbachia..... 119 4e-34 1 hit Bacteria:Proteobacteria:Alphaproteobacteria:Rickettsiales: RecName: Full=30S ribosomal protein S5
. . . Methylocella silvestris..... 159 2e-49 1 hit Bacteria:Proteobacteria:Alphaproteobacteria:Rhizobiales: RecName: Full=30S ribosomal protein S5
. . . Beijerinckia indica..... 156 3e-48 1 hit Bacteria:Proteobacteria:Alphaproteobacteria:Rhizobiales: RecName: Full=30S ribosomal protein S5
. . . Parvibaculum lavamentivorans..... 158 3e-49 1 hit Bacteria:Proteobacteria:Alphaproteobacteria:Rhizobiales: RecName: Full=30S ribosomal protein S5
. . . Bartonella quintana..... 156 1e-48 1 hit Bacteria:Proteobacteria:Alphaproteobacteria:Rhizobiales: RecName: Full=30S ribosomal protein S5
. . . Azorhizobium caulinodans..... 156 2e-48 1 hit Bacteria:Proteobacteria:Alphaproteobacteria:Rhizobiales: RecName: Full=30S ribosomal protein S5
. . . Bartonella henselae..... 155 3e-48 15 hits Bacteria:Proteobacteria:Alphaproteobacteria:Rhizobiales: RecName: Full=30S ribosomal protein S5
. . . Mesorhizobium..... 155 4e-48 1 hit Bacteria:Proteobacteria:Alphaproteobacteria:Rhizobiales: RecName: Full=30S ribosomal protein S5 [Mesorhizobiu...
. . . Bartonella..... 155 4e-48 1 hit Bacteria:Proteobacteria:Alphaproteobacteria:Rhizobiales: RecName: Full=30S ribosomal protein S5
. . . Xanthobacter autotrophicus..... 155 5e-48 1 hit Bacteria:Proteobacteria:Alphaproteobacteria:Rhizobiales: RecName: Full=30S ribosomal protein S5
. . . Ochrobactrum..... 154 1e-47 1 hit Bacteria:Proteobacteria:Alphaproteobacteria:Rhizobiales: RecName: Full=30S ribosomal protein S5
. . . Methylobacterium radiotolerans JCM 2831..... 154 1e-47 2 hits Bacteria:Proteobacteria:Alphaproteobacteria:Rhizobiales: RecName: Full=30S ribosomal protein S5
. . . Bartonella bacilliformis..... 154 2e-47 4 hits Bacteria:Proteobacteria:Alphaproteobacteria:Rhizobiales: RecName: Full=30S ribosomal protein S5
. . . Methylobacterium..... 154 2e-47 3 hits Bacteria:Proteobacteria:Alphaproteobacteria:Rhizobiales: RecName: Full=30S ribosomal protein S5
. . . Methylobacterium populi BJ001..... 154 2e-47 2 hits Bacteria:Proteobacteria:Alphaproteobacteria:Rhizobiales: RecName: Full=30S ribosomal protein S5
. . . Rhodopseudomonas palustris..... 153 3e-47 5 hits Bacteria:Proteobacteria:Alphaproteobacteria:Rhizobiales: RecName: Full=30S ribosomal protein S5
. . . Rhodopseudomonas palustris BisB18..... 153 3e-47 2 hits Bacteria:Proteobacteria:Alphaproteobacteria:Rhizobiales: RecName: Full=30S ribosomal protein S5
. . . Chelativorans sp. BNC1..... 153 3e-47 3 hits Bacteria:Proteobacteria:Alphaproteobacteria:Rhizobiales: RecName: Full=30S ribosomal protein S5
. . . Rhizobium/Agrobacterium group..... 153 4e-47 1 hit Bacteria:Proteobacteria:Alphaproteobacteria:Rhizobiales: RecName: Full=30S ribosomal protein S5
. . . Agrobacterium vitis..... 152 6e-47 13 hits Bacteria:Proteobacteria:Alphaproteobacteria:Rhizobiales: RecName: Full=30S ribosomal protein S5
. . . Rhizobium..... 152 6e-47 2 hits Bacteria:Proteobacteria:Alphaproteobacteria:Rhizobiales: RecName: Full=30S ribosomal protein S5
. . . Rhizobium leguminosarum bv. viciae 3841..... 152 7e-47 2 hits Bacteria:Proteobacteria:Alphaproteobacteria:Rhizobiales: RecName: Full=30S ribosomal protein S5
. . . Sinorhizobium/Ensifer group..... 152 9e-47 2 hits Bacteria:Proteobacteria:Alphaproteobacteria:Rhizobiales: RecName: Full=30S ribosomal protein S5
. . . Sinorhizobium meliloti 1021..... 152 1e-46 3 hits Bacteria:Proteobacteria:Alphaproteobacteria:Rhizobiales: RecName: Full=30S ribosomal protein S5
. . . Methylobacterium nodulans..... 151 2e-46 1 hit Bacteria:Proteobacteria:Alphaproteobacteria:Rhizobiales: RecName: Full=30S ribosomal protein S5
. . . Methylobacterium sp. 4-46..... 151 2e-46 2 hits Bacteria:Proteobacteria:Alphaproteobacteria:Rhizobiales: RecName: Full=30S ribosomal protein S5
. . . Brucella..... 150 4e-46 1 hit Bacteria:Proteobacteria:Alphaproteobacteria:Rhizobiales: RecName: Full=30S ribosomal protein S5
. . . Brucella ovis..... 150 6e-46 1 hit Bacteria:Proteobacteria:Alphaproteobacteria:Rhizobiales: RecName: Full=30S ribosomal protein S5
. . . Nitrobacter winogradskyi..... 150 8e-46 1 hit Bacteria:Proteobacteria:Alphaproteobacteria:Rhizobiales: RecName: Full=30S ribosomal protein S5
. . . Oligotropha carboxidovorans..... 149 1e-45 1 hit Bacteria:Proteobacteria:Alphaproteobacteria:Rhizobiales: RecName: Full=30S ribosomal protein S5
. . . Nitrobacter hamburgensis..... 148 5e-45 1 hit Bacteria:Proteobacteria:Alphaproteobacteria:Rhizobiales: RecName: Full=30S ribosomal protein S5
. . . Bradyrhizobium..... 147 6e-45 2 hits Bacteria:Proteobacteria:Alphaproteobacteria:Rhizobiales: RecName: Full=30S ribosomal protein S5
. . . Rhodopseudomonas palustris CGA009..... 146 1e-44 2 hits Bacteria:Proteobacteria:Alphaproteobacteria:Rhizobiales: RecName: Full=30S ribosomal protein S5; AltName: Ful...
```

..Bradyrhizobium diazoefficiens USDA 110..... 146 2e-44 4 hits Bacteria:Proteobacteria:Alphaproteobacteria:Rhizobiales: RecName: Full=30S ribosomal protein S5
 ..Rhodopseudomonas..... 146 3e-44 1 hit Bacteria:Proteobacteria:Alphaproteobacteria:Rhizobiales: RecName: Full=30S ribosomal protein S5
 ..Rhodopseudomonas palustris HaA2..... 145 3e-44 2 hits Bacteria:Proteobacteria:Alphaproteobacteria:Rhizobiales: RecName: Full=30S ribosomal protein S5
 ..Phenylobacterium zucineum..... 149 1e-45 1 hit Bacteria:Proteobacteria:Alphaproteobacteria:Caulobacterales: RecName: Full=30S ribosomal protein S5
 ..Caulobacter..... 149 2e-45 1 hit Bacteria:Proteobacteria:Alphaproteobacteria:Caulobacterales: RecName: Full=30S ribosomal protein S5 [Caulobacter ...
 ..Caulobacter sp. K31..... 148 6e-45 3 hits Bacteria:Proteobacteria:Alphaproteobacteria:Caulobacterales: RecName: Full=30S ribosomal protein S5
 ..Granulibacter bethesdensis..... 147 6e-45 9 hits Bacteria:Proteobacteria:Alphaproteobacteria:Rhodospirillales: RecName: Full=30S ribosomal protein S5
 ..Gluconobacter..... 144 1e-43 1 hit Bacteria:Proteobacteria:Alphaproteobacteria:Rhodospirillales: RecName: Full=30S ribosomal protein S5
 ..Acidiphilium..... 142 8e-43 1 hit Bacteria:Proteobacteria:Alphaproteobacteria:Rhodospirillales: RecName: Full=30S ribosomal protein S5
 ..Gluconacetobacter diazotrophicus..... 138 2e-41 1 hit Bacteria:Proteobacteria:Alphaproteobacteria:Rhodospirillales: RecName: Full=30S ribosomal protein S5
 ..Rhodospirillum rubrum..... 142 8e-43 1 hit Bacteria:Proteobacteria:Alphaproteobacteria:Rhodospirillales: RecName: Full=30S ribosomal protein S5
 ..Magnetospirillum..... 141 4e-42 1 hit Bacteria:Proteobacteria:Alphaproteobacteria:Rhodospirillales: RecName: Full=30S ribosomal protein S5
 ..Jannaschia sp. CCS1..... 146 2e-44 3 hits Bacteria:Proteobacteria:Alphaproteobacteria:Rhodobacterales: RecName: Full=30S ribosomal protein S5
 ..Dinoroseobacter shibae..... 146 2e-44 1 hit Bacteria:Proteobacteria:Alphaproteobacteria:Rhodobacterales: RecName: Full=30S ribosomal protein S5
 ..Ruegeria..... 144 8e-44 1 hit Bacteria:Proteobacteria:Alphaproteobacteria:Rhodobacterales: RecName: Full=30S ribosomal protein S5
 ..Ruegeria pomeroyi..... 144 1e-43 1 hit Bacteria:Proteobacteria:Alphaproteobacteria:Rhodobacterales: RecName: Full=30S ribosomal protein S5
 ..Paracoccus denitrificans..... 144 2e-43 3 hits Bacteria:Proteobacteria:Alphaproteobacteria:Rhodobacterales: RecName: Full=30S ribosomal protein S5
 ..Roseobacter denitrificans..... 143 3e-43 1 hit Bacteria:Proteobacteria:Alphaproteobacteria:Rhodobacterales: RecName: Full=30S ribosomal protein S5
 ..Rhodobacter sphaeroides..... 140 3e-42 4 hits Bacteria:Proteobacteria:Alphaproteobacteria:Rhodobacterales: RecName: Full=30S ribosomal protein S5
 ..Rhodobacter..... 140 4e-42 1 hit Bacteria:Proteobacteria:Alphaproteobacteria:Rhodobacterales: RecName: Full=30S ribosomal protein S5
 ..Hyphomonas..... 146 2e-44 1 hit Bacteria:Proteobacteria:Alphaproteobacteria:Rhodobacterales: RecName: Full=30S ribosomal protein S5
 ..Maricaulis maris..... 145 4e-44 1 hit Bacteria:Proteobacteria:Alphaproteobacteria:Rhodobacterales: RecName: Full=30S ribosomal protein S5
 ..Zymomonas mobilis..... 144 8e-43 1 hit Bacteria:Proteobacteria:Alphaproteobacteria:Sphingomonadales: RecName: Full=30S ribosomal protein S5
 ..Sphingopyxis alaskensis RB2256..... 132 3e-38 2 hits Bacteria:Proteobacteria:Alphaproteobacteria:Sphingomonadales: RecName: Full=30S ribosomal protein S5
 ..Sphingomonas..... 132 3e-38 1 hit Bacteria:Proteobacteria:Alphaproteobacteria:Sphingomonadales: RecName: Full=30S ribosomal protein S5
 ..Novosphingobium..... 125 8e-36 1 hit Bacteria:Proteobacteria:Alphaproteobacteria:Sphingomonadales: RecName: Full=30S ribosomal protein S5
 ..Erythrobacter litoralis HTCC2594..... 143 3e-42 2 hits Bacteria:Proteobacteria:Alphaproteobacteria:Sphingomonadales: RecName: Full=30S ribosomal protein S5
 ..Acidithiobacillus ferrooxidans ATCC 23270..... 127 3e-37 2 hits Bacteria:Proteobacteria:Gammaproteobacteria:Acidithiobacillales: RecName: Full=30S ribosomal protein S5
 ..Buchnera aphidicola..... 123 1e-35 2 hits Bacteria:Proteobacteria:Gammaproteobacteria:Enterobacterales: RecName: Full=30S ribosomal protein S5
 ..Buchnera aphidicola str. APS (Acyrtosiphon pis.. 121 7e-35 3 hits Bacteria:Proteobacteria:Gammaproteobacteria:Enterobacterales: RecName: Full=30S ribosomal protein S5
 ..Pseudomonas mesoacidophila..... 119 5e-34 1 hit Bacteria:Proteobacteria:Gammaproteobacteria:Pseudomonadales: RecName: Full=30S ribosomal protein S5
 ..Nitrosococcus oceani..... 119 8e-34 1 hit Bacteria:Proteobacteria:Gammaproteobacteria:Chromatiales: RecName: Full=30S ribosomal protein S5
 ..Shewanella amazonensis..... 118 1e-33 1 hit Bacteria:Proteobacteria:Gammaproteobacteria:Alteromonadales: RecName: Full=30S ribosomal protein S5
 ..Desulfovibrio salexigens..... 127 3e-37 1 hit Bacteria:Proteobacteria:Deltaproteobacteria:Desulfovibrionales: RecName: Full=30S ribosomal protein S5
 ..Pelobacter carbinolicus..... 122 2e-35 1 hit Bacteria:Proteobacteria:Deltaproteobacteria:Desulfuromonadales: RecName: Full=30S ribosomal protein S5
 ..Ralstonia solanacearum..... 121 8e-35 25 hits Bacteria:Proteobacteria:Betaproteobacteria:Burkholderiales: RecName: Full=30S ribosomal protein S5
 ..Ralstonia..... 120 2e-34 1 hit Bacteria:Proteobacteria:Betaproteobacteria:Burkholderiales: RecName: Full=30S ribosomal protein S5
 ..Ralstonia sp. GA3-3..... 119 4e-34 1 hit Bacteria:Proteobacteria:Betaproteobacteria:Burkholderiales: RecName: Full=30S ribosomal protein S5
 ..Cupriavidus pinatubonensis..... 119 3e-34 1 hit Bacteria:Proteobacteria:Betaproteobacteria:Burkholderiales: RecName: Full=30S ribosomal protein S5
 ..Burkholderia..... 119 5e-34 2 hits Bacteria:Proteobacteria:Betaproteobacteria:Burkholderiales: RecName: Full=30S ribosomal protein S5
 ..Burkholderiaceae..... 119 6e-34 5 hits Bacteria:Proteobacteria:Betaproteobacteria:Burkholderiales: RecName: Full=30S ribosomal protein S5
 ..Burkholderia phymatum STM815..... 119 7e-34 2 hits Bacteria:Proteobacteria:Betaproteobacteria:Burkholderiales: RecName: Full=30S ribosomal protein S5
 ..Burkholderia thailandensis E264..... 119 8e-34 3 hits Bacteria:Proteobacteria:Betaproteobacteria:Burkholderiales: RecName: Full=30S ribosomal protein S5
 ..Burkholderia pseudomallei K96243..... 119 8e-34 4 hits Bacteria:Proteobacteria:Betaproteobacteria:Burkholderiales: RecName: Full=30S ribosomal protein S5
 ..Cupriavidus..... 118 1e-33 2 hits Bacteria:Proteobacteria:Betaproteobacteria:Burkholderiales: RecName: Full=30S ribosomal protein S5
 ..Burkholderia phytofirmans PsJN..... 118 1e-33 2 hits Bacteria:Proteobacteria:Betaproteobacteria:Burkholderiales: RecName: Full=30S ribosomal protein S5
 ..Delftia..... 119 5e-34 1 hit Bacteria:Proteobacteria:Betaproteobacteria:Burkholderiales: RecName: Full=30S ribosomal protein S5
 ..Methylibium..... 119 9e-34 1 hit Bacteria:Proteobacteria:Betaproteobacteria:Burkholderiales: RecName: Full=30S ribosomal protein S5
 ..Polaromonas..... 118 1e-33 1 hit Bacteria:Proteobacteria:Betaproteobacteria:Burkholderiales: RecName: Full=30S ribosomal protein S5
 ..Syntrophomonas wolfei subsp. wolfei str. Goetti... 123 1e-35 2 hits Bacteria:Firmicutes:Clostridia:Clostridiales: RecName: Full=30S ribosomal protein S5
 ..Desulfotomaculum reducens..... 119 7e-34 1 hit Bacteria:Firmicutes:Clostridia:Clostridiales: RecName: Full=30S ribosomal protein S5
 ..Natranaerobius thermophilus..... 122 4e-35 1 hit Bacteria:Firmicutes:Clostridia:Natranaerobiales: RecName: Full=30S ribosomal protein S5
 ..Mycoplasma agalactiae..... 121 4e-34 1 hit Bacteria:Tenericutes:Mollicutes:Mycoplasmatales: RecName: Full=30S ribosomal protein S5

[Results end]

This output shows that the kingdom is **bacteria** and phylum is **Proteobacteria** with class **Alphaproteobacteria** and subclass **Rickettsiales**.

The output of the **Taxonomt List** tool table is below:

Taxonomy synthesis table

Classification	nb lines	nb hits	min E-value	max E-value
Bacteria:Proteobacteria:Alphaproteobacteria:Rickettsiales	14	21	6e-68	4e-34
Bacteria:Proteobacteria:Alphaproteobacteria:Rhizobiales	35	87	2e-49	3e-44
Bacteria:Proteobacteria:Alphaproteobacteria:Caulobacteriales	3	5	1e-45	6e-45
Bacteria:Proteobacteria:Alphaproteobacteria:Rhodospirillales	6	14	6e-45	2e-41
Bacteria:Proteobacteria:Alphaproteobacteria:Rhodobacterales	10	17	2e-44	4e-42
Bacteria:Proteobacteria:Alphaproteobacteria:Sphingomonadales	5	7	8e-43	8e-36
Bacteria:Proteobacteria:Gammaproteobacteria:Acidithiobacillales	1	2	3e-37	3e-37
Bacteria:Proteobacteria:Gammaproteobacteria:Enterobacteriales	1	2	1e-35	1e-35
Bacteria:Proteobacteria:Gammaproteobacteria:Pseudomonadales	1	1	5e-34	5e-34
Bacteria:Proteobacteria:Gammaproteobacteria:Chromatiales	1	1	8e-34	8e-34
Bacteria:Proteobacteria:Gammaproteobacteria:Alteromonadales	1	1	1e-33	1e-33
Bacteria:Proteobacteria:Deltaproteobacteria:Desulfovibrionales	1	1	3e-37	3e-37
Bacteria:Proteobacteria:Deltaproteobacteria:Desulfuromonadales	1	1	2e-35	2e-35
Bacteria:Proteobacteria:Betaproteobacteria:Burkholderiales	14	51	8e-35	1e-33
Bacteria:Firmicutes:Clostridia:Clostridiales	2	3	1e-35	7e-34
Bacteria:Firmicutes:Clostridia:Natranaerobiales	1	1	4e-35	4e-35
Bacteria:Tenericutes:Mollicutes:Mycoplasmatales	1	1	4e-34	4e-34

[Results end]

This synthesis table shows that the topmost row has the lowest E-value and out of the 17 records most of the records indicates the class **Alphaproteobacteria** and the top one corresponds to the subclass **Rickettsiales**.

The sequences below are in **FASTA format** and were obtained from the **TaxReports** tool:

>BPARPC1 [Bacteria Proteobacteria Alphaproteobacteria Rickettsiales Pelagibacterales Candidatus Pelagibacter]
Bacteria;Proteobacteria;Alphaproteobacteria;Rickettsiales;Pelagibacterales;Candidatus Pelagibacter; gi|494054731|ref|WP_006996828.1| 30S ribosomal
protein S5 [Candidatus Pelagibacter ubique]
MDFKDKDDGILEKLVHINRITKVVKGGRRFGFSALVVGNQAGKIGVAHAKAKQVPDAIKKANETARRVLHVPLREGR
TIHHDVYKGDGAGKIILRSAPKGTGIIAGGPVRAVCEVLGIKDIVAKSMGTSNPHNMIRATMKALSKQNSPKHIATIRNK
KISDVIEKRG

>BPARBM1 [Bacteria Proteobacteria Alphaproteobacteria Rhizobiales Beijerinckiaceae Methylocella]
Bacteria;Proteobacteria;Alphaproteobacteria;Rhizobiales;Beijerinckiaceae;Methylocella; gi|501585304|ref|WP_012589605.1| 30S ribosomal protein S5
[Methylocella silvestris]
MAREGEGGRDRREERDSEFMDRLVHINRVAKVVKGGRRFGAALVVGDQKGRVGFHGKAREVPEAIRKATESAKRSL
IRVPLREGRTLHHDVNGRHGAGRVVLAAPAGTGIIAGGPMRAIFETLGMHDVVAKSQSSNPYNMVRATFDALGRQDSP
RAVAARRNLKVSVLQGRRPGEAAEASEA

>BPARPP1 [Bacteria Proteobacteria Alphaproteobacteria Rhizobiales Phyllobacteriaceae Parvibaculum]
Bacteria;Proteobacteria;Alphaproteobacteria;Rhizobiales;Phyllobacteriaceae;Parvibaculum; gi|501060331|ref|WP_012111675.1| 30S ribosomal protein S5
[Parvibaculum lavamentivorans]
MARKENRDRDREERDSEFVDKLVHINRVAKVVKGGRRFGAALVVGDQKGRVGFHGKAREVPEAIRKATEAAKRAMIR
VPLREGRTLHHDVDRHGAGRVVLAAPPGTGIIAGGPMRAVETLGVQDVVAKSLGSSNPYNMVRATFDALKNEQSPRM
VAARRGKKVSEIVGRRSDAAAEAGLE

>BPARBB1 [Bacteria Proteobacteria Alphaproteobacteria Rhizobiales Bartonellaceae Bartonella]
Bacteria;Proteobacteria;Alphaproteobacteria;Rhizobiales;Bartonellaceae;Bartonella; gi|499492896|ref|WP_011179536.1| 30S ribosomal protein S5 [Bartonella
quintana]
MAQKERGERDERDGEFVDRLVHINRVAKVVKGGRRFGAALVVGDQKGRVGFHGKAREVPEAVRKATEAAKRGMIYVP
LRSGRTLHHDLEGRHGAGRVLLRSASAGTGIIAGGPMRAIFETLGMQDVVAKSLGSSNPYNMVRATFDALKHQMHPDIA
VQRGIKYSILQARRQHLVDAEG

>BPARXA1 [Bacteria Proteobacteria Alphaproteobacteria Rhizobiales Xanthobacteraceae Azorhizobium]
Bacteria;Proteobacteria;Alphaproteobacteria;Rhizobiales;Xanthobacteraceae;Azorhizobium; gi|501121910|ref|WP_012171063.1| 30S ribosomal protein S5
[Azorhizobium caulinodans]
MARDREHRAERDRDNTFVDKLVHINRVAKVVKGGRRFGAALVIVGDQKGRVGFHGKAREVPEAIRKATESAKRALIR
VPLREGRTLHHDVHGHGAGKVLRAAPAGTGIIAGGPMRAVETLGVHDVVAKSFGSSNPYNMVRATFDALTNQDSPRS
VAARRGLKVSQLSRRRVEDAEATD

>BPARBB2 [Bacteria Proteobacteria Alphaproteobacteria Rhizobiales Beijerinckiaceae Beijerinckia]
Bacteria;Proteobacteria;Alphaproteobacteria;Rhizobiales;Beijerinckiaceae;Beijerinckia; gi|501352733|ref|WP_012384368.1| 30S ribosomal protein S5
[Beijerinckia indica]
MAREGEGGRGGREERDSEFMDRLVHINRVAKVVKGGRRFGAALVVGDQKGRVGYGHGKAREVPEAIRKATESAKRAL
IRVPLREGRTLHHDVNGRHGAGRVVLAAPAGTGIIAGGPMRAVETLGMHDVVAKSQSSNPYNMIRATFDALQREDSP
RAVAARRSLKVSVLQGRRLGDTETAEG

>BPARBB3 [Bacteria Proteobacteria Alphaproteobacteria Rhizobiales Bartonellaceae Bartonella]
Bacteria;Proteobacteria;Alphaproteobacteria;Rhizobiales;Bartonellaceae;Bartonella; gi|499494257|ref|WP_011180897.1| 30S ribosomal protein S5 [Bartonella
henselae]
MAQKERSERDERESEFVDRLVHINRVAKVVKGGRRFGAALVVGDQKGRVGFHGKAREVPEAVRKATESAKRGMIIYVP
LRSGRTLHHDLEGGHAGRVLLRSASAGTGIIAGGPMRAIFETLGMQDVVAKSLGSSNPYNMVRATFDALKRQMHPDIA
AQRGIKYSTLQARRQHLVDAEG

>BPARPM1 [Bacteria Proteobacteria Alphaproteobacteria Rhizobiales Phyllobacteriaceae Mesorhizobium]
Bacteria;Proteobacteria;Alphaproteobacteria;Rhizobiales;Phyllobacteriaceae;Mesorhizobium; gi|696358248|ref|WP_032933400.1| MULTISPECIES: 30S
ribosomal protein S5 [Mesorhizobium]
MAQERREGGREREERDDGMVMDKLVHINRVAKVVKGGRRFGAALVVGDQKGRVGFHGKAREVPEAIRKATESAKRD
MIFVPLRSGRTLHHDVEGRWGAGRVLLRAAQGTGIIAGGPMRAVETLGMHDVVAKSMGSSNPYNMVRATFDALKSQMH
PKDVAAARGIKYSTLQARRGTAVAAEE

>BPARBB4 [Bacteria Proteobacteria Alphaproteobacteria Rhizobiales Bartonellaceae Bartonella]
 Bacteria;Proteobacteria;Alphaproteobacteria;Rhizobiales;Bartonellaceae;Bartonella; gi|492186901|ref|WP_005773287.1| MULTISPECIES: 30S ribosomal protein S5 [Bartonella]
 MAQKERGEREERDNEFVDRDLVHINRVAKVVKGGRRFGFAALVVVDQKGRVGFHGKAREVPEAVRKATESAKRGMIVVP
 LRSRTLHHDLEGRHGAGRLLRSASAGTGIIAGGPMRAIFEALGMQDVVAKSLGSSNPYNMVRATFDALKHQMHPRDIA
 AQRGIKYSTLQARRRHLVDVEG

>BPARXX1 [Bacteria Proteobacteria Alphaproteobacteria Rhizobiales Xanthobacteraceae Xanthobacter]
 Bacteria;Proteobacteria;Alphaproteobacteria;Rhizobiales;Xanthobacteraceae;Xanthobacter; gi|501065516|ref|WP_012116752.1| 30S ribosomal protein S5 [Xanthobacter autotrophicus]
 MAREREQRHEREDRDNTFVDKLVHINRVAKVVKGGRRFGFAALVVVDQKGRVGFHGKAREVPEAIRKATESAKRALIR
 VPLREGRTLHHDVHGHHGAGKVLRAAPAGTGIIAGGPMRAVFETLGMHDVVAKSLGSSNPYNMVRATFDALKNQDSPRS
 VAARRGIKVSQLSRRRVEDAEATD

>BPARBO1 [Bacteria Proteobacteria Alphaproteobacteria Rhizobiales Brucellaceae Ochrobactrum]
 Bacteria;Proteobacteria;Alphaproteobacteria;Rhizobiales;Brucellaceae;Ochrobactrum; gi|493512698|ref|WP_006467009.1| MULTISPECIES: 30S ribosomal protein S5 [Ochrobactrum]
 MAQRENRREDRGREERDSEFVDKLVHINRVAKVVKGGRRFGFAALVVVDQKGRVGFHGKAREVPEAIRKATEAAKRDM
 IFVPLRSRTLHHDVGRHGAGKVLRAAPAGTGIIAGGPMRAVFETLGMQDVVAKSLGSSNPYNMVRATFDALKHQMH
 KDIAAQRGIKYSTLQARRHDVVGSEE

>BPARMM1 [Bacteria Proteobacteria Alphaproteobacteria Rhizobiales Methylobacteriaceae Methylobacterium]
 Bacteria;Proteobacteria;Alphaproteobacteria;Rhizobiales;Methylobacteriaceae;Methylobacterium; gi|501276155|ref|WP_012319173.1| MULTISPECIES: 30S ribosomal protein S5 [Bacteria]
 MAREREGGGRREDREERDSEFVDKLVHINRVAKVVKGGRRFGFAALVVVDQKGRVGFHGKAREVPEAIRKATEAAK
 RGLIRVSLREGRTLHHDVNGRHHGAGKVLRAAPQGTGIIAGGPMRAVFETLGMQDVVAKSLGSSNPYNLVRATFEALKNE
 DSPRSVAARRGIKYSTLQARRRDADPADQSEAAVA

>BPARBB5 [Bacteria Proteobacteria Alphaproteobacteria Rhizobiales Bartonellaceae Bartonella]
 Bacteria;Proteobacteria;Alphaproteobacteria;Rhizobiales;Bartonellaceae;Bartonella; gi|492160661|ref|WP_005766947.1| 30S ribosomal protein S5 [Bartonella bacilliformis]
 MAQKERSDHNDSDERRNESGGEFIDKLVHINRVAKVVKGGRRFGFAALVVVDQKGRVGFHGKAREVPEAIRKATESAK
 REMIYVPLRSRTLHHDVGRHGAGRLLRSASAGTGIIAGGSMRAIFETLGMQDVVAKSLGSSNPYNMVRATFDALKRQ
 MHPRDIAAQRGIKYSTLQARRQYIVGMEE

>BPARMM2 [Bacteria Proteobacteria Alphaproteobacteria Rhizobiales Methylobacteriaceae Methylobacterium]
 Bacteria;Proteobacteria;Alphaproteobacteria;Rhizobiales;Methylobacteriaceae;Methylobacterium; gi|501210634|ref|WP_012253652.1| MULTISPECIES: 30S ribosomal protein S5 [Methylobacterium]
 MAREREGRRRDDREERDSEFVDKLVHINRVAKVVKGGRRFGFAALVVVDQKGRVGFHGKAREVPEAIRKATEAAKRGL
 IRVSLREGRTLHHDVNGRHHGAGKVLRAAPQGTGIIAGGPMRAVFETLGMQDVVAKSLGSSNPYNLVRATFDALKNEDSP
 RSVARRGLKVSALQARRRDADPADTSEAAVA

>BPARMM3 [Bacteria Proteobacteria Alphaproteobacteria Rhizobiales Methylobacteriaceae Methylobacterium]
 Bacteria;Proteobacteria;Alphaproteobacteria;Rhizobiales;Methylobacteriaceae;Methylobacterium; gi|489692953|ref|WP_003597131.1| MULTISPECIES: 30S ribosomal protein S5 [Methylobacterium]
 MAREREGRRRDDREERDSEFVDKLVHINRVAKVVKGGRRFGFAALVVVDQKGRVGFHGKAREVPEAIRKATEAAKRGL
 IRVSLREGRTLHHDVNGRHHGAGKVLRAAPQGTGIIAGGPMRAVFETLGMQDVVAKSLGSSNPYNLVRATFDALKNEDSP
 RSVARRGLKVSALQARRRDADPADTSDAAVA

>BPARBR1 [Bacteria Proteobacteria Alphaproteobacteria Rhizobiales Bradyrhizobiaceae Rhodopseudomonas]
 Bacteria;Proteobacteria;Alphaproteobacteria;Rhizobiales;Bradyrhizobiaceae;Rhodopseudomonas; gi|499984254|ref|WP_011664972.1| 30S ribosomal protein S5 [Rhodopseudomonas palustris]
 MAGERERGGDRGPREERDSEFVDKLVHINRVAKVVKGGKRFGFAALVVDQKGRVGFHGKAREVPEAIRKATESAKR
 NLTRVPLREGRTLHHDIAGRHGAGRVLRAAPAGTGIIAGGPMRAVFETLGIQDVVAKSIGSSNPYNMVRATFDALKHQD
 SPRSVAARRNIKVSTLQSRVGGDAEVAE

>BPARBR2 [Bacteria Proteobacteria Alphaproteobacteria Rhizobiales Bradyrhizobiaceae Rhodopseudomonas]
 Bacteria;Proteobacteria;Alphaproteobacteria;Rhizobiales;Bradyrhizobiaceae;Rhodopseudomonas; gi|499793124|ref|WP_011473858.1| 30S ribosomal protein S5 [Rhodopseudomonas palustris]

MAGERERGGGHRPREERDSEFVDKLVHINRVAKVVKGGRRFGFAALVVGDDQKGRVGFHGKAREVPEAIRKATESAKR
NLTRVPLREGRTLHHDHDIAGRHRGAGRVYLRAPAGTGIIAGGPMRAVFETLGIQDVVAKSIGSSNPYNMVRATFDALKHQD
SPRSVAARRNLKVSTLQSRVGGDAEVVAE

>BPAPPC2 [Bacteria Proteobacteria Alphaproteobacteria Rhizobiales Phyllobacteriaceae Chelativorans]
Bacteria;Proteobacteria;Alphaproteobacteria;Rhizobiales;Phyllobacteriaceae;Chelativorans; gi|499900265|ref|WP_011580999.1| 30S ribosomal protein S5
[Chelativorans sp. BNC1]

MAQPRRDRREREERDSEFVDKLVHINRVAKVVKGGRRFGFAALVVGDDQKGRVGFHGKAREVPEAIRKATESAKRDLI
FVPLREGRTLHHDVGGHHGAGRVLMRAAKPGTGIIAGGPMRAVFETLGVQDVVAKSLGSSNPYNMVRATFDALKRQMHPK
DIAAQRGIKYSTLQARRGTAVATEE

>BPARRR1 [Bacteria Proteobacteria Alphaproteobacteria Rhizobiales Rhizobiaceae Rhizobium/Agrobacterium group]
Bacteria;Proteobacteria;Alphaproteobacteria;Rhizobiales;Rhizobiaceae;Rhizobium/Agrobacterium group;Agrobacterium; gi|489590764|ref|WP_003495208.1|
MULTISPECIES: 30S ribosomal protein S5 [Rhizobium/Agrobacterium group]

MAQEKRSRDDRQNEREERDSEFVDKLVAINRVAKVVKGGRRFGFAALVVGDDQKGRVGFHGKAREVPEAIRKATESAKR
DLIFVPLRDGRTLHHDVNGRHGAGKVLRSKAGTGIIAGGPMRAVFETLGVHDVAKSTGSSNPYNMVRATFDALKHQV
HPKDIAAQRLKYATLQARRAASGNASEE

>BPARRR2 [Bacteria Proteobacteria Alphaproteobacteria Rhizobiales Rhizobiaceae Rhizobium/Agrobacterium group]
Bacteria;Proteobacteria;Alphaproteobacteria;Rhizobiales;Rhizobiaceae;Rhizobium/Agrobacterium group;Agrobacterium; gi|506396036|ref|WP_015915755.1|
30S ribosomal protein S5 [Agrobacterium vitis]

MAQEKRSRDDRQNEREERDSEFVDKLVAINRVAKVVKGGRRFGFAALVVGDDQKGRVGFHGKAREVPEAIRKATESAKR
DLIFVPLRDGRTLHHDVHGRHGAGKVLRSKAGTGIIAGGPMRAVFETLGMHDVAKSTGSSNPYNMVRATFDALKHQV
HPKDVAQAQRLKYATLQARRAASGNASEE

>BPARRR3 [Bacteria Proteobacteria Alphaproteobacteria Rhizobiales Rhizobiaceae Rhizobium/Agrobacterium group]
Bacteria;Proteobacteria;Alphaproteobacteria;Rhizobiales;Rhizobiaceae;Rhizobium/Agrobacterium group;Rhizobium; gi|489669518|ref|WP_003573784.1|
MULTISPECIES: 30S ribosomal protein S5 [Rhizobium]

MAQERRPQREDRQSEERDSEFVDKLVAINRVAKVVKGGRRFGFAALVVGDDQKGRVGFHGKAREVPEAIRKATEAAKR
ELIFVPLRDGRTLHHDVHGRHGAGKVLRSKAGTGIIAGGPMRAVFETLGMHDVAKSTGSSNPYNMVRATFDALKHQV
HPKDIAAQRGIKYATLQARRSASGNASEE

>BPARRR4 [Bacteria Proteobacteria Alphaproteobacteria Rhizobiales Rhizobiaceae Rhizobium/Agrobacterium group]
Bacteria;Proteobacteria;Alphaproteobacteria;Rhizobiales;Rhizobiaceae;Rhizobium/Agrobacterium group;Rhizobium; gi|489643125|ref|WP_003547565.1|
MULTISPECIES: 30S ribosomal protein S5 [Rhizobium]

MAQERRPQRDRQSEERDSEFVDKLVAINRVAKVVKGGRRFGFAALVVGDDQKGRVGFHGKAREVPEAIRKATEAAKR
ELIFVPLRDGRTLHHDVHGRHGAGKVLRSKAGTGIIAGGPMRAVFETLGMHDVAKSTGSSNPYNMVRATFDALKHQV
HPKDIAAQRGIKYATLQARRSASGNASEE

>BPARRS1 [Bacteria Proteobacteria Alphaproteobacteria Rhizobiales Rhizobiaceae Sinorhizobium/Ensifer group]
Bacteria;Proteobacteria;Alphaproteobacteria;Rhizobiales;Rhizobiaceae;Sinorhizobium/Ensifer group;Sinorhizobium; gi|502093970|ref|WP_012707768.1|
MULTISPECIES: 30S ribosomal protein S5 [Sinorhizobium/Ensifer group]

MAQERRSREDRQNEREERDSEFVDKLVAINRVAKVVKGGRRFGFAALVVGDDQKGRVGFHGKAREVPEAIRKATEAAKR
DLIFVPLRGRTLHHDVNGRHGAGKVLRSKAGTGIIAGGPMRAVFETLGVHDVAKSTGSSNPYNMVRATFDALKNQM
HPKDIAAQRMKYATLQARRVAAGAASEE

>BPARRS2 [Bacteria Proteobacteria Alphaproteobacteria Rhizobiales Rhizobiaceae Sinorhizobium/Ensifer group]
Bacteria;Proteobacteria;Alphaproteobacteria;Rhizobiales;Rhizobiaceae;Sinorhizobium/Ensifer group;Sinorhizobium; gi|489632070|ref|WP_003536510.1|
MULTISPECIES: 30S ribosomal protein S5 [Sinorhizobium/Ensifer group]

MAQERRSREDRQNEREERDSEFVDKLVAINRVAKVVKGGRRFGFAALVVGDDQKGRVGFHGKAREVPEAIRKATEAAKR
DLIFVPLRGRTLHHDVHGRHGAGKVLRSKAGTGIIAGGPMRAVFETLGVHDVAKSTGSSNPYNMVRATFDALKNQM
HPKDIAAQRMKYATLQSRVSAGVASEE

>BPARRM4 [Bacteria Proteobacteria Alphaproteobacteria Rhizobiales Methylobacteriaceae Methylobacterium]
Bacteria;Proteobacteria;Alphaproteobacteria;Rhizobiales;Methylobacteriaceae;Methylobacterium; gi|506408884|ref|WP_015928603.1| 30S ribosomal protein
S5 [Methylobacterium nodulans]

MAREREGRRDREREERDSEFVDRLVHINRVAKVVKGGRRFGFAALVVGDDQKGRVGFHGKAREVPEAIRKATEAAKRGL
VRVALREGRTLHHDVAGRHRGAGKVLRAAPQGTGIIAGGPMRAVFESLGMQDVVAKSLGSSNPYNLIRATFDALKREDSF
RAVAARRGLKVSTLQARRRDAEPGADAADA

>BPARMM5 [Bacteria Proteobacteria Alphaproteobacteria Rhizobiales Methylobacteriaceae Methylobacterium]
 Bacteria;Proteobacteria;Alphaproteobacteria;Rhizobiales;Methylobacteriaceae;Methylobacterium; gi|501287307|ref|WP_012330325.1| MULTISPECIES: 30S
 ribosomal protein S5 [Methylobacterium]
 MAREREGRRRDEREERDTEFVDRLVHINRVAKVVKGGRRFGFAALVVVDQKGRVGFHGKAREVPEAIRKATEAAKRL
 IRVALREGRTLHHDVAGRHSKVKILRAAPQGTGIIAGGPMRAVFESVGMQDVVAKSLGSSNPYNLIRATFDALKREDS
 RAVAARRGLKVSALQARRRDAEPGSADSADAA

>BPARBB6 [Bacteria Proteobacteria Alphaproteobacteria Rhizobiales Brucellaceae Brucella]
 Bacteria;Proteobacteria;Alphaproteobacteria;Rhizobiales;Brucellaceae;Brucella; gi|489054169|ref|WP_002964345.1| MULTISPECIES: 30S ribosomal protein S5
 [Brucella]
 MAQRERNREERGREERDSEFVDKLVHINRVAKVVKGGRRFGFAALVVVDQKGRVGFHGKAREVPEAIRKATEAAKRD
 IFVPLRSRRTLHHDVGRHAGKVLRAAPAGKGIAGGPMRAVFETLGVQDVVAKSLGSSNPYNMVRATFDALKHQMHP
 KDIAAQRGIKYSTLQARRHDVVGSEE

>BPARBB7 [Bacteria Proteobacteria Alphaproteobacteria Rhizobiales Brucellaceae Brucella]
 Bacteria;Proteobacteria;Alphaproteobacteria;Rhizobiales;Brucellaceae;Brucella; gi|492860458|ref|WP_006012776.1| 30S ribosomal protein S5 [Brucella ovis]
 MAQRERNREERGREERDSEFVDKLVHINRVAKVVKGGRRFGFAALVVVDQKGRVGFHGKAREVPEAIRKATEAAKRD
 IFVPLRSRRTLHHDVGRHAGKVLRAAPAGKGIAGGPMRAVFETLGVQDVVAKSLGSSNPYNMVRATFDALKHQMHP
 KDIAAQRGIKYSTLQARRHDVVGSEE

>BPARBN1 [Bacteria Proteobacteria Alphaproteobacteria Rhizobiales Bradyrhizobiaceae Nitrobacter]
 Bacteria;Proteobacteria;Alphaproteobacteria;Rhizobiales;Bradyrhizobiaceae;Nitrobacter; gi|499633921|ref|WP_011314655.1| 30S ribosomal protein S5
 [Nitrobacter winogradskyi]
 MAGERERGGRRSRDEREERDSEFVDKLVHINRVAKVVKGGRRFGFAALVVVDQKGRVGFHGKAREVPEAIRKATESAK
 RNLTRVPLREGRTLHHDVGRHAGRVYLRAPAGTGIAGGPMRAVFETLGIQDVVAKSVGSSNPYNMVRATFDALKHQ
 DSPRSVSARRNIKVSTLQSRVGGDAEAVAE

>BPAACP1 [Bacteria Proteobacteria Alphaproteobacteria Caulobacteriales Caulobacteraceae Phenyllobacterium]
 Bacteria;Proteobacteria;Alphaproteobacteria;Caulobacteriales;Caulobacteraceae;Phenyllobacterium; gi|501513868|ref|WP_012521805.1| 30S ribosomal
 protein S5 [Phenyllobacterium zucineum]
 MARRNEERGDRRREEERDSEIVEKLVHINRVAATVKGRRRFSFAALMVVDQKGRVGFHGKAREVPEAIRKATEEAKK
 SMIRVPLRESRTLHHDGAGRWGAGKVMVRAAPPPTGVIAGGPMRAVLETLGVQDVVKGSSNPYNMVRATFEALKAQS
 SPRQVAARKGKVSVDVIGRRADGASAAQPAADAEG

>BPARBO2 [Bacteria Proteobacteria Alphaproteobacteria Rhizobiales Bradyrhizobiaceae Oligotropha]
 Bacteria;Proteobacteria;Alphaproteobacteria;Rhizobiales;Bradyrhizobiaceae;Oligotropha; gi|501558352|ref|WP_012562851.1| 30S ribosomal protein S5
 [Oligotropha carboxidovorans]
 MAAERERGGRRGRDRDEREERDSEFVDKLVHINRVAKVVKGGRRFGFAALVVVDQKGRVGFHGKAREVPEAIRKATDSA
 KRNLTRVPLREGRTLHHDVGRHAGRVYLRAPAGTGIAGGPMRAVFETLGIQDVVAKSVGSSNPYNMIRATFDALKH
 QDSPRSVANRRNIKVSVLQARRVGGDAEASAD

>BPAACC1 [Bacteria Proteobacteria Alphaproteobacteria Caulobacteriales Caulobacteraceae Caulobacter]
 Bacteria;Proteobacteria;Alphaproteobacteria;Caulobacteriales;Caulobacteraceae;Caulobacter; gi|499221604|ref|WP_010919144.1| MULTISPECIES: 30S
 ribosomal protein S5 [Caulobacter]
 MARGEQQRGEQRRDRDRNAPEERVDSDIVEKLVHINRVAATVKGRRRFSFAALMVVDQKGRVGFHGKAREVPEAI
 RKATEEAKKTMRVPLRESRTLHHDGAGRWGAGKVMVRAAPPPTGVIAGGPMRAVLETLGVQDVVAKSTGSSNPYNMVR
 TFEALKVQSSPRQIAAKRGKVDLGRRADGASAPEAIEG

>BPARBN2 [Bacteria Proteobacteria Alphaproteobacteria Rhizobiales Bradyrhizobiaceae Nitrobacter]
 Bacteria;Proteobacteria;Alphaproteobacteria;Rhizobiales;Bradyrhizobiaceae;Nitrobacter; gi|499829336|ref|WP_011510070.1| 30S ribosomal protein S5
 [Nitrobacter hamburgensis]
 MAGERERGGHRGGREERDSEFVDKLVHINRVAKVVKGGRRFGFAALVVVDQKGRVGFHGKAREVPEAIRKATESAKR
 NLTRVALREGRTLHHDVGRHAGRVYLRAPAGTGIAGGPMRAVFETLGIQDVVAKSVGSSNPYNMIRATFDALKHQD
 SPRSVSARRNIKVSTLQSRVGGDAEAVAE

>BPARBB8 [Bacteria Proteobacteria Alphaproteobacteria Rhizobiales Bradyrhizobiaceae Bradyrhizobium]
 Bacteria;Proteobacteria;Alphaproteobacteria;Rhizobiales;Bradyrhizobiaceae;Bradyrhizobium; gi|493660508|ref|WP_006611855.1| MULTISPECIES: 30S
 ribosomal protein S5 [Bradyrhizobium]
 MAGERERGGDRKEREERDSEFVDKLVHINRVAKVVKGGRRFGFAALVVVDQKGRVGFHGKAREVPEAIRKATDSAKR

NLTRVALREGRTLHHDIFGRHGAGRVYLAAPAGTGIIAGGPMRAVFETLGIQDVVAKSIGSSNPYNMVRATFNALKHQD
SPRSVAARRNIKVSTLQARRVGGDAEAAAD

>BPACCC2 [Bacteria Proteobacteria Alphaproteobacteria Caulobacterales Caulobacteraceae Caulobacter]
Bacteria;Proteobacteria;Alphaproteobacteria;Caulobacterales;Caulobacteraceae;Caulobacter; gi|501242661|ref|WP_012285679.1| 30S ribosomal protein S5
[Caulobacter sp. K31]
MARGDQQRGEGGQNRGRDRNAPEERVDSIDIVEKLVHINVAATVKGGRRFSAALMVVDQKGRVGYGHGKAREVPEAI
RKATEEAKKTMVRVPLRESRTLHHDGYGRWGAGKVMMAAPPPTGVIAGGPMRAVLETGVDVAKSTGSSNPYNMVRA
TFEALKVQSSPRQIAAKRGKKGVDILGRRADGASAPDAIEG

>BPARAG1 [Bacteria Proteobacteria Alphaproteobacteria Rhodospirillales Acetobacteraceae Granulibacter]
Bacteria;Proteobacteria;Alphaproteobacteria;Rhodospirillales;Acetobacteraceae;Granulibacter; gi|499950544|ref|WP_011631278.1| 30S ribosomal protein S5
[Granulibacter bethesdensis]
MAREPREGRGGREREADDIIDKLVINRVAKVVKGGRRFAAALVVVDQKGRVGYGAGKAREVPEAIRKATERAKRSMI
RVPMKEGRTLHHDVYGHGAGKVVLRSATAGTGIIAGGPMRAVFESLIGDQVAKSLGSRNPHNMVKATFAALERCSSPR
SVASRRGKVSILGRRREPVAQGEEEHA

>BPARB3 [Bacteria Proteobacteria Alphaproteobacteria Rhizobiales Bradyrhizobiaceae Rhodopseudomonas]
Bacteria;Proteobacteria;Alphaproteobacteria;Rhizobiales;Bradyrhizobiaceae;Rhodopseudomonas; gi|499472138|ref|WP_011158778.1| 30S ribosomal protein
S5 [Rhodopseudomonas palustris]
MAAERERGGRRSREREERDSEFVDKLVHINRVAKVVKGGKRFGFAALVVVDQKGRVGFHGKAREVPEAIRKATESAK
RNLTRVALREGRTLHHDIAGRHGAGRVYLAAPAGTGIIAGGPMRAVFETLGIADVAKSVGSSNPYNMVRATFDALKHL
DSPRSVAARRNIKVSTLQARRVGGDAEVAE

>BPARB9 [Bacteria Proteobacteria Alphaproteobacteria Rhizobiales Bradyrhizobiaceae Bradyrhizobium]
Bacteria;Proteobacteria;Alphaproteobacteria;Rhizobiales;Bradyrhizobiaceae;Bradyrhizobium; gi|495411634|ref|WP_008136332.1| MULTISPECIES: 30S
ribosomal protein S5 [Bradyrhizobium]
MAAERERGGRRSREREERDSEFVDKLVHINRVAKVVKGGKRFGFAALVVVDQKGRAGFGHGKAREVPEAIRK
ATESAKRNLTRVSLREGRTLHHDIAGRHGAGRVYLAAPAGTGIIAGGPMRAVFETLGVQDVVAKSIGSSNPYNMVRATF
DALKHQDSPRSVAARRNIKVSTLQARRVGGDAEAAAD

>BPARR1 [Bacteria Proteobacteria Alphaproteobacteria Rhodobacterales Rhodobacteraceae Jannaschia]
Bacteria;Proteobacteria;Alphaproteobacteria;Rhodobacterales;Rhodobacteraceae;Jannaschia; gi|499772999|ref|WP_011453733.1| 30S ribosomal protein S5
[Jannaschia sp. CCS1]
MAERENRGRGRNRNEETPEFADRLVAINRVSKTVKGGKRFGFAALVVVDQGRVGFHGKAKEVPEAIRKATEQAKR
QLIRVPLKEGRTLHHDVKGHHGAGKVVMTAPEGTGIIAGGPMRAVFEMLVGKDVVSKSTGSQLPYNMIRATLDGLRNE
SPRQVASRRGKVVADILPKRDDHPQIDGEQAPVSEEA

>BPARRH1 [Bacteria Proteobacteria Alphaproteobacteria Rhodobacterales Hyphomonadaceae Hyphomonas]
Bacteria;Proteobacteria;Alphaproteobacteria;Rhodobacterales;Hyphomonadaceae;Hyphomonas; gi|499967091|ref|WP_011647809.1| MULTISPECIES: 30S
ribosomal protein S5 [Hyphomonas]
MAEEREGKRRRRDRENPRDRDESSELVDKLVGINRVAKTVKGGKRFGFAALVVVDQKGRAGFGKAKEVPEAIRKA
TEEAKRNLVRIPLREGRTLHHDGNGRHGAGKVVLAAPPPTGVIAGGPMRAVMEVLGVQDVVAKSLGSSNPYNMVRATFD
ALKGQANPRTVASKRGLKVQDIVGRRDTGASEAGMADSVN

>BPARRD1 [Bacteria Proteobacteria Alphaproteobacteria Rhodobacterales Rhodobacteraceae Dinoroseobacter]
Bacteria;Proteobacteria;Alphaproteobacteria;Rhodobacterales;Rhodobacteraceae;Dinoroseobacter; gi|501127973|ref|WP_012176981.1| 30S ribosomal
protein S5 [Dinoroseobacter shibae]
MAERDNRGRDRDRDETPEFADRLVAINRVSKTVKGGKRFGFAALVVVDQGRVGFHGKAKEVPEAIRKATEQAKRQM
IRVPLREGRTLHHDIEGRHGAGKVMRTAPQGTGIIAGGPMRAVFEMLVQDVVAKSIGSQNPYNMIRATLDGLRKETSP
RMVAQRRGKVSILKKGEPAAAAEPAEA

>BPARB4 [Bacteria Proteobacteria Alphaproteobacteria Rhizobiales Bradyrhizobiaceae Rhodopseudomonas]
Bacteria;Proteobacteria;Alphaproteobacteria;Rhizobiales;Bradyrhizobiaceae;Rhodopseudomonas; gi|499822833|ref|WP_011503567.1| MULTISPECIES: 30S
ribosomal protein S5 [Rhodopseudomonas]
MAAERERGGRRSREREERDSEFVDKLVHINRVAKVVKGGKRFGFAALVVVDQKGRVGFHGKAREVPEAIRKATEAAK
RNLTRVALREGRTLHHDIAGRHGAGRVYLAAPAGTGIIAGGPMRAVFETLGIQDVVAKSIGSSNPYNMVRATFDALKHL
DSPRAVAARRNIKVSTLQARRVGGDAEVAE

>BPARBR5 [Bacteria Proteobacteria Alphaproteobacteria Rhizobiales Bradyrhizobiaceae Rhodopseudomonas] Bacteria;Proteobacteria;Alphaproteobacteria;Rhizobiales;Bradyrhizobiaceae;Rhodopseudomonas; gi|499760468|ref|WP_011441202.1| 30S ribosomal protein S5 [Rhodopseudomonas palustris]
MAAERERGGRRSRDREERDSEFVDKLVHINRVAKVVKGGKRFGFAALVVIGDQKGRVGFHGKAREVPEAIRKATEAAK
RNLTRVALREGRTLHHDHDIAGRHRGAGRVYLAAPAGTGIIAGGPMRAVFETLGIQDVVAKSIGSSNPYNMVRATFDALKHV
DSPRSVAARRNIKVSTLQARRVGGDAEIVAE

>BPARHM1 [Bacteria Proteobacteria Alphaproteobacteria Rhodobacterales Hyphomonadaceae Maricaulis] Bacteria;Proteobacteria;Alphaproteobacteria;Rhodobacterales;Hyphomonadaceae;Maricaulis; gi|499962998|ref|WP_011643716.1| 30S ribosomal protein S5 [Maricaulis maris]
MAHQNEQRGGDRGRGRGRDRDQERDDELVDKLVHINRVAKTVKGRNRFQFAALVVVGDQKGRVGFQGGKAREVPEAI
RKATDEAKKNMIRVPLREGRTLHHDGKGRHGAGKVVLAAPPGTGVIIAGGPMRAVLETLGIQDVVAKSVGTSNPYNMIRA
TMDALKLQSSPRSVAAKRGKLVGDLVNRDDGASSPEAIEA

>BPARRR5 [Bacteria Proteobacteria Alphaproteobacteria Rhodobacterales Rhodobacteraceae Ruegeria] Bacteria;Proteobacteria;Alphaproteobacteria;Rhodobacterales;Rhodobacteraceae;Ruegeria; gi|496468029|ref|WP_009176874.1| MULTISPECIES: 30S ribosomal protein S5 [Ruegeria]
MAERENRRGNRRDREETPEFADRLVAINRVSKTVKGGKRFGFAALVVVGDQKGRVGFQGGKAKEVPEAIRKATEQAKRQM
IRVPLKEGRTLHHDHMYGRHGAGKVMRTAPEGTGIIAGGPMRAVFEMLGKIDVVSQSVGSQNPYNMIRATIDGLKKEQSP
RSVAQRRGKKVADILPKRDEAPAEAEA

>BPARRR6 [Bacteria Proteobacteria Alphaproteobacteria Rhodobacterales Rhodobacteraceae Ruegeria] Bacteria;Proteobacteria;Alphaproteobacteria;Rhodobacterales;Rhodobacteraceae;Ruegeria; gi|499357829|ref|WP_011046261.1| 30S ribosomal protein S5 [Ruegeria pomeroyi]
MAERENRRGNRRDREEAPEFADRLVAINRVSKTVKGGKRFGFAALVVVGDQKGRVGFQGGKAKEVPEAIRKATEQAKRQ
MIRVQLREGRTLHHDIEGRHGAGKVVMAAPEGTGIIAGGPMRAVFEMLGKIDVVSQSVGSANPNMIRATIDGLKKEQS
PRVAQRRGKKVADILPKRDEAPAEAEAEA

>BPARAG2 [Bacteria Proteobacteria Alphaproteobacteria Rhodospirillales Acetobacteraceae Gluconobacter] Bacteria;Proteobacteria;Alphaproteobacteria;Rhodospirillales;Acetobacteraceae;Gluconobacter; gi|499571166|ref|WP_011251949.1| MULTISPECIES: 30S ribosomal protein S5 [Gluconobacter]
MAREPREGRRGGREREQDLDLVLTINRVAKVVKGGRRFQFAALVVVGDQKGRVGFQGGKAKEVPEAIRKATERAKRT
MIRVPMKEGRTLHHDFTFGHYGAGKVVRAAEAGTGIIAGGPMRAVFESLGVNDVVAKSLGTRNPNHMIKATFAALEKASS
PRHVASRRGKKAELFGKREQGQTEAEVTNG

>BPARRP1 [Bacteria Proteobacteria Alphaproteobacteria Rhodobacterales Rhodobacteraceae Paracoccus] Bacteria;Proteobacteria;Alphaproteobacteria;Rhodobacterales;Rhodobacteraceae;Paracoccus; gi|500071103|ref|WP_011747116.1| 30S ribosomal protein S5 [Paracoccus denitrificans]
MAEREHRRGRREEREETPEFADRLVAINRVSKTVKGGKRFGFAALVVVGDQKGRVGFQGGKAKEVPEAIRKATEQAKRSL
IRVPLRDGRTLHHDIEGRHGAGKVMRTAVPGTGIAGGPMRAVFEMLGQDVVAKSQGSQNPYNMIRATIDGLKKEASP
RNVAQRRGKKVAEILPSNDKPAEAEAEA

>BPARRR7 [Bacteria Proteobacteria Alphaproteobacteria Rhodobacterales Rhodobacteraceae Roseobacter] Bacteria;Proteobacteria;Alphaproteobacteria;Rhodobacterales;Rhodobacteraceae;Roseobacter; gi|499886952|ref|WP_011567686.1| 30S ribosomal protein S5 [Roseobacter denitrificans]
MAERDNNRRGNRRDRDEAPEFADRLVAINRVSKTVKGGKRFGFAALVVVGDQKGRVGFQGGKAKEVPEAIRKATEQAKRQ
MIRVQLREGRTLHHDMEGRHGAGKVMMSAPEGTGIIAGGPMRAVFEMLGKIDVVSQSVGSQNPYNMIRATMDGLRKESS
PRVAQRRGKKVADILPKVDAAPAEAEAEA

>BPARRR8 [Bacteria Proteobacteria Alphaproteobacteria Rhodospirillales Rhodospirillaceae Rhodospirillum] Bacteria;Proteobacteria;Alphaproteobacteria;Rhodospirillales;Rhodospirillaceae;Rhodospirillum; gi|499709687|ref|WP_011390421.1| 30S ribosomal protein S5 [Rhodospirillum rubrum]
MARTPNTDRRQRGGDQNRNSPRSDERDEEFLDKLVHINRVAKVVKGGRRFQFAALVVVGDQKGRVGFQGGKAREVPEAI
RKATDQAKRTMIKVPPLREGRTLHHDNTHGFGAGVWTLRSAPAGTGIIISGGPMRAVFETMGVQDVVGKCTGTTNPYNMIKA
TFDALVNSQAPRQVAARRGKVGDIVARRDGAAGATE

>BPASSZ1 [Bacteria Proteobacteria Alphaproteobacteria Sphingomonadales Sphingomonadaceae Zymomonas] Bacteria;Proteobacteria;Alphaproteobacteria;Sphingomonadales;Sphingomonadaceae;Zymomonas; gi|499559657|ref|WP_011240440.1| 30S ribosomal protein S5 [Zymomonas mobilis]

MANESEIQKTENAEVANAANGTNPNNERRGRGGRGRGRDRDGRRRDRRNEEAGEELIEKLVHINRVSKTVKGGKRF
FAALVVVDGDKGRVGFHGKAREVPEAISKATAAAKTMIRVPLREGRTLHHDGRGFRGAGLVYLRAPSPTGIIAGGPM
RAIFESLGVADVTKSVGTSNPNMIRATFEALNDQKSPKAVARQRGKRIADLFRGGASAAKAEAEAAVVE
>BPARAA1 [Bacteria Proteobacteria Alphaproteobacteria Rhodospirillales Acetobacteraceae Acidiphilium]
Bacteria;Proteobacteria;Alphaproteobacteria;Rhodospirillales;Acetobacteraceae;Acidiphilium; gi|49466246|ref|WP_007424190.1| MULTISPECIES: 30S
ribosomal protein S5 [Acidiphilium]
MAREPREGRERERDRDDELIDKLVITINRAKVVKGGRRFAALVVVDQKGRVGFAGKAKEVPEAIRKATERAKR
GMIRVPMKEGRTLHHDIEGRFGAGHVVLRAAPPGTGIIAGGPMRAVFETLIGDQVVAKSLGSRNPHNMVKATVAALSQCA
SPRSVASRRGKVVADLFGPKREKPADV
>BPASEE1 [Bacteria Proteobacteria Alphaproteobacteria Sphingomonadales Erythrobacteraceae Erythrobacter]
Bacteria;Proteobacteria;Alphaproteobacteria;Sphingomonadales;Erythrobacteraceae;Erythrobacter; gi|122544181|sp|Q2N9C8.1|RS5_ERYLH RecName:
Full=30S ribosomal protein S5
MMADEKKTPEAETATPAVAVEDALKAETETLEAQKAKAEETPAVAETPSEAAANQSAQAEGQPRERGGHDRGGR
GGRGGNDRGRGGRDRNRGRRREEEDDIIIEKLVHINRVSKTVKGGKRFGFAALVVVDGQGRVGFAGKAKEVPEAIQ
KATAAARKMIRVALKEGRTLHHDGNGRFGAGKVTVRTAPPGTGIIAGGPMRAVFESLGVADVTKSVGTSNPNMIRAT
FDALQDQTSKSAQRRGKVVADLGRGGASEAEAEADAAIAE
>BPARRR9 [Bacteria Proteobacteria Alphaproteobacteria Rhodobacterales Rhodobacteraceae Rhodobacter]
Bacteria;Proteobacteria;Alphaproteobacteria;Rhodobacterales;Rhodobacteraceae;Rhodobacter; gi|500249131|ref|WP_011909473.1| 30S ribosomal protein
S5 [Rhodobacter sphaeroides]
MAERENRRDRDRSREETPEFADRLVAINRVSKTVKGGKRFGFAALVVVDQGRVGFAGKAKEVPEAIRKATEQAKR
QMIRVALRDGRTLHHDQEGRHGAGKVVRAAVPGTGIIAGGPMRAVFEMLIGDQVVAKSLGSRNPNMIRATMDGLKRES
SPRQVAQRRGKVAEILKPEAEVAAE
>BPARRR10 [Bacteria Proteobacteria Alphaproteobacteria Rhodobacterales Rhodobacteraceae Rhodobacter]
Bacteria;Proteobacteria;Alphaproteobacteria;Rhodobacterales;Rhodobacteraceae;Rhodobacter; gi|488810116|ref|WP_002722522.1| MULTISPECIES: 30S
ribosomal protein S5 [Rhodobacter]
MAERENRRDRDRSREETPEFADRLVAINRVSKTVKGGKRFGFAALVVVDQGRVGFAGKAKEVPEAIRKATEQAKR
QMIRVALRDGRTLHHDQEGRHGAGKVVRAAVPGTGIIAGGPMRAVFEMLIGDQVVAKSLGSRNPNMIRATMDGLKRES
SPRQVAQRRGKVVADILKPEAEVAAE
>BPARRM1 [Bacteria Proteobacteria Alphaproteobacteria Rhodospirillales Rhodospirillaceae Magnetospirillum]
Bacteria;Proteobacteria;Alphaproteobacteria;Rhodospirillales;Rhodospirillaceae;Magnetospirillum; gi|497554352|ref|WP_009868550.1| MULTISPECIES: 30S
ribosomal protein S5 [Magnetospirillum]
MARTPSSDRPERGGRGERDRPNRGRGAEQTPREERESEFVDKLVHINRAKVVKGGRRFAALVVVDGAKGRVGCSS
GKAREVPEAIRKATEQAKRNMKIALREGRTLHHDAYGHGAGRVILRAAPAGTGIIAGGPMRAVFETMGVQDVVAKCLG
TSNPHNMKATFDALINLASPRHVAARKGKVGVEIIGRRDGAAGV
>BPARAG3 [Bacteria Proteobacteria Alphaproteobacteria Rhodospirillales Acetobacteraceae Gluconacetobacter]
Bacteria;Proteobacteria;Alphaproteobacteria;Rhodospirillales;Acetobacteraceae;Gluconacetobacter; gi|501184853|ref|WP_012227931.1| 30S ribosomal
protein S5 [Gluconacetobacter diazotrophicus]
MAREPREGGRGDREREGDDLVDKLVITINRAKVVKGGRRFAALVVVDQKGRVGFAGKAREVPEAIRKATDRAKR
AMIRVPMKEGRTLHHDVAGHFGAGKVVLRADAGTGIIAGGPMRAVFESLGINDDVVAKSLGTRNPHNMVKATFAALERCA
SPRTVANRRGKVVSDILGRRDVAGTGEAAADV
>BPARRR11 [Bacteria Proteobacteria Alphaproteobacteria Rickettsiales Rickettsiaceae Rickettsiae]
Bacteria;Proteobacteria;Alphaproteobacteria;Rickettsiales;Rickettsiaceae;Rickettsiae;Rickettsia;spotted fever group; gi|504161968|ref|WP_014365550.1|
MULTISPECIES: 30S ribosomal protein S5 [spotted fever group]
MSKVKNDETLEVLVDVNRVTKVVGKRRFAFSAYVVVDKAGRVGAGHGKAKEVNEARGKAKQAAKRRMMKVPLYQNR
TIHHDVVGKSGAAKILRRKAGTGIIAGGSMRAIFDSLGVHDIVAKSIGSTNVYAMISATFDALNKLASPKSIAMRRDK
KVNEIFVKSSDIQVNE
>BPARRR12 [Bacteria Proteobacteria Alphaproteobacteria Rickettsiales Rickettsiaceae Rickettsiae]
Bacteria;Proteobacteria;Alphaproteobacteria;Rickettsiales;Rickettsiaceae;Rickettsiae;Rickettsia;spotted fever group; gi|501100019|ref|WP_012149888.1| 30S
ribosomal protein S5 [Rickettsia akari]
MSKVKNDETLEVLVDVNRVTKVVGKRRFAFSAYVVVDKAGRVGAGHGKAKEVNEARGKAKQAAKRRMMKVPLYQNR
TIHHDVVGKSGAAKILRRKAGTGIIAGGSMRAIFDSLGVHDIVAKSIGSTNVYAMISATFDALNKLASPKSIAMRRDK

KVNEISIKSSDIQVNE

>BPARRR13 [Bacteria Proteobacteria Alphaproteobacteria Rickettsiales Rickettsiaceae Rickettsieae]
 Bacteria;Proteobacteria;Alphaproteobacteria;Rickettsiales;Rickettsiaceae;Rickettsieae;Rickettsia;spotted fever group; gi|50222560|ref|WP_012736939.1| 30S ribosomal protein S5 [Rickettsia peacockii]
 MSKVKKNDSEILSEVLVDVNRVTKVVKGGRSFAFSAYVVVGDKAGRVGAGHGKAKEVNEARGKAKQAACKRMMKVPLYQNR
 TIHHDVVGKSGAAKILRRKAGTGVIIAGGSMRAIFDSLGVHDIVAKSIGSTNVYAMISATFDALNKLASPKSIAMRRDK
 KVNEISVKSADIQVNE

>BPARRR14 [Bacteria Proteobacteria Alphaproteobacteria Rickettsiales Rickettsiaceae Rickettsieae]
 Bacteria;Proteobacteria;Alphaproteobacteria;Rickettsiales;Rickettsiaceae;Rickettsieae;Rickettsia;spotted fever group; gi|499284519|ref|WP_010977579.1| 30S ribosomal protein S5 [Rickettsia conorii]
 MSKVKKNDSEILSEVLVDVNRVTKVVKGGRSFAFSAYVVVGDKAGRVGAGHGKAKEVNEARGKAKQAACKRMMKVPLYQNR
 TIHHDVVGKSGAAKILRRKAGTGIIAGGSMRAIFDSLGVHDIVAKSIGSTNVYAMISATFDALNKLASPKSIAMRRDK
 KVNEISVKSADIEIQVNE

>BPARRR15 [Bacteria Proteobacteria Alphaproteobacteria Rickettsiales Rickettsiaceae Rickettsieae]
 Bacteria;Proteobacteria;Alphaproteobacteria;Rickettsiales;Rickettsiaceae;Rickettsieae;Rickettsia;spotted fever group; gi|491139410|ref|WP_004997826.1| MULTISPECIES: 30S ribosomal protein S5 [spotted fever group]
 MSKVKKNDSEILSEVLVDVNRVTKVVKGGRSFAFSAYVVVGDKAGRVGAGHGKAKEVNEARGKAKQAACKRMMKVPLYQNR
 TIHHDVVGKSGAAKILRRKAGTGIIAGGSMRAIFDSLGVHDIVAKSIGSTNVYAMISATFDALNKLASPKSIAMRRDK
 KVNEISVKSADIQVNE

>BPARRR16 [Bacteria Proteobacteria Alphaproteobacteria Rickettsiales Rickettsiaceae Rickettsieae]
 Bacteria;Proteobacteria;Alphaproteobacteria;Rickettsiales;Rickettsiaceae;Rickettsieae;Rickettsia;spotted fever group; gi|499589872|ref|WP_011270644.1| MULTISPECIES: 30S ribosomal protein S5 [spotted fever group]
 MSKVKKNEETLSEVLVDVNRVTKVVKGGRRFAFSAYVVVGDKAGRVGAGHGKAKEVNEARGKAKQAACKRMMKVPLYQNR
 TIHHDVVGKSGAAKILRRKAGTGVIIAGGSMRAIFDSLGVHDIVAKSIGSTNVYAMISATFDALNKLASPKSIAMRRDK
 KVNEISVKSADIQVNE

>BPARRR17 [Bacteria Proteobacteria Alphaproteobacteria Rickettsiales Rickettsiaceae Rickettsieae]
 Bacteria;Proteobacteria;Alphaproteobacteria;Rickettsiales;Rickettsiaceae;Rickettsieae;Rickettsia;spotted fever group; gi|501101304|ref|WP_012151140.1| 30S ribosomal protein S5 [Rickettsia rickettsii]
 MSKVKKNDSEILSEVLVDVNRVTKVVKGGRSFAFSAYVVVGDKAGRVGAGHGKAKEVNEARGKAKQAACKRMMKVPLYQNR
 TIHHDVVGKSGAAKILRRKAGTGVIIAGGSMRAIFDSLGIHDVAKSIGSTNVYAMISATFDALNKLASPKSIAMRRDK
 KVNEISVKSADIQVNE

>BPASSS1 [Bacteria Proteobacteria Alphaproteobacteria Sphingomonadales Sphingomonadaceae Sphingopyxis]
 Bacteria;Proteobacteria;Alphaproteobacteria;Sphingomonadales;Sphingomonadaceae;Sphingopyxis; gi|122984737|sp|Q1GPB6.1|RS5_SPHAL RecName: Full=30S ribosomal protein S5
 MADENSTGPGNQPEATLAPEATNAEVTSQLAPRRGRGGGGRDGGRRDGGRRDRRPRDEDGGEELIEKLVHINRVSK
 TVKGGKRFGFAALVVVDGKGRAGFGHGKAREVPEISKATAAAKKAMIRVPLRDGRTLHHDGRGHFGAGNVTILRSAPAG
 TGIIAGGPMRAVFESLGVADVVTKSVGTSNPYNMIRATFEALGEQTSKPSVAQRRGKVKVSDLIKRGASDRAAEAEAAAV
 TE

>BPARRR18 [Bacteria Proteobacteria Alphaproteobacteria Rickettsiales Rickettsiaceae Rickettsieae]
 Bacteria;Proteobacteria;Alphaproteobacteria;Rickettsiales;Rickettsiaceae;Rickettsieae;Rickettsia;typhus group; gi|499504436|ref|WP_011191076.1| 30S ribosomal protein S5 [Rickettsia typhi]
 MSKVKKNEEALSEVLVDVNRVTKVVKGGRRFAFSAYVVVGDKAGRVGAGHGKAKEVNEARGKAKQAACKRMMKVPLYQNR
 TIHHDVVGKSGAAKILRRKAGTGVIIAGGSMRAIFDSLGVHDIVAKSIGSTNVYAMISATFDALNKLASPKSIARRGK
 KVHEISVKSADIEIQVNE

>BPARRR19 [Bacteria Proteobacteria Alphaproteobacteria Rickettsiales Rickettsiaceae Rickettsieae]
 Bacteria;Proteobacteria;Alphaproteobacteria;Rickettsiales;Rickettsiaceae;Rickettsieae;Rickettsia;typhus group; gi|490733890|ref|WP_004596227.1| 30S ribosomal protein S5 [Rickettsia prowazekii]
 MSKVKKNEEALSEVLVDVNRVTKVVKGGRRFAFSAYVVVGDKAGRVGAGHGKAKEVNEARGKAKQAACKRMMKVPLYQNR
 TIHHDVVGKSGAAKILRRKAGTGVIIAGGSMRAIFDSLGVHDIVAKSIGSTNVYAMISATFDALNKLASPKSIARRDK
 KVHEISVKSADIEIQVNE

>BPASS2 [Bacteria Proteobacteria Alphaproteobacteria Sphingomonadales Sphingomonadaceae Sphingomonas] Bacteria;Proteobacteria;Alphaproteobacteria;Sphingomonadales;Sphingomonadaceae;Sphingomonas; gi|500516432|ref|WP_011952179.1| MULTISPECIES: 30S ribosomal protein S5 [Sphingomonas]
MADETEIQAAAPAEAPGAEGQGERRGRGRRGNDRGGDRGRDRGRDRRRDSEEQEELIEKLVHINRVSKTVK
GGKRFGFAALVVVDGKGRAGFGHGKAREVPEAISKATAAAKAMVVRVPLREGRTLHHDGNGHFGAGRVTVRTAPPGTGI
IAGGPMRAIFESLGVADVTKSVGTSNPYNMIRATFEALDKDQTSKPSVSRQRGKIADLLGRGGASAPVAEAEAAATE

>BPARRR20 [Bacteria Proteobacteria Alphaproteobacteria Rickettsiales Rickettsiaceae Rickettsieae] Bacteria;Proteobacteria;Alphaproteobacteria;Rickettsiales;Rickettsiaceae;Rickettsieae;Rickettsia;typhus group; gi|501099085|ref|WP_012148977.1| 30S ribosomal protein S5 [Rickettsia canadensis]
MSKVKKNEETLSEVLVDVNRITKVVKGGRRFAFSACMVVGDKAGRVGAGHGKAKEVNEARGKAKQAAKKRMMKVPLYQNR
TIHHDVIGKSGAAKVLRRKAGTGVIAAGSMRAIFHSLGVHDIKAKSIGSTNVYAMISATFDALNKLASPKSIAMRRDK
RVNEISVKSVDV

>BPGAAA1 [Bacteria Proteobacteria Gammaproteobacteria Acidithiobacillales Acidithiobacillaceae Acidithiobacillus] Bacteria;Proteobacteria;Gammaproteobacteria;Acidithiobacillales;Acidithiobacillaceae;Acidithiobacillus; gi|501529610|ref|WP_012536093.1| MULTISPECIES: 30S ribosomal protein S5 [Bacteria]
MARENDRTPNDGMQEKLHINRVSKVVGGRQFGFAALMVVDGDKVGFGRGKAKEVPAGIQKATDQARRWMTSIPLM
RGGTIPYVEGRHGAARVMLRPAPEGSGVIAGGAMRAVCEAVGLRNVVAKSLGNNPINVVRATFDADFKLISPQAIAMK
RGKSLKEIRGQGGHDE

>BPDDDD1 [Bacteria Proteobacteria Deltaproteobacteria Desulfovibrionales Desulfovibrionaceae Desulfovibrio] Bacteria;Proteobacteria;Deltaproteobacteria;Desulfovibrionales;Desulfovibrionaceae;Desulfovibrio; gi|506331364|ref|WP_015851083.1| 30S ribosomal protein S5 [Desulfovibrio salexigenis]
MEQNDLGLIEKIVLNRVAKVVKGGRRFSFALVVVDGKQVGFGLGKANEVPEAIRKASEKARKEMISVPLDGLTPY
EVLGRYGAGRVMLKPKASKGTGIAGGPVRAVLEVVGVHDILTKAIGTNNPHNVLRAIAGLASLRSADVEVQLRGKVVV
PRK

>BPASSN1 [Bacteria Proteobacteria Alphaproteobacteria Sphingomonadales Sphingomonadaceae Novosphingobium] Bacteria;Proteobacteria;Alphaproteobacteria;Sphingomonadales;Sphingomonadaceae;Novosphingobium; gi|499764190|ref|WP_011444924.1| MULTISPECIES: 30S ribosomal protein S5 [Novosphingobium]
MADETNLEGVAAEATGGEPQREGRRGRGRRGNDRGGRRGRRDRRRGRGNNDDEEGEELIEKLVHINRVSKTVKGG
KRFGFAALVVVDGKGRAGFGKAREVPEAITKATAAAKRAMVVRVPLKEGRTLHHDGKGRFGAGKVNVRAPAGTGIIA
GGPMRAVFEVLGSDVTKSVGTSNPYNMIRATFDALDQTSKPSVAQRRGKVVADLLGRGGASQVEAEAAEAAEIAE

>BPGEEB1 [Bacteria Proteobacteria Gammaproteobacteria Enterobacteriales Enterobacteriaceae Buchnera] Bacteria;Proteobacteria;Gammaproteobacteria;Enterobacteriales;Enterobacteriaceae;Buchnera; gi|499366419|ref|WP_011053997.1| 30S ribosomal protein S5 [Buchnera aphidicola]
MANIEKNNSELQEKLITVNRVSKTVKGGRRFSFALTIVVNGEGRVGFYGGKAREVPAAIQKAMEKARRNMVTIPLVNK
TLQHSLKGSHTGSNIFMKPASDGTGIAGGAMRAVLEVAGIHNVLAKTYGSTNPINVVRATMNGLTNMKSPEMIAAKRNK
LIEDILG

>BPARRR21 [Bacteria Proteobacteria Alphaproteobacteria Rickettsiales Rickettsiaceae Rickettsieae] Bacteria;Proteobacteria;Alphaproteobacteria;Rickettsiales;Rickettsiaceae;Rickettsieae;Orientia; gi|500486818|ref|WP_011944423.1| 30S ribosomal protein S5 [Orientia tsutsugamushi]
MFKEIIVVHLVNVNRVTKVVKGGRRFSFACVVLGNKAGKVGYGHGKAKEVGEARLKALQNAKKYMIEVPLYKGRTHY
DVYGKSGAAKVMMLRRARAGTGIAGGAMRFIFDSLGIQDVVAKSFGSSNSYLMIAATLNALKQLETPRVIAERRGLRLNE
LSVTIYKTFHEG

>BFCCSS1 [Bacteria Firmicutes Clostridia Clostridiales Syntrophomonadaceae Syntrophomonas] Bacteria;Firmicutes;Clostridia;Clostridiales;Syntrophomonadaceae;Syntrophomonas; gi|122317299|sp|Q0AUJ7.1|RS5_SYNWW RecName: Full=30S ribosomal protein S5
MIDKEQAAPEMIEKVVVIRVAKVVKGGRRFSFALVVVDGEGKVGTKGKATEVPEAIRKAIENAKKNMNSVPLIDGR
TIPHEILGIFGAGQVLLKPASEGTVIAGGPVRAVLEAAGIKDILTKSLGSDNATNIVHATMEGLRSLKRVEDVARQRGK
TIDEIMN

>BPDDPP1 [Bacteria Proteobacteria Deltaproteobacteria Desulfuromonadales Pelobacteraceae Pelobacter] Bacteria;Proteobacteria;Deltaproteobacteria;Desulfuromonadales;Pelobacteraceae;Pelobacter; gi|499659686|ref|WP_011340420.1| 30S ribosomal protein S5 [Pelobacter carbinolicus]

MLRIDPNELELTDREVHINRCAKVVKGGRRFSFALSALVVVDGQGVYGHGKAKEVPEAIRKGVQAKKNLIRVPLKDRS
 IPFDVIGKFGAGRLLKPASAGTGVIAAGAVRAVLEVSGVDILSKCIGSNPHNVVRATIDALSRLKSAEELRALRGAD
 TEE

>BFCNN1 [Bacteria Firmicutes Clostridia Natranaerobiales Natranaerobiaceae Natranaerobius]
 Bacteria;Firmicutes;Clostridia;Natranaerobiales;Natranaerobiaceae;Natranaerobius; gi|501422029|ref|WP_012446699.1| 30S ribosomal protein S5
 [Natranaerobius thermophilus]

MRRKDFDPNKLDSLSEKVEINRVTKVVKGGRRFSFALSALVVVDGQGVYGHGKAKEVPEAIRKGVQAKKNLIRVPLKDRS
 TTPHIEVGRDGGGQVLLKPAYEGTGVIAAGPVRAVLEQAGIRDILTSLGSSNNAKSMVNATIAGLNQLKTAEQVRLRG
 ISVKDLQG

>BPGEEB2 [Bacteria Proteobacteria Gammaproteobacteria Enterobacterales Enterobacteriaceae Buchnera]
 Bacteria;Proteobacteria;Gammaproteobacteria;Enterobacterales;Enterobacteriaceae;Buchnera; gi|497560274|ref|WP_009874458.1| 30S ribosomal protein S5
 [Buchnera aphidicola]

MANIEKKNHNDLQEKLITVNRVSKTVKGGRRFSFALTALVVVDGQGVYGHGKAKEVPAIQKAMEQARRNMITIPLVVK
 TLQHSKLSGSHTSNVFMKPSADGTGIIAGGAMRAVLEVAGIHNVLAKTYGSTNPINVVVRATMGLVNMKSPEMIAAKRNK
 RIKDILG

>BPBBB1 [Bacteria Proteobacteria Betaproteobacteria Burkholderiales Burkholderiaceae Ralstonia]
 Bacteria;Proteobacteria;Betaproteobacteria;Burkholderiales;Burkholderiaceae;Ralstonia; gi|499312131|ref|WP_011002906.1| 30S ribosomal protein S5
 [Ralstonia solanacearum]

MAKIQPKVQDERDDGLREKMIANVRVTKVVKGGRRILGFAALTVVDGQGVYGHGKAKEVPAVQKAMDEARRKMVKV
 PLKNGTLQHEVIGKHGAAKVLMAPAKEGTGVIAGGPMRAIFEVMGVTVNIVTKSHGSTNPYNMVRATLDGLRKMSTPAEIA
 AKRGKSVVEILG

>BPBBB2 [Bacteria Proteobacteria Betaproteobacteria Burkholderiales Burkholderiaceae Ralstonia]
 Bacteria;Proteobacteria;Betaproteobacteria;Burkholderiales;Burkholderiaceae;Ralstonia; gi|490772271|ref|WP_004634493.1| MULTISPECIES: 30S ribosomal
 protein S5 [Ralstonia]

MAKIQPKVQDERDDGLREKMIANVRVTKVVKGGRRILGFAALTVVDGQGVYGHGKAKEVPAVQKAMDEARRKMVKV
 PLKNGTLQHEVIGKHGAAKVQMMPAKDGTGVIAGGPMRAIFEVMGVTVNIVTKSHGSTNPYNMVRATLDGLRKMSTPAEIA
 AKRGKSVVEILG

>BPARAW1 [Bacteria Proteobacteria Alphaproteobacteria Rickettsiales Anaplasmataceae Wolbachieae]
 Bacteria;Proteobacteria;Alphaproteobacteria;Rickettsiales;Anaplasmataceae;Wolbachieae;Wolbachia; gi|499575740|ref|WP_011256523.1| 30S ribosomal
 protein S5 [Wolbachia endosymbiont of Brugia malayi]

MTIKNFQNNNDWSELLSVRRVTVTKGGRRFSFALSALVVVDGQGVYGHGKAKEVPAVQKAMDEARRKMVKV
 RTLHHDIAKAFCSGEIVLRAARAGTGIIAGGAIRSVFEVLGIKDVVAKSTRSNNPHNVICAVFKAFDNMLSPRQVASKRG
 KKISEVVGNR

>BPBBB3 [Bacteria Proteobacteria Betaproteobacteria Burkholderiales Burkholderiaceae Cupriavidus]
 Bacteria;Proteobacteria;Betaproteobacteria;Burkholderiales;Burkholderiaceae;Cupriavidus; gi|499618568|ref|WP_011299302.1| 30S ribosomal protein S5
 [Cupriavidus pinatubonensis]

MAKMQAKVQDERDDGLREKMIANVRVTKVVKGGRRILGFAALTVVDGQGVYGHGKAKEVPAVQKAMDEARRKMVKV
 SLKNGTLQHEVIGKHGAAKVLMMPAKEGTGVIAGGPMRAIFEVMGVTVNIVTKSHGSTNPYNMVRATLDGLQKMPSTPSEIA
 AKRGKSVEDILG

>BTMMM1 [Bacteria Tenericutes Mollicutes Mycoplasmataceae Mycoplasma] Bacteria;Tenericutes;Mollicutes;Mycoplasmataceae;Mycoplasma;
 gi|500502176|ref|WP_011949693.1| 30S ribosomal protein S5 [Mycoplasma agalactiae]

MADLENKTVKANVENKPGAQTSVSPKRTESGAKKQIWEKRSADSKDMPKKSVDRAKVNKRNTRFGEANNEFSEKVV
 NISRVTKVVKGGRRFSFALSALVVVDGQGVYGHGKAKEVPAIKKAVKDARNHLISVPIQNKITVPEIHAKFLASKVM
 LKPAPKGGIVASGTVRAVELAGYTDIYTKYGSRSKANIVRATLALQQLRTPAQIAEIRKDKVDKLLG

>BPBBB4 [Bacteria Proteobacteria Betaproteobacteria Burkholderiales Burkholderiaceae Cupriavidus]
 Bacteria;Proteobacteria;Betaproteobacteria;Burkholderiales;Burkholderiaceae;Cupriavidus; gi|493147187|ref|WP_006160449.1| MULTISPECIES: 30S
 ribosomal protein S5 [Cupriavidus]

MAKMQAKVQDERDDGLREKMIANVRVTKVVKGGRRILGFAALTVVDGQGVYGHGKAKEVPAVQKAMDEARRKMVKV
 SLKNGTLQHEVIGKHGAAKVLMMPAKEGTGVIAGGPMRAIFEVMGVTVNIVTKSHGSTNPYNMVRATLDGLQKMPSTPGEIA
 AKRGKSVEDILG

>BPARAW2 [Bacteria Proteobacteria Alphaproteobacteria Rickettsiales Anaplasmataceae Wolbachieae]
 Bacteria;Proteobacteria;Alphaproteobacteria;Rickettsiales;Anaplasmataceae;Wolbachieae;Wolbachia; gi|499265355|ref|WP_010962748.1| MULTISPECIES:
 30S ribosomal protein S5 [Wolbachia]
 MAIKNLQNNNDLSELLVSVRRVTTVTKGRRRFSFSLVVGDEKGRVCGIGKHAIEV AEARVKAVNAAKKSMIRVYLREG
 RTLHHDIAKAFCSGEIVLRTARAGTGIIAGGAIRSVFEVLGIKDVVAKSTRSNNPHNVICAVFKAFDSMLS PRQVASKRG
 KKISEIVGNR

>BPBBBB1 [Bacteria Proteobacteria Betaproteobacteria Burkholderiales Burkholderiaceae Burkholderia]
 Bacteria;Proteobacteria;Betaproteobacteria;Burkholderiales;Burkholderiaceae;Burkholderia;Burkholderia
 cepacia complex;
 gi|493528878|ref|WP_006482895.1| MULTISPECIES: 30S ribosomal protein S5 [Burkholderia]
 MAKMQAKVQADERDDGLREKISMISVNRVTKVVKGGRI LGFAALTVVGDGDGRIGMGKGAKEVPVAVQKAMEQARRNMFKV
 PLKNGTLQHEVHGKHGASAVLLAPAKAGTGV IAGGPMRAVFDVMGVQNVVAKSHGSTNPYNLVRATLDGLRKQSTPADIA
 AKRGKSV EILG

>BPBBBB2 [Bacteria Proteobacteria Betaproteobacteria Burkholderiales Burkholderiaceae Burkholderia]
 Bacteria;Proteobacteria;Betaproteobacteria;Burkholderiales;Burkholderiaceae;Burkholderia;Burkholderia
 cepacia complex;
 gi|493805102|ref|WP_006752931.1| MULTISPECIES: 30S ribosomal protein S5 [Burkholderia]
 MAKMQAKVQADERDDGLREKISMISVNRVTKVVKGGRI LGFAALTVVGDGDGRIGMGKGAKEVPVAVQKAMEQARRNMFKV
 PLKNGTLQHEVHGKHGASAVLLAPAKAGTGV IAGGPMRAVFDVMGVQNVVAKSHGSTNPYNLVRATLDGLRKQSTPADIA
 AKRGKSV EILG

>BPBBCD1 [Bacteria Proteobacteria Betaproteobacteria Burkholderiales Comamonadaceae Delftia]
 Bacteria;Proteobacteria;Betaproteobacteria;Burkholderiales;Comamonadaceae;Delftia; gi|501158502|ref|WP_012202971.1| MULTISPECIES: 30S ribosomal
 protein S5 [Delftia]
 MAKFSKPVQDEGRDDGLREKMIANVRVTKVVKGGRI LGFAALTVVGDGDGRVGMGKGSKEVPAAVQKAMEESRRNMIKV
 SLKHGTIHTVMGQHGASAVMLNPAPKGTGIIAGGPMRAVFEVLGITDIVAKSHGSSNPYNMVRATFDALKHSTTPSEVA
 AKRGKSV EIDFTA

>BPBBBB3 [Bacteria Proteobacteria Betaproteobacteria Burkholderiales Burkholderiaceae Burkholderia]
 Bacteria;Proteobacteria;Betaproteobacteria;Burkholderiales;Burkholderiaceae;Burkholderia; gi|494323647|ref|WP_007180127.1| MULTISPECIES: 30S
 ribosomal protein S5 [Burkholderiaceae]
 MAKMQAKVQADERDDGLREKISMISVNRVTKVVKGGRI LGFAALTVVGDGDGRIGMGKGAKEVPVAVQKAMEQARRNMFKV
 PLKNGTLQHEVHGKHGASAVLLAPAKDGTGVIAGGPMRAVFDVMGVQNVVAKSHGSTNPYNLVRATLDGLRKQSTPGDIA
 AKRGKSV EILG

>BFCCPD1 [Bacteria Firmicutes Clostridia Clostridiales Peptococcaceae Desulfotomaculum]
 Bacteria;Firmicutes;Clostridia;Clostridiales;Peptococcaceae;Desulfotomaculum; gi|500206400|ref|WP_011876621.1| 30S ribosomal protein S5
 [Desulfotomaculum reducens]
 MARIDASKLELTKVYINRAKVVKGGRRFSFSLVVGDNHGHVAGLGAEEVPEAIRKGI EDAKKNMIKVLNGTT
 ITHQIQGRFGAGKVLMPAAKGTGVIAGGPMRAILELAGVRDILTKSLGTNNANNMVNATMEGLKQLKTP EEARLRGKT
 VEELLG

>BPBBBB4 [Bacteria Proteobacteria Betaproteobacteria Burkholderiales Burkholderiaceae Burkholderia]
 Bacteria;Proteobacteria;Betaproteobacteria;Burkholderiales;Burkholderiaceae;Burkholderia; gi|495004590|ref|WP_007730604.1| MULTISPECIES: 30S
 ribosomal protein S5 [Burkholderiaceae]
 MAKMQAKVQADERDDGLREKISMISVNRVTKVVKGGRI LGFAALTVVGDGDGRVGMGKGAKEVPVAVQKAMEQARRNMFKV
 PLKNGTLQHEVHGKHGASTVLLAPAKDGTGVIAGGPMRAVFDVMGVQNVVAKSHGSTNPYNLVRATLDGLRKQSTPADIA
 AKRGKSV EILG

>BPBBBB5 [Bacteria Proteobacteria Betaproteobacteria Burkholderiales Burkholderiaceae Burkholderia]
 Bacteria;Proteobacteria;Betaproteobacteria;Burkholderiales;Burkholderiaceae;Burkholderia;pseudomallei group; gi|497574218|ref|WP_009888402.1|
 MULTISPECIES: 30S ribosomal protein S5 [Burkholderiaceae]
 MAKMQAKVQADERDDGLREKISMISVNRVTKVVKGGRI LGFAALTVVGDGDGRVGMGKGAKEVPVAVQKAMEQARRNMFKV
 PLKNGTLQHEVHGKHGASTVLLAPAKDGTGVIAGGPMRAVFDVMGVQNVVAKSHGSTNPYNLVRATLDGLRKQSTPGDIA
 AKRGKSV EILG

>BPBBBB6 [Bacteria Proteobacteria Betaproteobacteria Burkholderiales Burkholderiaceae Burkholderia]
 Bacteria;Proteobacteria;Betaproteobacteria;Burkholderiales;Burkholderiaceae;Burkholderia;pseudomallei group; gi|490302563|ref|WP_004197945.1|
 MULTISPECIES: 30S ribosomal protein S5 [Burkholderiaceae]

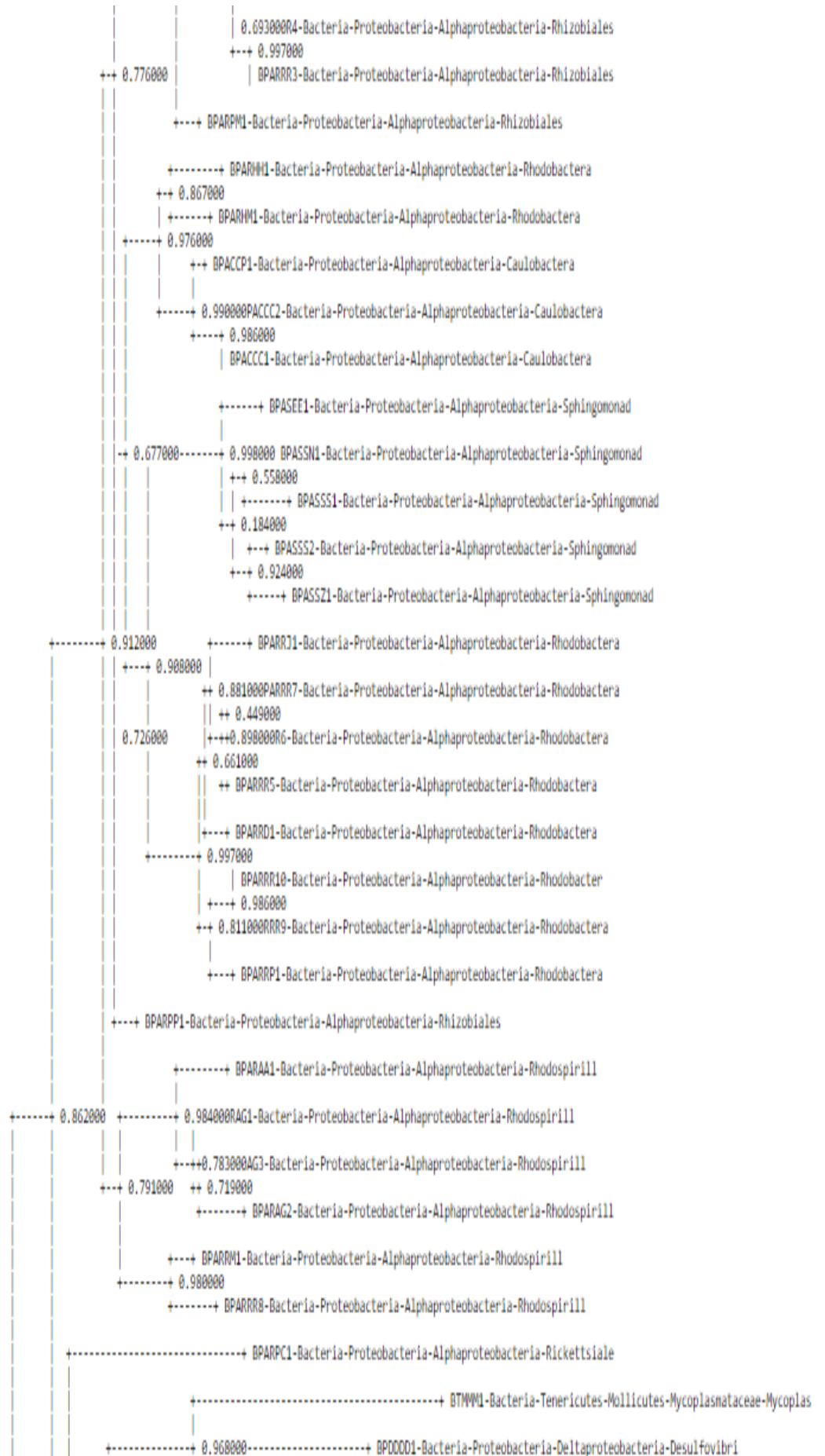
MAKMQAKVQADERDDGLREKMISVNRVTKVVKGGRIKFAALTVVGDGDGRVGMGKGAKEVPVAVQKAMEQARRNMFKV
 PLKNGTLQHEVHGKHGASTVLLAPAKDGTGVIAGGPMRAVFDVMGVQNVVAKSHGSTNPYNLVRATLDGLRKQSTPADIA
 AKRGKSVEEILG
 >BPGCCN1 [Bacteria Proteobacteria Gammaproteobacteria Chromatiales Chromatiaceae Nitrosococcus]
 Bacteria;Proteobacteria;Gammaproteobacteria;Chromatiales;Chromatiaceae;Nitrosococcus; gi|499650182|ref|WP_011330916.1| 30S ribosomal protein S5
 [Nitrosococcus oceani]
 MAQTDRRATGDGGLLEKLVGRRVAKVVKGGRRQFGFSALTVVGDGKGRVGFGRGKAREVPVAIQKAMENARKNMISVPLEG
 DTLQHPITARHGAAKVHMQPASEGTGIIAGGAMRAVFEVGVHNVLAKEIGSANPVNVVQATVKGLVQMSSPEIAAKRG
 KNLEEIVGGG
 >BPBBM1 [Bacteria Proteobacteria Betaproteobacteria Burkholderiales Methylibium]
 Bacteria;Proteobacteria;Betaproteobacteria;Burkholderiales;Methylibium; gi|500156328|ref|WP_011831000.1| MULTISPECIES: 30S ribosomal protein S5
 [Methylibium]
 MAKFPRIADDGRDDGLKEKMIQVNRVTKVVKGGRTMSFAALTVVGDGDGRVGMGKGAKEVPVAVTKAMDAARRDMVKV
 SLKNGTVHHNVTGEHGAALKVLLAPAAGTGIAGGPMRAVFEVGMGITDIVAKSLGSSNPYNMVRATFNALRRSTTPSEVA
 SKRGKSVEEIFN
 >BPBBP1 [Bacteria Proteobacteria Betaproteobacteria Burkholderiales Comamonadaceae Polaromonas]
 Bacteria;Proteobacteria;Betaproteobacteria;Burkholderiales;Comamonadaceae;Polaromonas; gi|500123761|ref|WP_011799766.1| MULTISPECIES: 30S
 ribosomal protein S5 [Polaromonas]
 MAKFQAKSQNDAPDDGLKEKMIQVNRVTKVVKGGRIKFAALTVVGDGDGRVGMGKGSKEVPAAVQKAMEEARRNMTKV
 SLKNGTLHHNVFGHHGAANVMAPAPKGTGIIAGGPMRAVFEVGMGITDIVAKSHGSSNPYNMVRATMDALKNSTTASDIA
 AKRGKSVEEIFG
 >BPGASS1 [Bacteria Proteobacteria Gammaproteobacteria Alteromonadales Shewanellaceae Shewanella]
 Bacteria;Proteobacteria;Gammaproteobacteria;Alteromonadales;Shewanellaceae;Shewanella; gi|500082338|ref|WP_011758351.1| 30S ribosomal protein S5
 [Shewanella amazonensis]
 MAKVEAQKQKDDLEKLIQVNRVSKVVKGGRIKFAALTVVGDGNGRVGYGYGKAREVPAAIQKAMEKARRNMTVELNE
 GTLHPVKGRHTGSRVYMQPASQGTGIIAGGAMRAVLEVAGVHNVLSKAYGSTNPINIVRATVDALTHMKSPAQVAAKRG
 LSVVEIRG
 >BPBBBC3 [Bacteria Proteobacteria Betaproteobacteria Burkholderiales Burkholderiaceae Cupriavidus]
 Bacteria;Proteobacteria;Betaproteobacteria;Burkholderiales;Burkholderiaceae;Cupriavidus; gi|495918372|ref|WP_008642951.1| MULTISPECIES: 30S
 ribosomal protein S5 [Cupriavidus]
 MAKMQAKVQADERDDGLREKMISVNRVTKVVKGGRIKFAALTVVGDGDGRIGMGKGAKEVPVAVQKAMDEARRKMVKV
 ALKNGTLQHEVVRHGAALKVLMMPAKEGTGVIAGGPMRAIFEVGMVTVNVTKSHGSTNPYNMVRATLDGLTKMSTPAEIA
 AKRGKSVEDILG
 >BPBBB7 [Bacteria Proteobacteria Betaproteobacteria Burkholderiales Burkholderiaceae Burkholderia]
 Bacteria;Proteobacteria;Betaproteobacteria;Burkholderiales;Burkholderiaceae;Burkholderia; gi|492938078|ref|WP_006052219.1| MULTISPECIES: 30S
 ribosomal protein S5 [Burkholderiaceae]
 MAKMQAKVQADERDDGLREKMISVNRVTKVVKGGRIKFAALTVVGDGDGRVGMGKGAKEVPVAVQKAMEQARRNMFKV
 PLKNGTLQHEVHGKHGASVMVLLAPAKDGTGVIAGGPMRAVFDVMGVQNVVAKSHGSTNPYNLVRATLDGLRKQSTPGDIA
 AKRGKSVEDILG
 >BFBBB1 [Bacteria Firmicutes Bacillales Bacillaceae Bacillus] Bacteria;Firmicutes;Bacillales;Bacillaceae;Bacillus; gi|499198794|ref|WP_010896334.1|
 MULTISPECIES: 30S ribosomal protein S5 [Bacillus]
 MRRIDPNTLEEEKVVAINRVAKVVKGGRRFRFAALVVVDKNGRVGFGMGKAQEVPEAIRKAVEDAKKNLIEVPIVGT
 IPHEIVGRFGAGRVLLKPASEGTVIAGGPRVAVLDLAGVDILSKSLGSSNPINMVRATVKGLQELKRAEDVAKLRGKT
 VEELG
 [Results end]

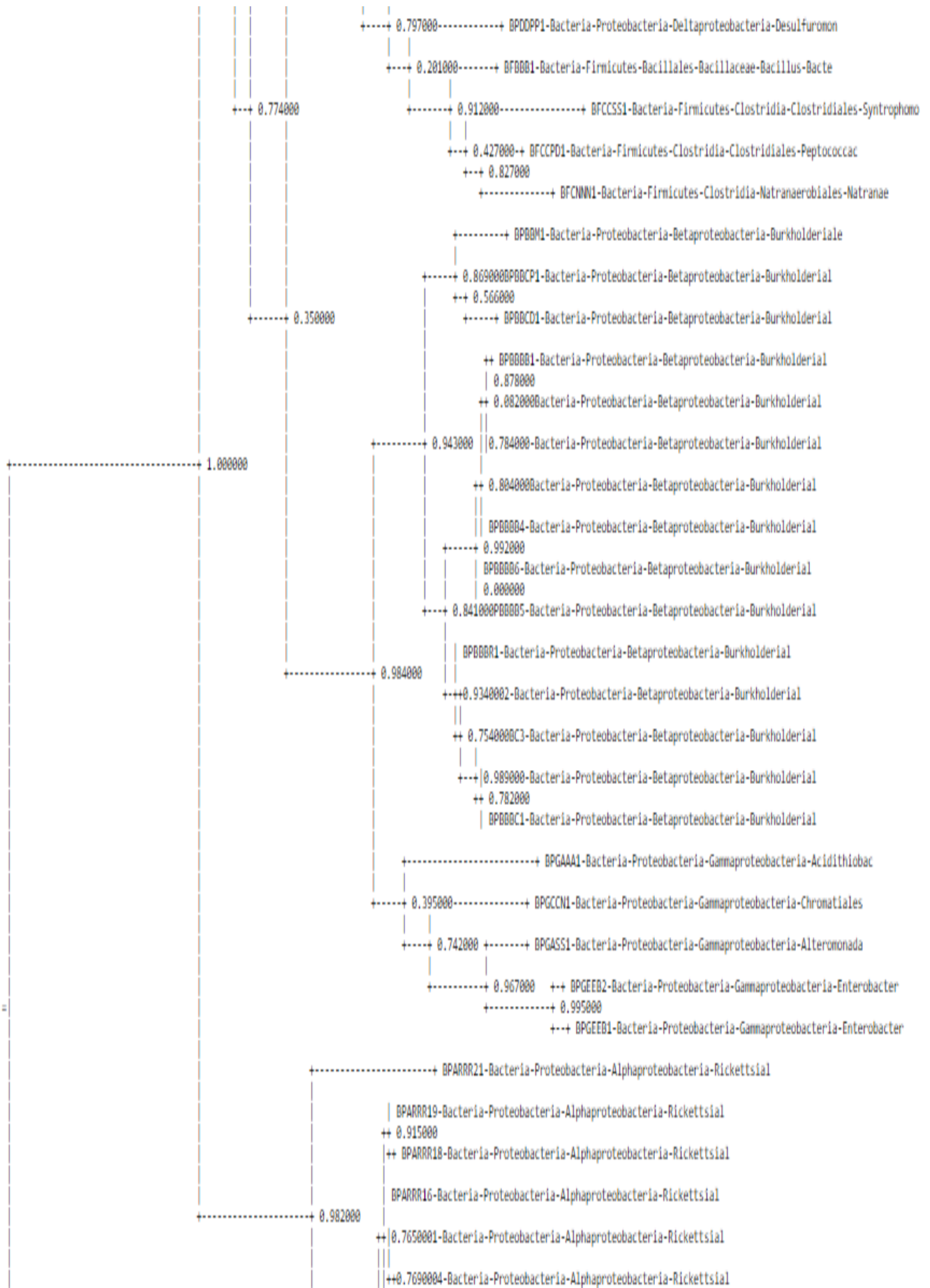
These sequences in FASTA format with legible labels are used as the input to the **ClustalW** tool to find the multiple sequence alignment and with those alignments in FASTA format we get the following phylogenetic tree with the help of **PhyML**.

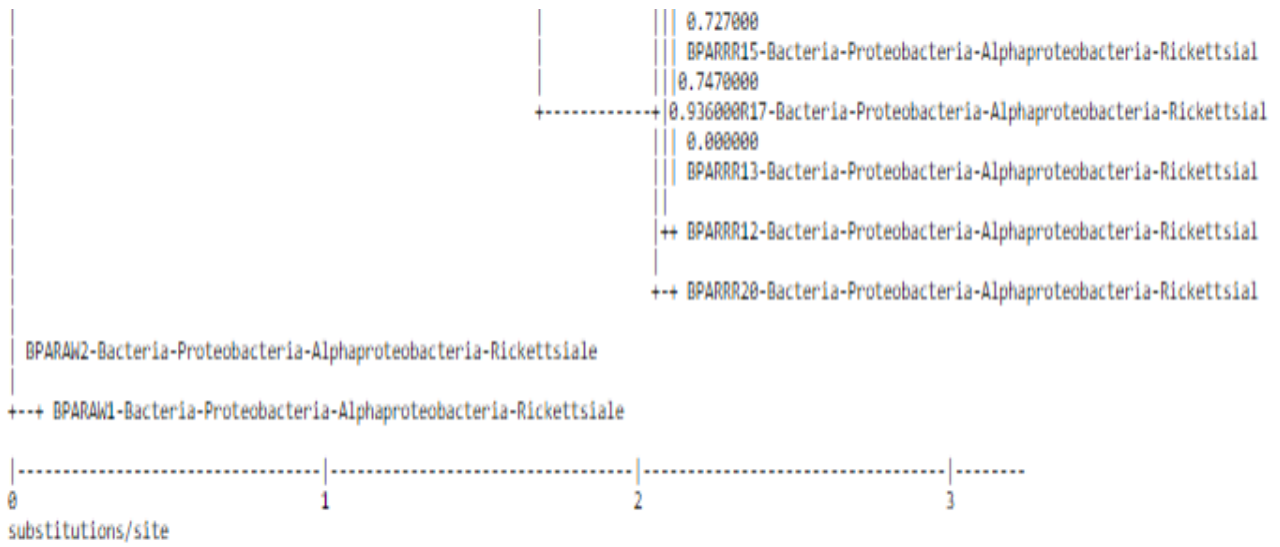
PHYLOGENETIC TREE:

```

    +-+ BPARBN2-Bacteria-Proteobacteria-Alphaproteobacteria-Rhizobiales
      | 0.820000
    + BPARBN1-Bacteria-Proteobacteria-Alphaproteobacteria-Rhizobiales
      |
    +-+ BPARBR3-Bacteria-Proteobacteria-Alphaproteobacteria-Rhizobiales
      | 0.743000
    +| 0.947000-Bacteria-Proteobacteria-Alphaproteobacteria-Rhizobiales
    +-+ 0.117000
      | 0.7550004-Bacteria-Proteobacteria-Alphaproteobacteria-Rhizobiales
      |
    +-+ BPARBR2-Bacteria-Proteobacteria-Alphaproteobacteria-Rhizobiales
    +---+ 0.9880000
      | | BPARBR1-Bacteria-Proteobacteria-Alphaproteobacteria-Rhizobiales
      | |
    +-+ BPARBB9-Bacteria-Proteobacteria-Alphaproteobacteria-Rhizobiales
      | + 0.335000
    +-+ 0.85900070008-Bacteria-Proteobacteria-Alphaproteobacteria-Rhizobiales
      | |
    +-+ BPARB02-Bacteria-Proteobacteria-Alphaproteobacteria-Rhizobiales
      | |
    +---+ BPARBB2-Bacteria-Proteobacteria-Alphaproteobacteria-Rhizobiales
    +---+ 0.977000
    +-+ 0.853000ARB01-Bacteria-Proteobacteria-Alphaproteobacteria-Rhizobiales
      | |
      | BPARXX1-Bacteria-Proteobacteria-Alphaproteobacteria-Rhizobiales
    +---+ 0.992000
      | |
      | +-+ BPARXA1-Bacteria-Proteobacteria-Alphaproteobacteria-Rhizobiales
    +-+ 0.723000M1-Bacteria-Proteobacteria-Alphaproteobacteria-Rhizobiales
      | |
    +-+ 0.988000PARW5-Bacteria-Proteobacteria-Alphaproteobacteria-Rhizobiales
      | +---+ 1.000000
      | | BPARW4-Bacteria-Proteobacteria-Alphaproteobacteria-Rhizobiales
    +-+ 0.887000
      | BPARW3-Bacteria-Proteobacteria-Alphaproteobacteria-Rhizobiales
      | 0.000000
      | BPARW2-Bacteria-Proteobacteria-Alphaproteobacteria-Rhizobiales
    +---+ 0.952000
      | |
      | BPARB01-Bacteria-Proteobacteria-Alphaproteobacteria-Rhizobiales
      |
    +-+| 0.922000-Bacteria-Proteobacteria-Alphaproteobacteria-Rhizobiales
      | +-+ 0.776000
      | | BPARBB6-Bacteria-Proteobacteria-Alphaproteobacteria-Rhizobiales
    +-+ 0.911000
      | | +-+ BPARBB5-Bacteria-Proteobacteria-Alphaproteobacteria-Rhizobiales
      | |
    +-+| 0.947000BB3-Bacteria-Proteobacteria-Alphaproteobacteria-Rhizobiales
      | |
      | +-+| 0.9230004-Bacteria-Proteobacteria-Alphaproteobacteria-Rhizobiales
      | | 0.743000
      | +-+ BPARBB1-Bacteria-Proteobacteria-Alphaproteobacteria-Rhizobiales
    +---+ 0.997000
      | +-+ BPARPC2-Bacteria-Proteobacteria-Alphaproteobacteria-Rhizobiales
      | |
      | +-+ BPARRS2-Bacteria-Proteobacteria-Alphaproteobacteria-Rhizobiales
    +-+ 0.715000.832000
      | | | BPARRS1-Bacteria-Proteobacteria-Alphaproteobacteria-Rhizobiales
      | | +---+ 0.997000
      | | | BPARRR1-Bacteria-Proteobacteria-Alphaproteobacteria-Rhizobiales
      | |
    +-+ 0.062000+0.889000-Bacteria-Proteobacteria-Alphaproteobacteria-Rhizobiales
  
```







Phylogenetic Tree

The **tree data** is also here:

Tree Data:

```
((((((((((BPARB2_Bacteria_Proteobacteria_Alphaproteobacteria_Rhizobiales:0.040679,
BPARB1_Bacteria_Proteobacteria_Alphaproteobacteria_Rhizobiales:0.007194)
0.820000:0.014261,
((BPARB3_Bacteria_Proteobacteria_Alphaproteobacteria_Rhizobiales:0.033600,
(BPARB5_Bacteria_Proteobacteria_Alphaproteobacteria_Rhizobiales:0.011362,
BPARB4_Bacteria_Proteobacteria_Alphaproteobacteria_Rhizobiales:0.010680)
0.117000:0.007451)0.947000:0.033878,
(BPARB2_Bacteria_Proteobacteria_Alphaproteobacteria_Rhizobiales:0.023412,
BPARB1_Bacteria_Proteobacteria_Alphaproteobacteria_Rhizobiales:0.000001)
0.835000:0.015876)0.755000:0.008244)0.743000:0.007735,
((BPARB9_Bacteria_Proteobacteria_Alphaproteobacteria_Rhizobiales:0.082347,
BPARB8_Bacteria_Proteobacteria_Alphaproteobacteria_Rhizobiales:0.031239)
0.335000:0.031548,
BPARB2_Bacteria_Proteobacteria_Alphaproteobacteria_Rhizobiales:0.073693)
0.697000:0.010997)0.988000:0.102707,
(BPARB2_Bacteria_Proteobacteria_Alphaproteobacteria_Rhizobiales:0.110162,
BPARB1_Bacteria_Proteobacteria_Alphaproteobacteria_Rhizobiales:0.054951)
0.977000:0.097765)0.859000:0.036463,
((BPARX1_Bacteria_Proteobacteria_Alphaproteobacteria_Rhizobiales:0.000001,
BPARX1_Bacteria_Proteobacteria_Alphaproteobacteria_Rhizobiales:0.058206)
0.992000:0.135582,
(BPARM1_Bacteria_Proteobacteria_Alphaproteobacteria_Rhizobiales:0.014705,
((BPARM5_Bacteria_Proteobacteria_Alphaproteobacteria_Rhizobiales:0.039115,
BPARM4_Bacteria_Proteobacteria_Alphaproteobacteria_Rhizobiales:0.016766)
1.000000:0.127042,
(BPARM3_Bacteria_Proteobacteria_Alphaproteobacteria_Rhizobiales:0.007204,
BPARM2_Bacteria_Proteobacteria_Alphaproteobacteria_Rhizobiales:0.000001)
0.000000:0.000001)0.887000:0.038088)0.988000:0.094480)0.723000:0.024891)
0.853000:0.044004,
(((BPARB1_Bacteria_Proteobacteria_Alphaproteobacteria_Rhizobiales:0.000001,
(BPARB7_Bacteria_Proteobacteria_Alphaproteobacteria_Rhizobiales:0.007886,
BPARB6_Bacteria_Proteobacteria_Alphaproteobacteria_Rhizobiales:0.000001)
0.776000:0.007915)0.922000:0.080878,
(BPARB5_Bacteria_Proteobacteria_Alphaproteobacteria_Rhizobiales:0.102638,
(BPARB3_Bacteria_Proteobacteria_Alphaproteobacteria_Rhizobiales:0.026225,
(BPARB4_Bacteria_Proteobacteria_Alphaproteobacteria_Rhizobiales:0.043204,
BPARB1_Bacteria_Proteobacteria_Alphaproteobacteria_Rhizobiales:0.034167)
0.743000:0.012110)0.923000:0.070620)0.947000:0.110881)0.911000:0.084898,
((BPARPC2_Bacteria_Proteobacteria_Alphaproteobacteria_Rhizobiales:0.074685,
((BPARRS2_Bacteria_Proteobacteria_Alphaproteobacteria_Rhizobiales:0.037218,
BPARRS1_Bacteria_Proteobacteria_Alphaproteobacteria_Rhizobiales:0.000001)
0.832000:0.029253,
(BPARRR1_Bacteria_Proteobacteria_Alphaproteobacteria_Rhizobiales:0.000001,
(BPARRR2_Bacteria_Proteobacteria_Alphaproteobacteria_Rhizobiales:0.014742,
(BPARRR4_Bacteria_Proteobacteria_Alphaproteobacteria_Rhizobiales:0.000001,
BPARR3_Bacteria_Proteobacteria_Alphaproteobacteria_Rhizobiales:0.007261)
0.997000:0.070240)0.693000:0.014939)0.889000:0.048999)0.997000:0.176837)
0.715000:0.036722,
BPARPM1_Bacteria_Proteobacteria_Alphaproteobacteria_Rhizobiales:0.111461)
0.062000:0.027857)0.997000:0.208357)0.952000:0.091502,
(((BPARHH1_Bacteria_Proteobacteria_Alphaproteobacteria_Rhodobacteria:0.260285,
BPARHM1_Bacteria_Proteobacteria_Alphaproteobacteria_Rhodobacteria:0.219949)
0.867000:0.062315,
(BPACCP1_Bacteria_Proteobacteria_Alphaproteobacteria_Caulobacteria:0.054012,
(BPACCC2_Bacteria_Proteobacteria_Alphaproteobacteria_Caulobacteria:0.048275,
```

BPACCC1_Bacteria_Proteobacteria_Alphaproteobacteria_Caulobactera:0.000001)
0.986000:0.151408)0.990000:0.186281)0.976000:0.168534,
((BPASEE1_Bacteria_Proteobacteria_Alphaproteobacteria_Sphingomonad:0.195801,
((BPASSN1_Bacteria_Proteobacteria_Alphaproteobacteria_Sphingomonad:0.129816,
BPASSS1_Bacteria_Proteobacteria_Alphaproteobacteria_Sphingomonad:0.226268)
0.558000:0.047435,
(BPASSS2_Bacteria_Proteobacteria_Alphaproteobacteria_Sphingomonad:0.069384,
BPASSZ1_Bacteria_Proteobacteria_Alphaproteobacteria_Sphingomonad:0.145711)
0.924000:0.084110)0.184000:0.072052)0.998000:0.356359,
(((BPARRJ1_Bacteria_Proteobacteria_Alphaproteobacteria_Rhodobactera:0.205469,
((BPARRR7_Bacteria_Proteobacteria_Alphaproteobacteria_Rhodobactera:0.106295,
BPARRR6_Bacteria_Proteobacteria_Alphaproteobacteria_Rhodobactera:0.048344)
0.449000:0.028225,
BPARRR5_Bacteria_Proteobacteria_Alphaproteobacteria_Rhodobactera:0.037389)
0.898000:0.058617)0.881000:0.038363,
BPARRD1_Bacteria_Proteobacteria_Alphaproteobacteria_Rhodobactera:0.110809)
0.661000:0.007983,
((BPARRR10_Bacteria_Proteobacteria_Alphaproteobacteria_Rhodobacter:0.000002,
BPARRR9_Bacteria_Proteobacteria_Alphaproteobacteria_Rhodobactera:0.015915)
0.986000:0.127447,
BPARRP1_Bacteria_Proteobacteria_Alphaproteobacteria_Rhodobactera:0.112956)
0.811000:0.034161)0.997000:0.256693)0.908000:0.138153)0.677000:0.032408,
BPARPP1_Bacteria_Proteobacteria_Alphaproteobacteria_Rhizobiales:0.104871)
0.726000:0.014690)0.776000:0.055703,
((BPARAA1_Bacteria_Proteobacteria_Alphaproteobacteria_Rhodospirill:0.258489,
(BPARAG1_Bacteria_Proteobacteria_Alphaproteobacteria_Rhodospirill:0.078643,
(BPARAG3_Bacteria_Proteobacteria_Alphaproteobacteria_Rhodospirill:0.087362,
BPARAG2_Bacteria_Proteobacteria_Alphaproteobacteria_Rhodospirill:0.253709)
0.719000:0.028393)0.783000:0.079048)0.984000:0.267255,
(BPARRM1_Bacteria_Proteobacteria_Alphaproteobacteria_Rhodospirill:0.138290,
BPARRR8_Bacteria_Proteobacteria_Alphaproteobacteria_Rhodospirill:0.235596)
0.980000:0.232756)0.791000:0.097383)0.912000:0.245785,
(BPARPC1_Bacteria_Proteobacteria_Alphaproteobacteria_Rickettsiale:0.891404,
((BTMMM1_Bacteria_Tenericutes_Mollicutes_Mycoplasmataceae_Mycoplas:1.250646,
(BPDDDD1_Bacteria_Proteobacteria_Deltaproteobacteria_Desulfovibri:0.725521,
(BPDDPP1_Bacteria_Proteobacteria_Deltaproteobacteria_Desulfuromon:0.503624,
(BFBBB1_Bacteria_Firmicutes_Bacillales_Bacillaceae_Bacillus_Bacte:0.267833,
(BFCCSS1_Bacteria_Firmicutes_Clostridia_Clostridiales_Syntrophomo:0.651293,
(BFCCPD1_Bacteria_Firmicutes_Clostridia_Clostridiales_Peptococcac:0.226618,
BFCNNN1_Bacteria_Firmicutes_Clostridia_Natranaerobiales_Natranae:0.387531)
0.827000:0.071105)0.427000:0.110140)0.912000:0.214554)0.201000:0.104752)
0.797000:0.149685)0.968000:0.416260,
(((BPBBM1_Bacteria_Proteobacteria_Betaproteobacteria_Burkholderiale:0.296172,
(BPBBCP1_Bacteria_Proteobacteria_Betaproteobacteria_Burkholderial:0.165274,
BPBBCD1_Bacteria_Proteobacteria_Betaproteobacteria_Burkholderial:0.170255)
0.566000:0.056805)0.869000:0.163700,
((((BPBBBB1_Bacteria_Proteobacteria_Betaproteobacteria_Burkholderial:0.010034,
BPBBBB2_Bacteria_Proteobacteria_Betaproteobacteria_Burkholderial:0.000001)
0.878000:0.020416,
BPBBBB3_Bacteria_Proteobacteria_Betaproteobacteria_Burkholderial:0.000001)
0.082000:0.018399,
BPBBBB7_Bacteria_Proteobacteria_Betaproteobacteria_Burkholderial:0.005573)
0.784000:0.016675,
BPBBBB4_Bacteria_Proteobacteria_Betaproteobacteria_Burkholderial:0.000001)
0.804000:0.010050,
(BPBBBB6_Bacteria_Proteobacteria_Betaproteobacteria_Burkholderial:0.000001,
BPBBBB5_Bacteria_Proteobacteria_Betaproteobacteria_Burkholderial:0.010075)
0.000000:0.000001)0.992000:0.184315,

```

(BPBBBBR1_Bacteria_Proteobacteria_Betaproteobacteria_Burkholderial:0.000001,
(BPBBBBR2_Bacteria_Proteobacteria_Betaproteobacteria_Burkholderial:0.032758,
(BPBBBC3_Bacteria_Proteobacteria_Betaproteobacteria_Burkholderial:0.026095,
(BPBBBC2_Bacteria_Proteobacteria_Betaproteobacteria_Burkholderial:0.009491,
BPBBBC1_Bacteria_Proteobacteria_Betaproteobacteria_Burkholderial:0.000667)
0.782000:0.015136)0.989000:0.087532)0.754000:0.019280)0.934000:0.076072)
0.841000:0.104854)0.943000:0.296288,
(BPGAAA1_Bacteria_Proteobacteria_Gammaproteobacteria_Acidithiobac:0.738794,
(BPGCCN1_Bacteria_Proteobacteria_Gammaproteobacteria_Chromatiales:0.539482,
(BPGASS1_Bacteria_Proteobacteria_Gammaproteobacteria_Alteromonada:0.243365,
(BPGEEB2_Bacteria_Proteobacteria_Gammaproteobacteria_Enterobacter:0.050450,
BPGEEB1_Bacteria_Proteobacteria_Gammaproteobacteria_Enterobacter:0.067747)
0.995000:0.386382)0.967000:0.298251)0.742000:0.147678)0.395000:0.187864)
0.984000:0.460303)0.350000:0.226036)0.774000:0.061629)0.862000:0.212914,
(BPARRR21_Bacteria_Proteobacteria_Alphaproteobacteria_Rickettsial:0.672995,
(((BPARRR19_Bacteria_Proteobacteria_Alphaproteobacteria_Rickettsial:0.007895,
BPARRR18_Bacteria_Proteobacteria_Alphaproteobacteria_Rickettsial:0.018004)
0.915000:0.026725,
((BPARRR16_Bacteria_Proteobacteria_Alphaproteobacteria_Rickettsial:0.000001,
BPARRR11_Bacteria_Proteobacteria_Alphaproteobacteria_Rickettsial:0.016705,
((BPARRR14_Bacteria_Proteobacteria_Alphaproteobacteria_Rickettsial:0.008237,
BPARRR15_Bacteria_Proteobacteria_Alphaproteobacteria_Rickettsial:0.000001)
0.727000:0.008228,
(BPARRR17_Bacteria_Proteobacteria_Alphaproteobacteria_Rickettsial:0.016573,
BPARRR13_Bacteria_Proteobacteria_Alphaproteobacteria_Rickettsial:0.008305)
0.000000:0.000001)0.481000:0.016685)0.769000:0.008195)0.000000:0.000001,
BPARRR12_Bacteria_Proteobacteria_Alphaproteobacteria_Rickettsial:0.016454)
0.747000:0.008453)0.765000:0.019617,
BPARRR20_Bacteria_Proteobacteria_Alphaproteobacteria_Rickettsial:0.046399)
0.936000:0.365605)0.982000:0.641598)1.000000:1.050406,
BPARAW2_Bacteria_Proteobacteria_Alphaproteobacteria_Rickettsiale:0.000002,
BPARAW1_Bacteria_Proteobacteria_Alphaproteobacteria_Rickettsiale:0.081578);

```

This is called the **NEWICK** tree format. It gives us complete details of a tree in text format.

4.2 DISCUSSION:

Now we see in case of our first sequence we get the **biological process** as **translation**. We also get to know that its **molecular functions** are **RNA binding** and secondly, it helps in contributing the **structural integrity of the ribosome**. Our analysis also shows that the **cellular component** here is **ribosome**.

The translation defines the cellular metabolic process in which a protein is formed using the sequence of a mature mRNA molecule to specify the sequence of amino acids in a polypeptide chain. Translation is mediated by the ribosome and begins with the formation of a ternary complex between aminoacylated initiator methionine tRNA, GTP, and initiation factor 2, which subsequently associates with the small subunit of the ribosome and an mRNA. Translation ends with the release of a polypeptide chain from the ribosome[7].

RNA binding defines interacting selectively and non-covalently with an RNA molecule or a portion thereof.

And the ribosome is an intracellular organelle, about 200 Å in diameter, consisting of RNA and protein. It is the site of protein biosynthesis resulting from translation of messenger RNA (mRNA). It consists of two subunits, one large and one small, each containing only protein and RNA. Both the ribosome and its subunits are characterized by their sedimentation coefficients, expressed in Svedberg units (symbol: S). Hence, the prokaryotic ribosome (70S) comprises a large (50S) subunit and a small (30S) subunit, while the eukaryotic ribosome (80S) comprises a large (60S) subunit and a small (40S) subunit. Two sites on the ribosomal large subunit are involved in translation, namely the aminoacyl site (A site) and peptidyl site (P site). Ribosomes from prokaryotes, eukaryotes, mitochondria, and chloroplasts have characteristically distinct ribosomal proteins.

We have applied the same methods for the other three sequences also and now we see the results.

FOR SEQUENCE 2

INTERPRO RESULT:

Here we have used ORF1 where the protein sequence is:

MIKKIAILGSTGSIGKSLKIIIEADRSFKIILLTANKDYTSLIKQANKFNVKNIITDQKNYNI AKLKIKNKKIKIFNNYNQFKNIFKN
KIDYAMSAITGLDGLKPTIDIKYTKNIAIANKESI CAWDIIKTELKKNKTFIPVDSEHFSI WYGLKDFNNSIEKIYLTASGGPFY
ATPIKLLKNIKIQALNHPNWKMGGKISIDSATMINKIYESIEAKNIFNIPYKKIEIIIIHPKSYVHALIKFNSGPIKIIAHDTTMMIPY
LILCI

Submitted Export =

Length 273 amino acids

Protein family membership
1-deoxy-D-xylulose 5-phosphate reductoisomerase (IPR003821)

Homologous superfamilies
Homologous superfamily

Domains and repeats
Domain

Detailed signature matches

- IPR036291 NAD(P)-binding domain superfamily SSF51735 (NAD(P)-bl...)
- IPR003821 1-deoxy-D-xylulose 5-phosphate reductoisomerase PTHR30525 (1-DEOXY...)
- IPR013512 1-deoxy-D-xylulose 5-phosphate reductoisomerase, N-terminal PF02670 (DXP_reduct...)
- IPR013644 1-deoxy-D-xylulose 5-phosphate reductoisomerase, C-terminal PF08436 (DXP_reduc...)
- no IPR Unintegrated signatures G3DSA:3.40.50... PTHR30525:SF0 (1-DE...) SSF55347 (glycerald...)

GO term prediction

Biological Process

- GO:0008299 isoprenoid biosynthetic process
- GO:0055114 oxidation-reduction process

Molecular Function

- GO:0005515 protein binding
- GO:0030604 1-deoxy-D-xylulose-5-phosphate reductoisomerase activity
- GO:0046872 metal ion binding
- GO:0070402 NADPH binding

Cellular Component

None predicted.

INTERPRO result for ORF1 of sequence-2

Here we see that protein family membership is **1-deoxy-D-xylulose 5-phosphate reductoisomerase**. By GO term prediction we see that it has the cellular component not predicted. Here the biological processes are **isoprenoid biosynthetic process**(GO:0008299) and **oxidation reduction process**(GO:0055114) and molecular functions are **protein binding**(GO:0005515), **NADPH binding**(GO:0070402), **metal ion binding**(GO:0046872) and **1-deoxy-D-xylulose-5-phosphate reductoisomerase activity**(GO:0030604).

BLAST RESULT:

BLAST® » blastp suite » RID-FSRXZJA401R [Home](#) [Recent Results](#) [Saved Strategies](#) [Help](#)

BLAST Results

[Edit and Resubmit](#) [Save Search Strategies](#) [Formatting options](#) [Download](#) [YouTube](#) [How to read this page](#) [Blast report description](#)

Job title: lc|ORF1 (273 letters)

RID [FSRXZJA401R](#) (Expires on 05-18 00:43 am)

Query ID	lc Query_124313	Database Name	swissprot
Description	lc ORF1	Description	Non-redundant UniProtKB/SwissProt sequences
Molecule type	amino acid	Program	BLASTP 2.8.0+ Citation
Query Length	273		


Other reports: [Search Summary](#) [Taxonomy reports](#) [Distance tree of results](#) [Multiple alignment](#) [MSA viewer](#)

New Analyze your query with [SmartBLAST](#)

Graphic Summary

[Show Conserved Domains](#)

Putative conserved domains have been detected, click on the image below for detailed results.



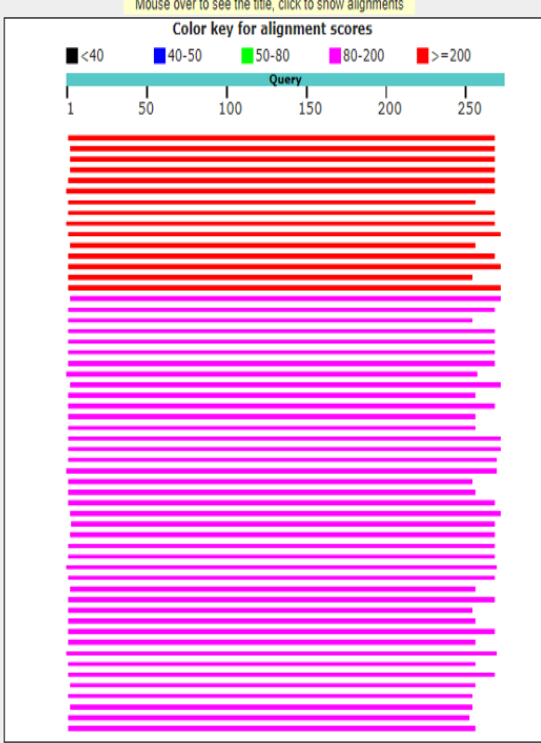
Query seq, Superfamilies: DXP_redisom_C superfamily

Distribution of the top 100 Blast Hits on 100 subject sequences

Mouse over to see the title, click to show alignments

Color key for alignment scores

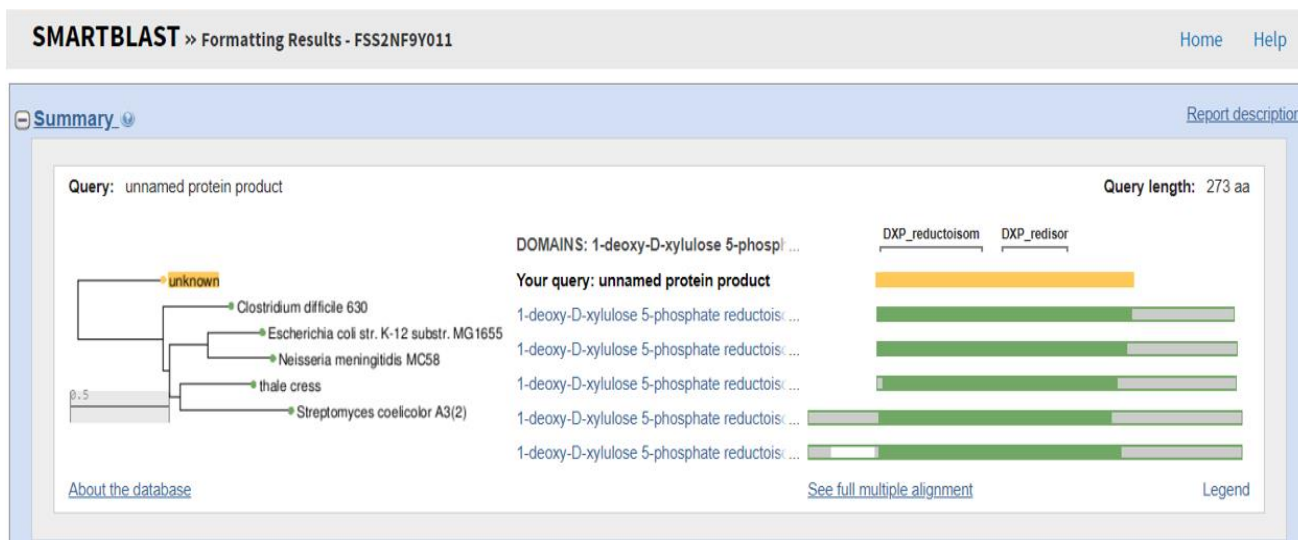
■ <40	■ 40-50	■ 50-80	■ 80-200	■ >=200
-------	---------	---------	----------	---------



Result of BLAST

This blast result shows that it has the **DXP_redisom_C** superfamily. And the distribution for the top 100 BLAST hits on 100 subject sequences are also shown.

SMARTBLAST RESULT:



SMARTBLAST result

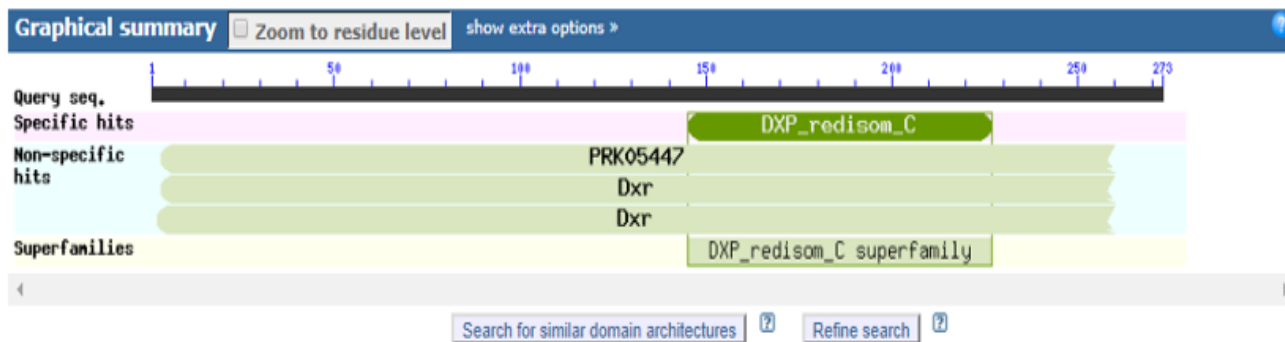
This image shows that the domain is **1-deoxy-D-xylulose 5-phosphate reductoisomerase**.

CONSERVED PROTEIN DOMAINS:

Conserved domains on [lc|ORF1]

View Standard Results ?

Local query sequence



List of domain hits					
	Name	Accession	Description	Interval	E-value
[+]	PRK05447	PRK05447	1-deoxy-D-xylulose 5-phosphate reductoisomerase; Provisional	3-260	1.28e-94
[+]	Dxr	COG0743	1-deoxy-D-xylulose 5-phosphate reductoisomerase [Lipid transport and metabolism];	3-260	6.08e-82
[+]	Dxr	TIGR00243	1-deoxy-D-xylulose 5-phosphate reductoisomerase; 1-deoxy-D-xylulose 5-phosphate is converted ...	2-260	4.03e-67
[+]	DXP_redisor_C	pfam08436	1-deoxy-D-xylulose 5-phosphate reductoisomerase C-terminal; This domain is found to the ...	145-227	1.42e-37

Blast search parameters

Data Source: Live blast search RID = FSRXZ4MG01R
 User Options: Database: CDSEARCH/odd v3.16 Low complexity filter: no Composition Based Adjustment: yes E-value threshold: 0.01 Maximum number of hits: 500

Conserved Protein Domain

From the above results we can say that the protein family membership of the sequence-2 is **1-deoxy-D-xylulose 5-phosphate reductoisomerase**.

FOR SEQUENCE 3

INTERPRO RESULT:

Here we have used ORF2 where the protein sequence is:

MTMKDFFQKVNSYVKEAFGASKYLFDGITVTFDHLKRRPVTVQYPYEKLIPSERYGRRIHYEFDKCIACEVCVRVCPINLPVVD
WVMNKETKKKELRNYSIDFGACIFCGNCVEYCPNCLSMTEEYELATFDRHNLNYDNVALGRLPTNVTSDP TVKSLKELTYLP
KRVMDPHE

InterPro Protein sequence analysis & classification

Search InterPro... 1
Examples: IPR020405, kinase, P51587, PF02932, GO:0007165

Home Search Release notes Download About InterPro Help Contact

Submitted Export =

Length 175 amino acids

Protein family membership
NADH-quinone oxidoreductase, chain I (IPR010226)
NADH-plastoquinone oxidoreductase, subunit I (IPR004497)

Homologous superfamilies
None predicted.

Domains and repeats

Detailed signature matches

IPR ID	Signature Name	Matched Signatures
IPR010226	NADH-quinone oxidoreductase, chain I	MF_01351 (NDH1_Nuel) TIGR01971
IPR004497	NADH-plastoquinone oxidoreductase, subunit I	TIGR00403 PTHR10849-SF23 (NAD...)
IPR017896	4Fe-4S ferredoxin-type, iron-sulphur binding domain	PF13237 (Fe4_10) PS51379 (4FE4S_FER_2)
IPR017900	4Fe-4S ferredoxin, iron-sulphur binding, conserved site	PS00198 (4FE4S_FER_1)
no IPR	Unintegrated signatures	G3DSA:3.30.70... PTHR10849 (NADH DEH...) SSF54862 (4Fe-4S fe...)

GO term prediction

Biological Process
[GO:0055114](#) oxidation-reduction process

Molecular Function
[GO:0008137](#) NADH dehydrogenase (ubiquinone) activity
[GO:0016651](#) oxidoreductase activity, acting on NAD(P)H
[GO:0051539](#) 4 iron, 4 sulfur cluster binding

Cellular Component
[GO:0016020](#) membrane

INTERPRO result for ORF2 of sequence-3

Here we see that protein family membership is **NADH-quinone oxidoreductase / NADH-plastoquinone oxidoreductase**. By GO term prediction we see that it has the cellular component membrane(GO:0016020). Here the biological process is **oxidation reduction process**(GO:0055114) and molecular functions are **NADH dehydrogenase (ubiquinone) activity**(GO:0008137), **oxidoreductase activity, acting on NAD(P)H**(GO:0016651), **4 iron, 4 sulfur cluster binding**(GO:0051539).

BLAST RESULT:

BLAST [®] » blastp suite » RID-FU23Y3KX014 [Home](#) [Recent Results](#) [Saved Strategies](#) [Help](#)

BLAST Results

[Edit and Resubmit](#) [Save Search Strategies](#) [Formatting options](#) [Download](#) [YouTube](#) [How to read this page](#) [Blast report description](#)

Job title: Icl|ORF2:406:930 unnamed protein product, partial (175 letters)

RID [FU23Y3KX014](#) (Expires on 05-18 12:43 pm)

Query ID	Icl Query_154175	Database Name	swissprot
Description	Icl ORF2:406:930 unnamed protein product, partial	Description	Non-redundant UniProtKB/SwissProt sequences
Molecule type	amino acid	Program	BLASTP 2.8.0+ Citation
Query Length	175		

Other reports: [Search Summary](#) [Taxonomy reports](#) [Distance tree of results](#) [Multiple alignment](#) [MSA viewer](#)

New Analyze your query with [SmartBLAST](#)

Graphic Summary

[Show Conserved Domains](#)

Putative conserved domains have been detected, click on the image below for detailed results.

Query seq. Superfamilies

Distribution of the top 100 Blast Hits on 100 subject sequences [⌵](#)

Mouse over to see the title, click to show alignments

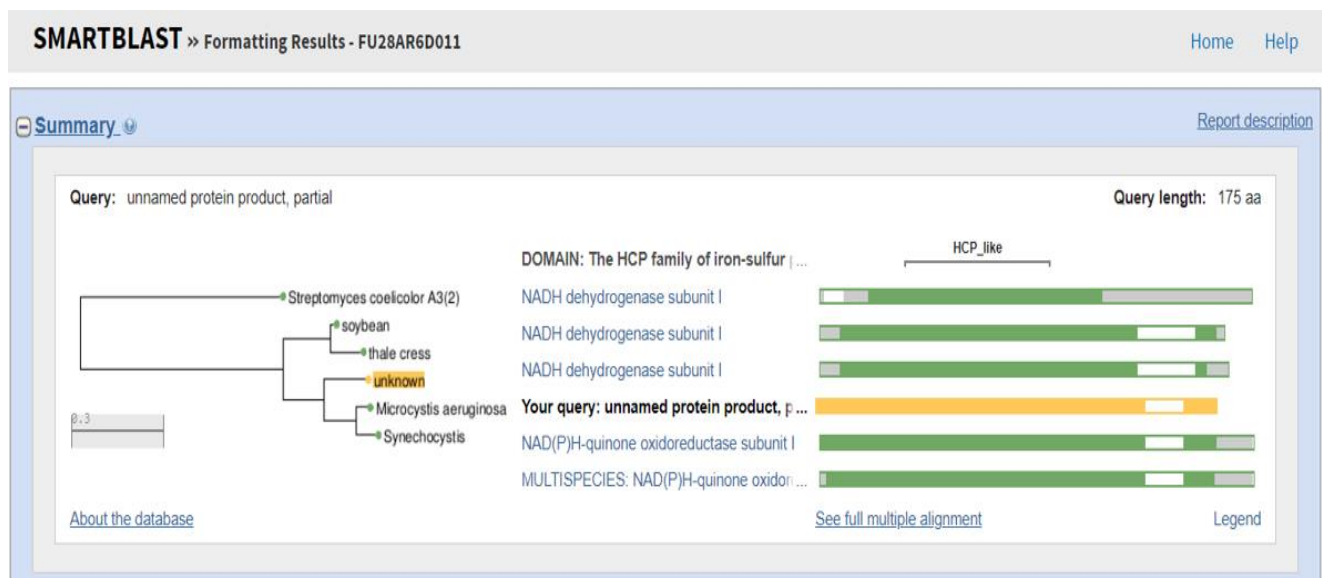
Color key for alignment scores

■ <40	■ 40-50	■ 50-80	■ 80-200	■ >=200
-------	---------	---------	----------	---------

Result of BLAST

This blast result shows that it has the **NuoI** superfamily. And the distribution for the top 100 BLAST hits on 100 subject sequences are also shown.

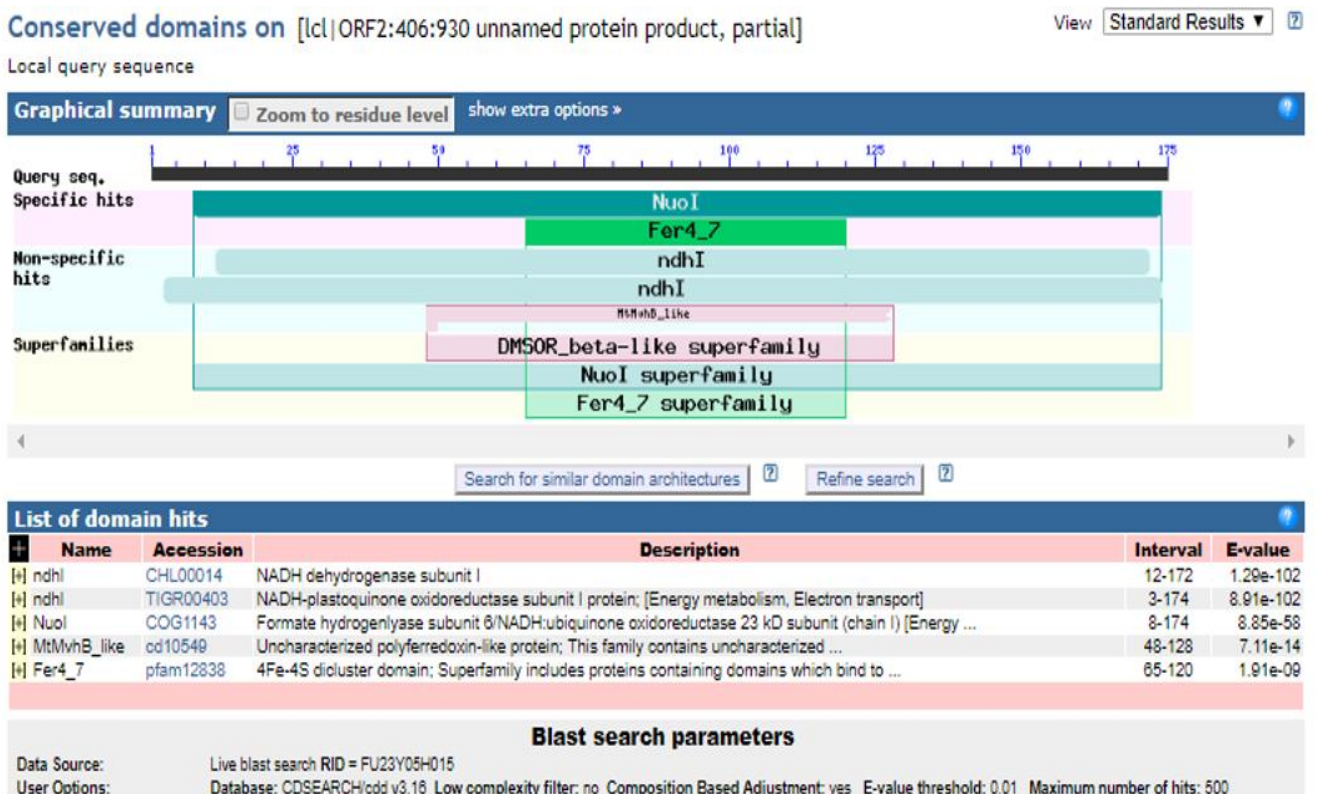
SMARTBLAST RESULT:



SMARTBLAST result

This image shows that the domain is the HCP family of iron-sulfur.

CONSERVED PROTEIN DOMAINS:



Conserved Protein Domain

Now, we see from the above results that the protein family membership of the sequence-3 is NADH-quinone oxidoreductase/NADH-plastoquinone oxidoreductase.

FOR SEQUENCE 4

INTERPRO RESULT:

Here we have used ORF2 where the protein sequence is:

MKYREFSDLGWKVKVSEIGIGCWAIGSEWGDVSIEDAKEYALYSIDHGINFFDTADVYGDGRSEKIISEVLKNSNEQIYVATKAGR
RLNPHNADGYNLKNIESFIDRSLSNLGVDIIDLVLQLHCPPSEICGPKETYEMMDEIVKKGKIKYGVSVSEKLVSEALDAIQHPNVK
SIQIIFNIFRQKPNEIFFQEAKKN

The screenshot displays the InterPro web interface for a protein sequence. The main content area is titled "Submitted" and shows the protein length as 193 amino acids. The "Protein family membership" section identifies the protein as belonging to the "Aldo/keto reductase (IPR020471)" family. The "Homologous superfamilies" section shows a single bar representing the "Homologous superfamily" from residue 1 to 193. The "Domains and repeats" section shows a single bar representing the "Domain" from residue 1 to 193. The "Detailed signature matches" section lists several matches, including "IPR036812 NADP-dependent oxidoreductase domain superfamily" (with sub-matches SSF51430 and G3DSA.3.20.20...), "IPR020471 Aldo/keto reductase" (with sub-match PR00066), "IPR023210 NADP-dependent oxidoreductase domain" (with sub-matches cd00660 and PF00248), and "no IPR Unintegrated signatures" (with sub-match PTHR43312). The "Residue annotation" section shows an "active site cd00660 catalytic tetrad c...". The "GO term prediction" section lists "Biological Process" with "GO:0055114 oxidation-reduction process", "Molecular Function" with "GO:0016491 oxidoreductase activity", and "Cellular Component" with "None predicted".

INTERPRO result for ORF2 of sequence-4

Here we see that protein family membership is **Aldo / keto reductase**. By GO term prediction we see that it has the cellular component not predicted. Here the biological process is **oxidation reduction process**(GO:0055114) and molecular function is **oxidoreductase** activity(GO:0016491).

BLAST RESULT:

BLAST [®] » blastp suite » RID-FU3EDDU1014 [Home](#) [Recent Results](#) [Saved Strategies](#) [Help](#)

BLAST Results

[Edit and Resubmit](#) [Save Search Strategies](#) [Formatting options](#) [Download](#) [YouTube](#) [How to read this page](#) [Blast report description](#)

Job title: lcl|ORF2:344:922 unnamed protein product, partial (193 letters)

RID [FU3EDDU1014](#) (Expires on 05-18 13:06 pm)

Query ID lcl Query_241163	Database Name swissprot
Description lcl ORF2:344:922 unnamed protein product, partial	Description Non-redundant UniProtKB/SwissProt sequences
Molecule type amino acid	Program BLASTP 2.8.0+ Citation
Query Length 193	


Other reports: [Search Summary](#) [Taxonomy reports](#) [Distance tree of results](#) [Multiple alignment](#) [MSA viewer](#)

New Analyze your query with [SmartBLAST](#)

Graphic Summary

[Show Conserved Domains](#)


Putative conserved domains have been detected, click on the image below for detailed results.



The diagram shows a query sequence of 193 amino acids. A yellow bar highlights a conserved domain labeled 'Aldo_ket_red superfamily' spanning from approximately position 25 to 180. Within this domain, a 'catalytic tetrad' is indicated by four yellow triangles at positions 30, 45, 60, and 75. A specific hit 'Aldo_ket_red' is also shown within the superfamily bar.

Distribution of the top 42 Blast Hits on 42 subject sequences

Mouse over to see the title, click to show alignments



Color key for alignment scores

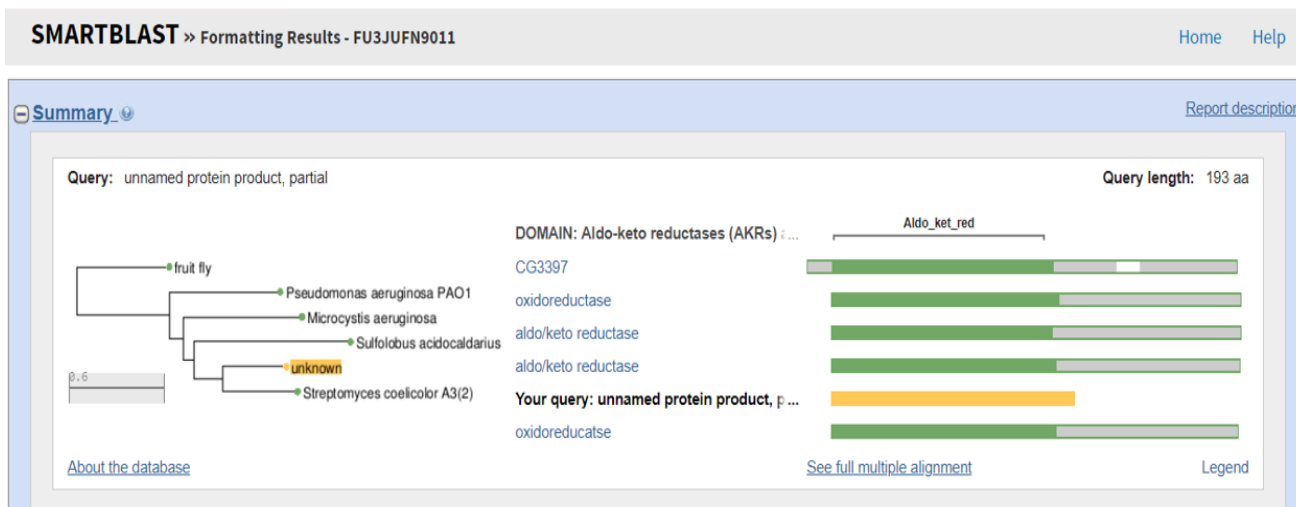
■ <40	■ 40-50	■ 50-80	■ 80-200	■ >=200
-------	---------	---------	----------	---------

The chart shows the distribution of alignment scores for 42 subject sequences. The x-axis represents the query sequence position from 1 to 180. The 'Query' sequence is shown as a light blue bar at the top. The subject sequences are represented by horizontal bars of various colors corresponding to the color key. Most subject sequences have scores between 50 and 80 (green bars), with a few having scores between 80 and 200 (magenta bars). A few subject sequences have scores less than 40 (black bars).

Result of BLAST

This blast result shows that it has the **Aldo_ket_red** superfamily. And the distribution for the top 100 BLAST hits on 100 subject sequences are also shown.

SMARTBLAST RESULT:



SMARTBLAST result

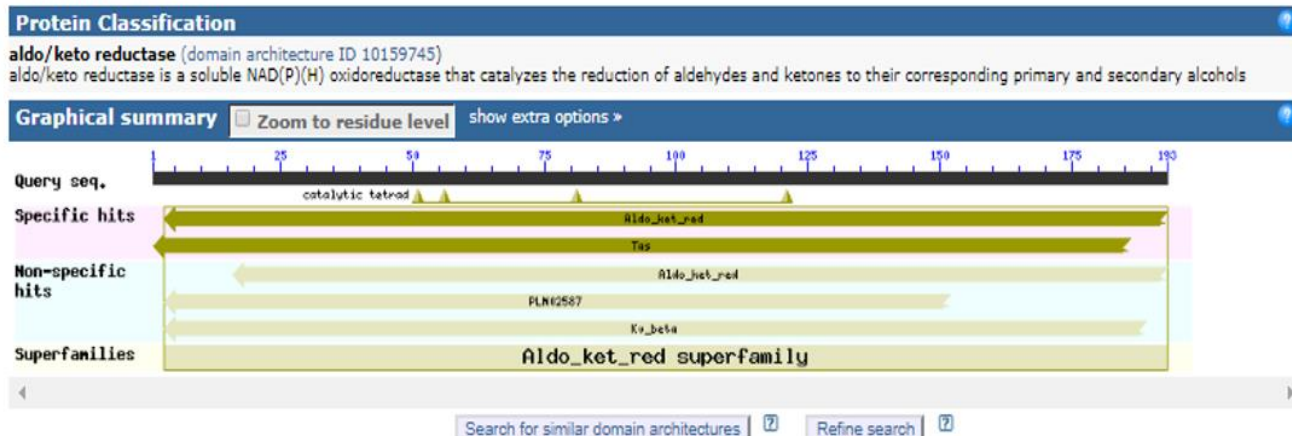
This image shows that the domain is Aldo-keto reductase.

CONSERVED PROTEIN DOMAINS:

Conserved domains on [lcl|ORF2:344:922 unnamed protein product, partial]

View **Standard Results** ?

Local query sequence



List of domain hits

#	Name	Accession	Description	Interval	E-value
[+]	Aldo_ket_red	cd06880	Aldo-keto reductases (AKRs) are a superfamily of soluble NAD(P)(H) oxidoreductases whose chief ...	3-193	2.72e-80
[+]	Tas	COG0687	Predicted oxidoreductase (related to aryl-alcohol dehydrogenase) [General function prediction ...	1-188	8.05e-52
[+]	Aldo_ket_red	pfam00248	Aldo/keto reductase family; This family includes a number of K+ ion channel beta chain ...	18-193	2.00e-45
[+]	PLN02587	PLN02587	L-galactose dehydrogenase	3-152	6.32e-21
[+]	Kv_beta	TIGR01293	voltage-dependent potassium channel beta subunit, animal; This model describes the conserved ...	3-189	1.08e-18

Blast search parameters

Data Source: Live blast search RID = FU3EC1FB014
 User Options: Database: CDSEARCH/cdd v3.16 Low complexity filter: no Composition Based Adjustment: yes E-value threshold: 0.01 Maximum number of hits: 500

Conserved Protein Domain

The above results show that the protein family membership of the sequence-4 is Aldo / keto reductase.

CHAPTER 5: RESULT SUMMERY

5.1 CONCLUSION:

In this project we have tried to analyze some sequences and also found the corresponding activities, functionalities, biological information about them. By analysing those sequences we finally reach the conclusion that from which sequence pool we have chosen them, in most of the cases we find the nearby results which has been produced before. For the first sequence we analysed we have seen that its **cellular component** is **ribosome**(GO:0005840) with the **biological process** as **translation**(GO:0006412) and it performs **RNA binding**(GO:0003723) and helps in **constituting structural integrity of the ribosome**(GO:0003735).

For the all other three sequences we got the common biological process as **oxidation reduction process**(GO:0055114) whereas the cellular component for both 2nd and 3rd sequence fragment is unpredicted and that for the 4th sequence fragment is **membrane**(GO:0016020). And we noticed for the 3rd and 4th sequence we found a common molecular function as **oxidoreductase activity**(GO:0016491) acting on NAD(P)H.

Thus we use metagenomics and bioinformatics methods and tools to analyse a completely unknown oceanic bacteria sequence fragment and get sequential results from homology prediction, conserved protein domain detection, multiple sequence alignment and phylogenetic tree analysis and at last stage we identify the taxonomic id for the unknown sequences and their biological functions including biological processes, molecular functions and cellular component annotations, the sequence and functions and get our required result.

REFERENCES:

- [1] <https://en.wikipedia.org/wiki/Metagenomics>
- [2] www.taraoceans-dataportal.org/
- [3] <https://www.ncbi.nlm.nih.gov/orffinder/>
- [4] <http://www.ebi.ac.uk/interpro/search/sequence-search>
- [5] <https://www.ebi.ac.uk/Tools/msa/clustalo/>
- [6] <https://www.ebi.ac.uk/QuickGO/>
- [7] <https://www.ncbi.nlm.nih.gov/genbank/>
- [8] <https://bioinformatics.org/sms/>
- [9] Yergeau, E. et al. Metagenomic survey of the taxonomic and functional microbial communities of sea water and sea ice from the Canadian Arctic. *Sci. Rep.* 7, 42242; doi: 10.1038/srep42242 (2017).
- [10] Wegley, L., Edwards, R., Rodriguez-Brito, B., Liu, H., & Rohwer, F. (2007). Metagenomic analysis of the microbial community associated with the coral *Porites astreoides*. *Environmental microbiology*, 9 11, 2707-19
- [11] Christian S. Riesenfeld, Patrick D. Schloss, Jo Handelsman, Metagenomics: Genomic Analysis of Microbial Communities, *Annual Review of Genetics* (2004) 38:1, 525-552
- [12] Thibault Varin, Connie Lovejoy, Anne D. Jungblut, Warwick F. Vincent, Jacques Corbeil, Metagenomic profiling of Arctic microbial mat communities as nutrient scavenging and recycling systems, *LIMNOLOGY AND OCEANOGRAPHY*, Volume 55, Issue 5, September 2010, Pages 1901-1911
- [13] Phillip B. Pope, Bharat K. C. Patel; Metagenomic analysis of a freshwater toxic cyanobacteria bloom, *FEMS Microbiology Ecology*, Volume 64, Issue 1, 1 April 2008, Pages 9-27