

BIOS 477/877 Bioinformatics and Molecular Evolution

Instructor: Etsuko Moriyama
(School of Biological Sciences)

Spring 2026 Lecture 12

BIOS477/877 L12 - 1

1

Today's topics

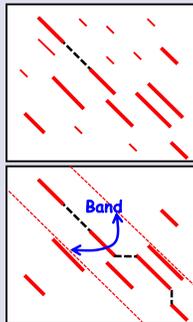
- Assignment 2 Review
- Similarity Search
 - BLAST algorithm
 - BLAST website and options
- Assignment 6

BIOS477/877 L12 - 2

2

FASTA algorithm: summary

1. Find identities using **k-tuples**
2. Join diagonals without gaps
3. Choose top 10 diagonals using a scoring matrix (e.g., BLOSUM62) (init1: the top diagonal score)
4. Join again with gaps (initn: score of the longer diagonal)
5. A **diagonal band** is defined (width: 32 if k=1, 16 if k=2 for protein)
6. Find optimal local alignment using dynamic programming algorithm within the **band** (opt: the final score)

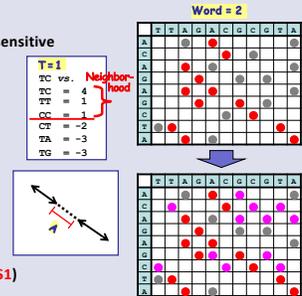


BIOS477/877 L12 - 3

3

BLAST algorithm: summary

- Word-matching size (**W**)
 - Longer words: faster but less sensitive
- Neighborhood threshold (**T**)
 - Lower T:
 - detects weaker similarities
 - slower but more sensitive
- Extension
 - Drop-off score threshold (**X**)
 - Two-hit method (**A**: distance b/w 2 hits)
- HSP selection (ungapped alignment threshold: **S1**)
- Gapped HSP extension (gapped alignment threshold: **S2**)



BIOS477/877 L12 - 4

4

BLAST similarity search

BIOS477/877 L12 - 5

5

blastp protein similarity search

BIOS477/877 L12 - 6

6

blastp protein similarity search

Query sequence(s)* can be in:

- FASTA format
- bare sequence(s)
- accession number(s) (e.g., P01013)
- gi number(s) (e.g., 129295)

*Multiple sequences can be searched

Default database is ClusteredNR:

- based on the protein NR database
- Clustered at 90% identity
- Redundancy can be reduced

BIOS477/877 L12 - 7

7

BLAST databases (protein)

Protein databas ^{default}	Description
Clustered non-redundant	Derived from the protein NR database by clustering sequences at 90% identity and 90% length.
RefSeq Select	NCBI RefSeq protein sequences from human, mouse, and prokaryotes
RefSeq proteins	NCBI Protein Reference Sequences
Model Organisms (landmark)	Proteome of 27 model organisms spanning a wide taxonomic range
UniProtKB/Swiss-Prot	Non-redundant UniProtKB/SwissProt sequences
Non-redundant (nr)	Non-redundant protein sequences from GenPept, Swissprot, PIR, PDF, PDB, and NCBI RefSeq
Patented	Protein sequences derived from the Patent division of GenBank
Protein Data Bank (PDB)	Sequences from the Protein Data Bank
Metagenomic proteins	Proteins from WGS metagenomic projects
Transcriptome Shotgun Assembly proteins	CDS features on mRNA sequences in the Transcriptome Shotgun Assembly sequences

RefSeq: A comprehensive, integrated, non-redundant, well-annotated set of reference sequences including genomic DNA, transcripts, and proteins.

Model Organisms (landmark) database: used by SmartBLAST

BIOS477/877 L12 - 8

8

BLAST landmark database (protein)

➤ Proteome from 27 representative genomes

#	Organism	Assembly	Superkingdom
1	<i>Escherichia coli</i>	NC_009416.1	Bacteria
2	<i>Escherichia coli</i>	NC_009416.1	Bacteria
3	<i>Escherichia coli</i>	NC_009416.1	Bacteria
4	<i>Mycobacterium tuberculosis</i>	H37Rv	Bacteria
5	<i>Mycobacterium tuberculosis</i>	H37Rv	Bacteria
6	<i>Mycobacterium tuberculosis</i>	H37Rv	Bacteria
7	<i>Neisseria meningitidis</i>	CGSC_009009205.1	Bacteria
8	<i>Neisseria meningitidis</i>	CGSC_009009205.1	Bacteria
9	<i>Streptococcus pneumoniae</i>	ATCC_49619	Bacteria
10	<i>Streptococcus pneumoniae</i>	ATCC_49619	Bacteria
11	<i>Streptococcus pneumoniae</i>	ATCC_49619	Bacteria
12	<i>Synschiocystis</i>	ATCC_11488	Bacteria
13	<i>Trichomonas axosporus</i>	ATCC_30210	Archaea
14	<i>Methanostreptococcus thermophilus</i>	ATCC_35061	Archaea
15	<i>Halobacterium salinarum</i>	R1	Eukaryota
16	<i>Drosophila melanogaster</i>	BDGP_5	Eukaryota
17	<i>Drosophila melanogaster</i>	BDGP_5	Eukaryota
18	<i>Drosophila melanogaster</i>	BDGP_5	Eukaryota
19	<i>Drosophila melanogaster</i>	BDGP_5	Eukaryota
20	<i>Drosophila melanogaster</i>	BDGP_5	Eukaryota
21	<i>Arabidopsis thaliana</i>	Col-0	Eukaryota
22	<i>Arabidopsis thaliana</i>	Col-0	Eukaryota
23	<i>Arabidopsis thaliana</i>	Col-0	Eukaryota
24	<i>Arabidopsis thaliana</i>	Col-0	Eukaryota
25	<i>Arabidopsis thaliana</i>	Col-0	Eukaryota
26	<i>Arabidopsis thaliana</i>	Col-0	Eukaryota
27	<i>Schistosoma mansoni</i>	SM96	Eukaryota

➔ Used for a quick search across all the kingdoms (SmartBLAST) BIOS477/877 L12 - 9

9

blastp protein similarity search

Search can be limited against (or excluding) specific organism(s). But remember how it affects the database size and E-values!

Click to change the algorithm parameters

BIOS477/877 L12 - 10

10

blastp search parameters

Available combinations of gap penalties are different depending on the scoring matrix

Only a limited set of scoring matrix and gap penalty combinations are available: BLAST uses a lookup table for K and lambda

BIOS477/877 L12 - 11

11

Estimation of K and lambda in BLAST

➤ K and lambda → from Altschul & Gish (1996) table.

a	b	A	K	H (nats)	a	b	A	K	H (nats)
0-∞	0	0.232	0.11	0.34	11	8-11	0.197	0.05	0.21
16	4-16	0.222	0.08	0.31	11	6-7	0.190	0.04	0.19
16	3	0.213	0.06	0.27	11	5	0.184	0.04	0.17
16	2	0.207	0.05	0.24	11	4	0.177	0.03	0.15
16	1	0.180	0.024	0.15	11	3	0.167	0.028	0.11
15	8-15	0.222	0.09	0.31	11	2	0.130	0.009	0.08
15	6-7	0.219	0.08	0.29	10	8-10	0.183	0.04	0.17
15	4	0.216	0.07	0.28	10	6-7	0.178	0.035	0.16
15	3	0.210	0.06	0.25	10	5	0.168	0.026	0.13
15	2	0.202	0.05	0.22	10	4	0.156	0.020	0.10
15	1	0.166	0.018	0.11	10	3	0.139	0.013	0.07
14	8-14	0.218	0.08	0.29	10	2	0.099	0.007	0.03
14	5-7	0.214	0.07	0.27	10	1	0.070	0.002	0.01
14	4	0.205	0.05	0.24	9	7-9	0.164	0.029	0.13
14	3	0.201	0.05	0.22	9	5-6	0.152	0.021	0.10
14	2	0.188	0.034	0.17	9	4	0.134	0.014	0.07
14	1	0.140	0.009	0.07	9	3	0.107	0.008	0.04
13	8-13	0.211	0.06	0.27	9	1,2	0.070	0.002	0.01
13	5-7	0.205	0.05	0.24	8	8	0.139	0.017	0.08
13	4	0.202	0.05	0.22	8	7	0.134	0.015	0.07
13	3	0.188	0.034	0.15	8	6	0.127	0.013	0.06
13	2	0.174	0.025	0.13	8	5	0.117	0.011	0.05
13	1	0.114	0.006	0.04	8	4	0.101	0.009	0.03
12	7-12	0.205	0.06	0.24	8	1-3	0.070	0.002	0.01
12	5-6	0.197	0.05	0.21	7	7	0.105	0.010	0.04
12	4	0.192	0.04	0.18	7	6	0.094	0.010	0.03
12	3	0.176	0.028	0.15	7	1-5	0.070	0.002	0.01
12	2	0.158	0.019	0.10	7	1-6	0.070	0.002	0.01
12	1	0.108	0.009	0.04	7	1-6	0.070	0.002	0.01

← For BLOSUM50;
a: gap opening penalty
b: gap extension penalty (affine gap penalty)

Based on 10,000 random sequence pairs,
→ calculate optimal local alignment scores, S,
→ estimate K and lambda by fitting the distribution of S
with $P(S \geq x) \approx Km e^{-\lambda x}$
This is what BLAST does

BIOS477/877 L11 - 12

12

blastp search parameters

Algorithm parameters

General Parameters Restore default search parameters

Max target sequences: 100

Short queries: Automatically adjust parameters for short input sequences

Expect threshold: 0.05

Word size: 3

Max matches in a query range: 0

Scoring matrix is adjusted based on amino acid composition, yielding more accurate E-values

To mask off segments of low compositional complexity

Matrix: BLOSUM62

Gap Costs: Existence: 11 Extension: 1

Compositional adjustments: Conditional compositional score matrix adjustment

Filters and Masking

Filter: Low complexity regions

Mask: Mask for lookup table only

Mask lower case letters

BLAST Search database ClusteredNR using Blastp (protein-protein BLAST)

Show results in a new window

BIOS477/877 L12 - 13

13

blastp similarity search (example)

Query: Q58746.1 (AGLUS_METJIA)
Archaeal glutamate synthase [Methanocaldococcus jannaschii DSM 2661]

BLAST® » blastp suite » results for RID-TAS17E2W014

blastp **blastp** blastx tblastn tblastx Standard Protein BLAST

Enter Query Sequence: **Q58746.1** ← Accession number as an input

Database: ClusteredNR (nr_cluster_test)

Program Selection: blastp (protein-protein BLAST)

BLAST Search database ClusteredNR using Blastp (protein-protein BLAST)

BIOS477/877 L12 - 14

14

blastp similarity search: result page

BLAST® » blastp suite » results for RID-TAS17E2W014

Job Title: sp|Q58746.1

RID: TAS17E2W014

Program: BLASTP

Database: ClusteredNR

Query ID: Q58746.1

Description: ReName: Full-Archaeal glutamate synthase [NADPH] A...

Molecule type: amino acid

Query Length: 510

Other reports: Distance tree of results Multiple alignment MSA viewer

Clusters producing significant alignments

Cluster Composition	Cluster Ancestor	Cluster Representative Sequence	Max Score	Total Score	Query Cover	E Value	Per. Ident	Acc. Len	Accession
[E] (members): 1 organism(s)	survivochaetidae	glutamate synthase-related protein [Methanocaldococcus sp.]	1027	1027	100%	0.0	97.65%	510	WP_456372201.1
[E] (members): 8 organism(s)	survivochaetidae	glutamate synthase-related protein [Methanocaldococcus infernalis]	947	947	100%	0.0	91.18%	508	WP_013100277.1
[E] (members): 1 organism(s)	survivochaetidae	glutamate synthase-related protein [Methanocaldococcus indolis]	916	916	100%	0.0	88.43%	503	WP_423796411.1
[E] (members): 1 organism(s)	survivochaetidae	glutamate synthase-related protein [Methanocaldococcus thibaultii]	913	913	100%	0.0	88.02%	503	WP_306599331.1
[E] (members): 1 organism(s)	survivochaetidae	glutamate synthase [NADPH] (GDS) subunit [Methanocaldococcus sp.]	910	910	89%	0.0	88.24%	454	AL030553.1

BIOS477/877 L12 - 15

15

blastp similarity search: result page

BLAST® » blastp suite » results for RID-TAS17E2W014

Job Title: sp|Q58746.1

RID: TAS17E2W014

Program: BLASTP

Database: ClusteredNR

Query ID: Q58746.1

Description: ReName: Full-Archaeal glutamate synthase [NADPH] A...

Molecule type: amino acid

Query Length: 510

Other reports: Distance tree of results Multiple alignment MSA viewer

Clusters producing significant alignments

Cluster Composition	Cluster Ancestor	Cluster Representative Sequence	Max Score	Total Score	Query Cover	E Value	Per. Ident	Acc. Len	Accession
[E] (members): 1 organism(s)	survivochaetidae	glutamate synthase-related protein [Methanocaldococcus sp.]	1027	1027	100%	0.0	97.65%	510	WP_456372201.1
[E] (members): 8 organism(s)	survivochaetidae	glutamate synthase-related protein [Methanocaldococcus infernalis]	947	947	100%	0.0	91.18%	508	WP_013100277.1
[E] (members): 1 organism(s)	survivochaetidae	glutamate synthase-related protein [Methanocaldococcus indolis]	916	916	100%	0.0	88.43%	503	WP_423796411.1
[E] (members): 1 organism(s)	survivochaetidae	glutamate synthase-related protein [Methanocaldococcus thibaultii]	913	913	100%	0.0	88.02%	503	WP_306599331.1
[E] (members): 1 organism(s)	survivochaetidae	glutamate synthase [NADPH] (GDS) subunit [Methanocaldococcus sp.]	910	910	89%	0.0	88.24%	454	AL030553.1

BIOS477/877 L12 - 16

16

blastp similarity search: result page

BLAST® » blastp suite » results for RID-TAS17E2W014

Job Title: sp|Q58746.1

RID: TAS17E2W014

Program: BLASTP

Database: ClusteredNR

Query ID: Q58746.1 Query name

Description: ReName: Full-Archaeal glutamate synthase [NADPH] A...

Molecule type: amino acid

Query Length: 510 Query length

Other reports: Distance tree of results Multiple alignment MSA viewer

Clusters producing significant alignments

Cluster Composition	Cluster Ancestor	Cluster Representative Sequence	Max Score	Total Score	Query Cover	E Value	Per. Ident	Acc. Len	Accession
[E] (members): 1 organism(s)	survivochaetidae	glutamate synthase-related protein [Methanocaldococcus sp.]	1027	1027	100%	0.0	97.65%	510	WP_456372201.1
[E] (members): 8 organism(s)	survivochaetidae	glutamate synthase-related protein [Methanocaldococcus infernalis]	947	947	100%	0.0	91.18%	508	WP_013100277.1
[E] (members): 1 organism(s)	survivochaetidae	glutamate synthase-related protein [Methanocaldococcus indolis]	916	916	100%	0.0	88.43%	503	WP_423796411.1
[E] (members): 1 organism(s)	survivochaetidae	glutamate synthase-related protein [Methanocaldococcus thibaultii]	913	913	100%	0.0	88.02%	503	WP_306599331.1
[E] (members): 1 organism(s)	survivochaetidae	glutamate synthase [NADPH] (GDS) subunit [Methanocaldococcus sp.]	910	910	89%	0.0	88.24%	454	AL030553.1

BIOS477/877 L12 - 17

17

blastp similarity search: result page

BLAST® » blastp suite » results for RID-TAS17E2W014

Job Title: sp|Q58746.1

RID: TAS17E2W014

Program: BLASTP

Database: ClusteredNR

Query ID: Q58746.1

Description: ReName: Full-Archaeal glutamate synthase [NADPH] A...

Molecule type: amino acid

Query Length: 510

Other reports: Distance tree of results Multiple alignment MSA viewer

Clusters producing significant alignments

Cluster Composition	Cluster Ancestor	Cluster Representative Sequence	Max Score	Total Score	Query Cover	E Value	Per. Ident	Acc. Len	Accession
[E] (members): 1 organism(s)	survivochaetidae	glutamate synthase-related protein [Methanocaldococcus sp.]	1027	1027	100%	0.0	97.65%	510	WP_456372201.1
[E] (members): 8 organism(s)	survivochaetidae	glutamate synthase-related protein [Methanocaldococcus infernalis]	947	947	100%	0.0	91.18%	508	WP_013100277.1
[E] (members): 1 organism(s)	survivochaetidae	glutamate synthase-related protein [Methanocaldococcus indolis]	916	916	100%	0.0	88.43%	503	WP_423796411.1
[E] (members): 1 organism(s)	survivochaetidae	glutamate synthase-related protein [Methanocaldococcus thibaultii]	913	913	100%	0.0	88.02%	503	WP_306599331.1
[E] (members): 1 organism(s)	survivochaetidae	glutamate synthase [NADPH] (GDS) subunit [Methanocaldococcus sp.]	910	910	89%	0.0	88.24%	454	AL030553.1

BIOS477/877 L12 - 18

18

blastp similarity search result: Graphic summary and conserved domains

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

Conserved domains: PreA, Glu_synthase

Distribution of the top 100 Blast Hits on 100 subject clusters

BIOS477/877 L12 - 19

19

blastp similarity search result: Alignments

glutamate synthase-related protein [Methanocaldococcus sp.]

Sequence ID: WP_456372.001. Length: 510. Number of Matches: 1

Alignment view: Pairwise

BIOS477/877 L12 - 20

20

blastp similarity search result: Taxonomic view

Organism	Blast Name	Score	Number of Hits	Description
cellular organisms			100	
Archaea	archaea	98		
... Methanobacteriia	archaea	94		
... Methanobacteriota	surcharaeotes	92		
... Methanococcales	surcharaeotes	13		
... Methanocaldococceae	surcharaeotes	6		
... Methanocaldococcus	surcharaeotes	5		
... Methanocaldococcus sp.	surcharaeotes	1027	1	Methanocaldococcus sp.
... Methanocaldococcus infernus	surcharaeotes	947	1	Methanocaldococcus infernus
... Methanocaldococcus indusii	surcharaeotes	916	1	Methanocaldococcus indusii
... Methanocaldococcus villosus	surcharaeotes	913	1	Methanocaldococcus villosus
... Methanocaldococcus bathoederseni	surcharaeotes	910	1	Methanocaldococcus bathoederseni
... Methanocaldococcus formicicus	surcharaeotes	875	1	Methanocaldococcus formicicus
... Methanocaldococcus abysus	surcharaeotes	869	1	Methanocaldococcus abysus
... Methanothermococcus sp. A23	surcharaeotes	866	1	Methanothermococcus sp. A23
... Methanothermococcus	surcharaeotes	865	1	Methanothermococcus
... Methanococcus aeolicus	surcharaeotes	860	1	Methanococcus aeolicus
... Methanococcus marisnigri	surcharaeotes	848	1	Methanococcus marisnigri

BIOS477/877 L12 - 21

21

blastp similarity search result: Tree view

Select sequences to be included in the tree

Sequence	Max Score	Total Score	Query Cover	E value	Per. Ident	Acc. Len	Accession
Methanocaldococcus sp.	1027	1027	100%	0.0	97.69%	510	WP_456372.001
Methanocaldococcus infernus	947	947	100%	0.0	91.18%	506	WP_45130277.1
Methanocaldococcus indusii	916	916	100%	0.0	88.43%	503	WP_45130261.1
Methanocaldococcus villosus	913	913	100%	0.0	88.20%	503	WP_04939333.1
Methanocaldococcus bathoederseni	910	910	89%	0.0	88.24%	454	MAR08102.1

BIOS477/877 L12 - 22

22

blastp similarity search result: Tree view

Phylogeny based on pairwise distance from BLAST pairwise alignments

→ Approximated tree. For a more accurate phylogeny, distances need to be estimated from the multiple alignment.

BIOS477/877 L12 - 23

23

blastp similarity search result: How to download the results

Download all or selected BLAST results:

- BLAST search result in text format
- Sequences and alignments in FASTA format
- BLAST hit statistics in "Hit Table (csv)" [Can be imported to any spread sheet program (e.g., Excel)]

BIOS477/877 L12 - 24

24

blastp similarity search: ClusteredNR vs. regular NR databases

from ClusteredNR

Cluster	Representative Sequence	Max Score	Total Score	Query Cover	E value	Per. Ident	Acc. Len	Accession
[B] (to members): 8 organisms	slu04616.ayrbaase-related.protein (Methanocaldococcus sp.)	1027	1027	100%	0.0	97.85%	510	WP_456373703.1
[B] (to members): 8 organisms	slu04616.ayrbaase-related.protein (Methanocaldococcus infernalis)	947	947	100%	0.0	91.18%	508	WP_013100272.1
[B] (to members): 1 organism	slu04616.ayrbaase-related.protein (Methanocaldococcus indicus)	916	916	100%	0.0	88.43%	503	WP_423739241.1
[B] (to members): 1 organism	slu04616.ayrbaase-related.protein (Methanocaldococcus rufus)	913	913	100%	0.0	89.02%	503	WP_004992021.1
[B] (to members): 3 organisms	slu04616.ayrbaase-related.protein (Methanocaldococcus sp.)	910	910	89%	0.0	98.24%	494	GG020282.1
[B] (to members): 3 organisms	slu04616.ayrbaase-related.protein (Methanocaldococcus slovac)	889	889	100%	0.0	87.03%	510	WP_133007298.1
[B] (to members): 4 organisms	slu04616.ayrbaase-related.protein (Methanocaldococcus sp.)	0.0	86.44%	510	WP_423739241.1	86.44%	510	WP_015870869.1
[B] (to members): 4 organisms	slu04616.ayrbaase-related.protein (Methanocaldococcus sp.)	0.0	85.85%	510	WP_015870869.1	85.85%	510	WP_015870869.1
[B] (to members): 1 organism	slu04616.ayrbaase-related.protein (Methanocaldococcus sp.)	0.0	85.07%	510	WP_013125009.1	85.07%	510	WP_013125009.1
[B] (to members): 2 organisms	slu04616.ayrbaase-related.protein (Methanocaldococcus sp.)	0.0	87.23%	510	WP_007546828.1	87.23%	510	WP_007546828.1

From Regular NR

Cluster	Representative Sequence	Max Score	Total Score	Query Cover	E value	Per. Ident	Acc. Len	Accession
[B] (to members): 8 organisms	slu04616.ayrbaase-related.protein (Methanocaldococcus sp.)	1027	1027	100%	0.0	100.00%	510	WP_015870869.1
[B] (to members): 8 organisms	slu04616.ayrbaase-related.protein (Methanocaldococcus infernalis)	1027	1027	100%	0.0	98.43%	510	WP_013100272.1
[B] (to members): 8 organisms	slu04616.ayrbaase-related.protein (Methanocaldococcus infernalis)	1027	1027	100%	0.0	98.43%	510	WP_013100272.1
[B] (to members): 8 organisms	slu04616.ayrbaase-related.protein (Methanocaldococcus infernalis)	1027	1027	100%	0.0	97.84%	510	WP_214406072.1
[B] (to members): 8 organisms	slu04616.ayrbaase-related.protein (Methanocaldococcus infernalis)	1027	1027	100%	0.0	97.84%	510	WP_015290758.1
[B] (to members): 8 organisms	slu04616.ayrbaase-related.protein (Methanocaldococcus infernalis)	1027	1027	100%	0.0	97.84%	510	WP_048201303.1
[B] (to members): 8 organisms	slu04616.ayrbaase-related.protein (Methanocaldococcus infernalis)	1027	1027	100%	0.0	97.85%	510	CA63287833.1
[B] (to members): 8 organisms	slu04616.ayrbaase-related.protein (Methanocaldococcus sp.)	1027	1027	100%	0.0	97.85%	510	WP_456373703.1
[B] (to members): 8 organisms	slu04616.ayrbaase-related.protein (Methanocaldococcus sp.)	1028	1028	100%	0.0	97.45%	510	WP_456373703.1
[B] (to members): 8 organisms	slu04616.ayrbaase-related.protein (Methanocaldococcus sp.)	1028	1028	100%	0.0	97.45%	510	WP_203137711.1
[B] (to members): 8 organisms	slu04616.ayrbaase-related.protein (Methanocaldococcus sp.)	1028	1028	100%	0.0	97.45%	510	WP_203137711.1

BIOS477/877 L12 - 25

25

blastp similarity search: ClusteredNR vs. regular NR databases

Search result from ClusteredNR

Taxonomy	Number of hits	Number of Organisms
Prokaryota	100	71
Bacteria	96	67
Firmicutes	96	66
Methanobacteriota	96	66
Methanococcales	13	13
Methanocaldococcus group	78	50
Methanococcus	73	49
Methanocaldococcus	2	1
Methanocaldococcus	3	1
Methanocaldococcus	2	2
Methanocaldococcus	2	1
Methanocaldococcus	4	4

Search result from regular NR

Taxonomy	Number of hits	Number of Organisms
Prokaryota	210	98
Bacteria	109	45
Methanobacteriota	101	65

BIOS477/877 L12 - 26

26

blastp search parameters: Max target sequences

Algorithm parameters

General Parameters

Max target sequences: 100

Short queries: Automatically adjust parameters for short input sequences

Expect threshold: 0.05

Word size: 5

Max matches in a query range: 0

Scoring Parameters

Matrix: BLOSUM62

Gap Costs: Existence: 11 Extension: 1

Compositional adjustments: Conditional compositional score matrix adjustment

Filters and Masking

Filter: Low complexity regions

Mask: Mask for lookup table only, Mask lower case letters

BLAST Search database: ClusteredNR using Blastp (protein-protein BLAST)

BIOS477/877 L12 - 27

27

blastp search parameters: Max target sequences

Query: Q58746.1 (AGLUS_METJIA)

Sequences producing significant alignments

select all 100 sequences selected

Description	Score	Max Total	Query E	Per. Ident	Acc. Len	Accession
slu04616.ayrbaase-related.protein (Methanocaldococcus sp.)	1027	1027	0.0	97.85%	510	WP_456373703.1
slu04616.ayrbaase-related.protein (Methanocaldococcus infernalis)	947	947	0.0	91.18%	508	WP_013100272.1
slu04616.ayrbaase-related.protein (Methanocaldococcus indicus)	916	916	0.0	88.43%	503	WP_423739241.1
slu04616.ayrbaase-related.protein (Methanocaldococcus rufus)	913	913	0.0	89.02%	503	WP_004992021.1
slu04616.ayrbaase-related.protein (Methanocaldococcus sp.)	910	910	0.0	98.24%	494	GG020282.1
slu04616.ayrbaase-related.protein (Methanocaldococcus slovac)	889	889	0.0	87.03%	510	WP_133007298.1
slu04616.ayrbaase-related.protein (Methanocaldococcus sp.)	0.0	86.44%	510	86.44%	510	WP_015870869.1
slu04616.ayrbaase-related.protein (Methanocaldococcus sp.)	0.0	85.85%	510	85.85%	510	WP_015870869.1
slu04616.ayrbaase-related.protein (Methanocaldococcus sp.)	0.0	85.07%	510	85.07%	510	WP_013125009.1
slu04616.ayrbaase-related.protein (Methanocaldococcus sp.)	0.0	87.23%	510	87.23%	510	WP_007546828.1

BIOS477/877 L12 - 28

28

blastp search parameters: Max target sequences

Algorithm parameters

General Parameters

Max target sequences: 5000

Short queries: Automatically adjust parameters for short input sequences

Expect threshold: 0.05

Word size: 5

Max matches in a query range: 0

Scoring Parameters

Matrix: BLOSUM62

Gap Costs: Existence: 11 Extension: 1

Compositional adjustments: Conditional compositional score matrix adjustment

Filters and Masking

Filter: Low complexity regions

Mask: Mask for lookup table only, Mask lower case letters

BLAST Search database: ClusteredNR using Blastp (protein-protein BLAST)

BIOS477/877 L12 - 29

29

blastp search parameters: Max target sequences

Query: Q58746.1 (AGLUS_METJIA)

Sequences producing significant alignments

select all 5000 sequences selected

Description	Score	Max Total	Query E	Per. Ident	Acc. Len	Accession
slu04616.ayrbaase-related.protein (Methanocaldococcus sp.)	1027	1027	0.0	97.85%	510	WP_456373703.1
slu04616.ayrbaase-related.protein (Methanocaldococcus infernalis)	947	947	0.0	91.18%	508	WP_013100272.1
slu04616.ayrbaase-related.protein (Methanocaldococcus indicus)	916	916	0.0	88.43%	503	WP_423739241.1
slu04616.ayrbaase-related.protein (Methanocaldococcus rufus)	913	913	0.0	89.02%	503	WP_004992021.1
slu04616.ayrbaase-related.protein (Methanocaldococcus sp.)	910	910	0.0	98.24%	494	GG020282.1
slu04616.ayrbaase-related.protein (Methanocaldococcus slovac)	889	889	0.0	87.03%	510	WP_133007298.1
slu04616.ayrbaase-related.protein (Methanocaldococcus sp.)	0.0	86.44%	510	86.44%	510	WP_015870869.1
slu04616.ayrbaase-related.protein (Methanocaldococcus sp.)	0.0	85.85%	510	85.85%	510	WP_015870869.1
slu04616.ayrbaase-related.protein (Methanocaldococcus sp.)	0.0	85.07%	510	85.07%	510	WP_013125009.1
slu04616.ayrbaase-related.protein (Methanocaldococcus sp.)	0.0	87.23%	510	87.23%	510	WP_007546828.1

BIOS477/877 L12 - 30

30