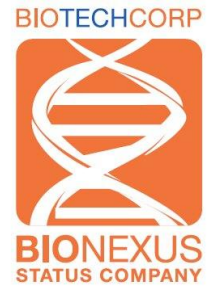




Malaysian
Genomics
Resource
Centre



SynaBlast-Mega™

User Guide

Table of Contents

1. SynaBlast-Mega.....	1
2. Operation Flow	1
3. Parameter Descriptions	2
4. Previewing and Retrieving Results.....	4
5. Output Format.....	6
6. Nucleotide-nucleotide.....	10
7. Protein-protein	11
8. Nucleotide-protein	12

All titles and intellectual property rights in and to Synamatix, SynaBASE, SynaBlast, and the accompanying printed materials (save where expressly indicated otherwise) are owned by Synamatix Sdn Bhd. No unauthorised reproduction of this document or any part therein is allowed without the permission of the owner. All third party intellectual property rights remain the property of their respective owners.

1. SynaBlast-Mega

SynaBlast-Mega™ is a multiple sequence alignment tool that enables users to search for multiple protein and nucleotide sequences. This tool makes it easier for researchers to analyse sequences that they have submitted in bulk. Comprehensive alignment options are provided and the final result is presented in various alignment output formats. SynaBlast-Mega allows nucleotide-nucleotide, protein-protein, and nucleotide-protein searches in SynaBASE. The application accepts FASTA format as its query sequence input.

2. Operation Flow

SYNABLAST-MEGA Search Result (by Job ID)

Query

1 Query type Nucleotide - Nucleotide

2 Query sequences Query Sequence Input (FASTA format) Total query : 0

Query input (FASTA format) Clear

(Paste query sequence(s) here)

Job title Job ID: 488233086121

Target database

3 SynaBASE H_sapiens Genome NCBI-GRC Release 37.1 Word size 15

Report options

4 Bit score threshold 0 E-value cut-off 10.0 Format Pairwise alignment view Maximum results reported per query 10

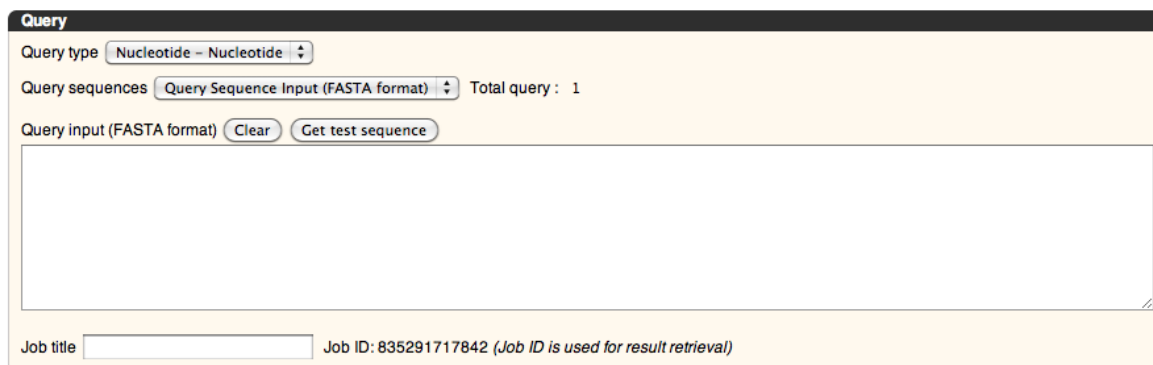
Advanced options

5 [- Show](#)

6 Submit Be notified by email (An email will be sent to the address provided when the results are available)

1. Select a 'Query type'
2. Enter 'Query sequences' to search
3. Choose a SynaBASE
4. Set 'Report options'
5. Set 'Advanced options' (optional)
6. Run SynaBlast-Mega. Select the checkbox for email notification

3. Parameter Descriptions



The screenshot shows a web interface titled "Query". It features a dropdown menu for "Query type" set to "Nucleotide - Nucleotide". Below it, "Query sequences" is set to "Query Sequence Input (FASTA format)" with a "Total query : 1" indicator. A "Query input (FASTA format)" text area is present, with "Clear" and "Get test sequence" buttons. At the bottom, there is a "Job title" input field and a "Job ID: 835291717842 (Job ID is used for result retrieval)" label.

Query type

Specify the query type. Choose one of the three options: nucleotide-nucleotide, nucleotide-protein or protein-protein searches.

Query sequences

Submit query sequence(s). Choose one of the following modes to submit a query sequence:

1. **Query Sequence Input (FASTA format)**
Paste a sequence (in FASTA or raw format).
2. **Upload Query File**
Upload a sequence (in FASTA or raw format) contained in a file on your computer.
3. **Sequence Search by Keyword**
Select a SynaBASE and search for a sequence by typing a keyword or sequence header expression that pulls out the reference sequence(s).

Clear

Used to delete content in the **Query input** box.

Get test sequence

Used to select a sample sequence for showcasing the capability of the application(s).



The screenshot shows a "Target database" section with a dropdown menu for "SynaBASE" set to "H_sapiens Genome NCBI-GRC Release 37.1" and a "Word size" input field set to "15".

SynaBASE

Select a suitable target SynaBASE based on query type. Each SynaBASE has its own set of default search parameters. These can be changed to suit the user's specific requirements.

Word size

Word size is the minimum initial pattern length to search for matches in a SynaBASE. Generally, short word sizes return more results, requiring a longer search time.



Report options

Bit score threshold E-value cut-off Format Maximum results reported per query

Bit score threshold

Alignments with bit scores lower than this value will not be reported.

E-value cut-off

Alignments with an E-value (expectation value) greater than this will not be reported.

Format

Output format of submission results:

1. Pairwise alignment view (Default)
2. Hit table alignment view
3. Blat PSL format

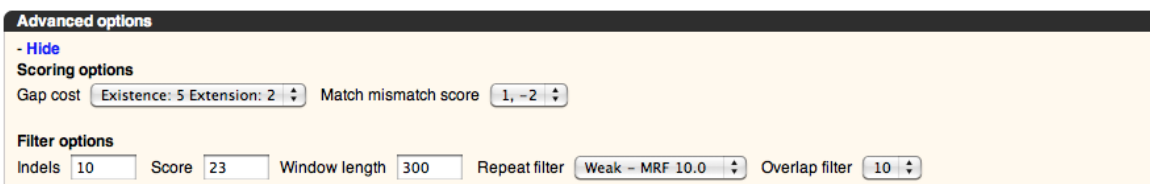
Maximum results reported per query

This setting limits the number of top scoring alignments that are reported per query.



Advanced options

[- Show](#)



Advanced options

[- Hide](#)

Scoring options

Gap cost Match mismatch score

Filter options

Indels Score Window length Repeat filter Overlap filter

Gap cost

This is the cost to create and extend a gap in an alignment.

Match mismatch score

A reward and penalty scoring system for matching and mismatching bases.

InDels

InDels are the allowable sizes of gaps in the searches.

Score

Score is the cumulative count of matching character patterns within a specified window length in an alignment.

Window length

Window length is the length of the query sequence to be used to initialise searches.

Repeat filter

The repeat filter removes high frequency patterns from the seeding process. Turning the filter off will allow the use of all patterns in the query in the seeding process and may return unwanted alignments to low complexity and repeats. Values are: Off 0.0, Weak 10.0, Low 3.0, Medium 2.0, Strong 1.0. For example, if the ‘Weak’ repeat filter is selected, patterns with frequency 10 times average are filtered out before seeding.

Overlap filter

This refers to the percentage of overlapping alignments to be filtered.

4. Previewing and Retrieving Results


Query summary

Job ID : 295870102094
 Query type : Nucleotide/Nucleotide
 Total query : 5
 SynaBASE : H_sapiens Genome NCBI-GRC Release 37.1
 Word size : 15

Report options	Bit score threshold	E-value cut-off	Format	Maximum results reported per query
	0	10.0	Pairwise alignment view	10

Scoring options	Gap existence	Gap extension	Match rewards	Mismatch penalty
	5	2	1	2

Filter options	Indels	Score	Window length	Repeat filter	Overlap filter
	10	23	300	Weak - Maximum Relative Frequency (MRF) 10.0	10

Submission progress :
 Completed!

The result file will be available online for 7 days, before being deleted.

Result preview (Please click on the "Download result file" button to get the complete result file)

```

# SynaBASE      : Synabase API Version: ; Build: Dec 14 2010, 01:51:40
# Database      : hsr_grc_ncbi-37.1_sb3.0.9
# Min. Length   : 15
# Max. Length   : 0
# Max. RelFreq  : 10.0
# Filter Length : 300

```

Query summary

Summary overview that includes the Job title and/or Job ID used for the submission, as well as the query parameters explained in Chapter 3.

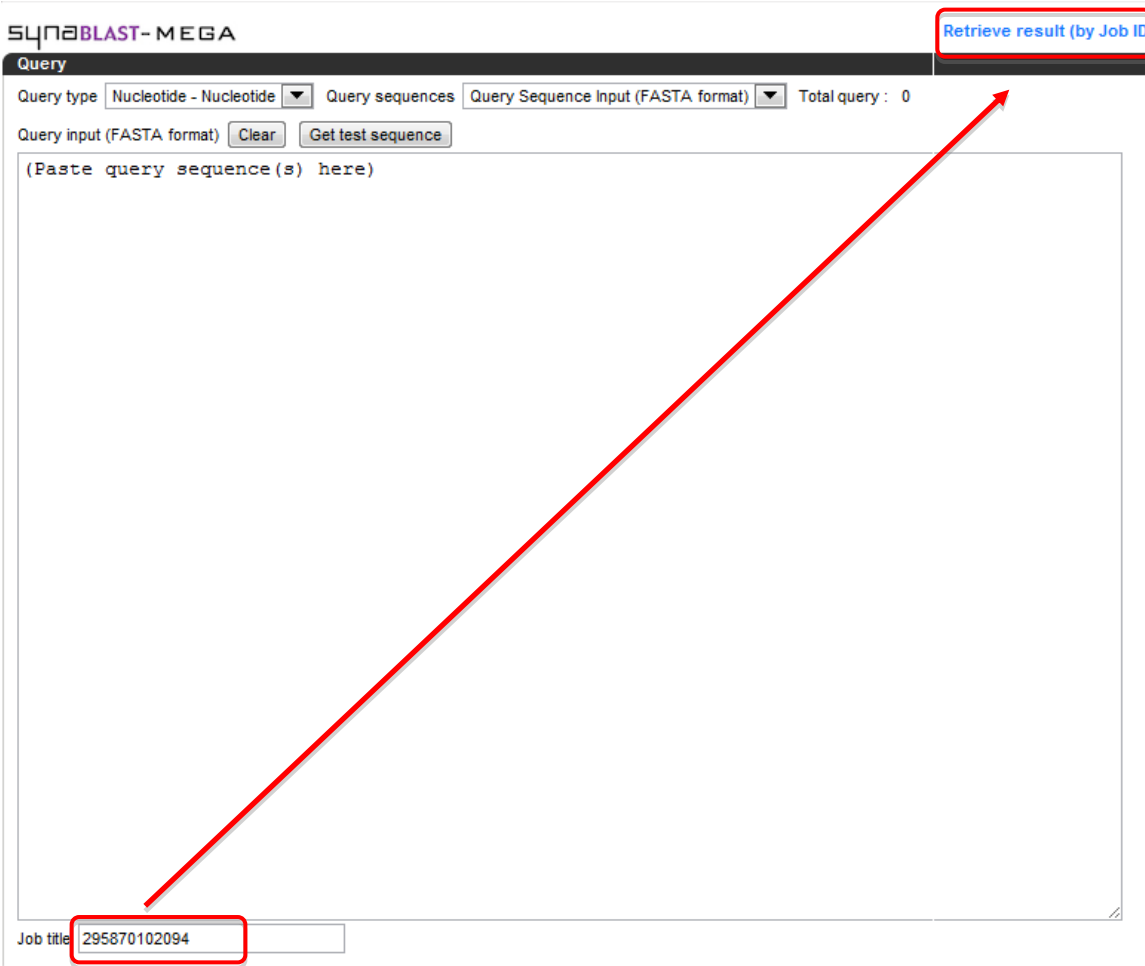
Submission progress

The status of the submission process can be viewed from the progress bar available on the result page. The submission process can be cancelled by clicking on the “Cancel Process” button found

below the submission progress bar. Upon completion the user can retrieve and download the result file by clicking the “Download Result File” button.

Result preview

This section provides a preview of the result file that is generated from the submission process. The 500-line overview enables the user to review a sample of the result file before downloading the full file.

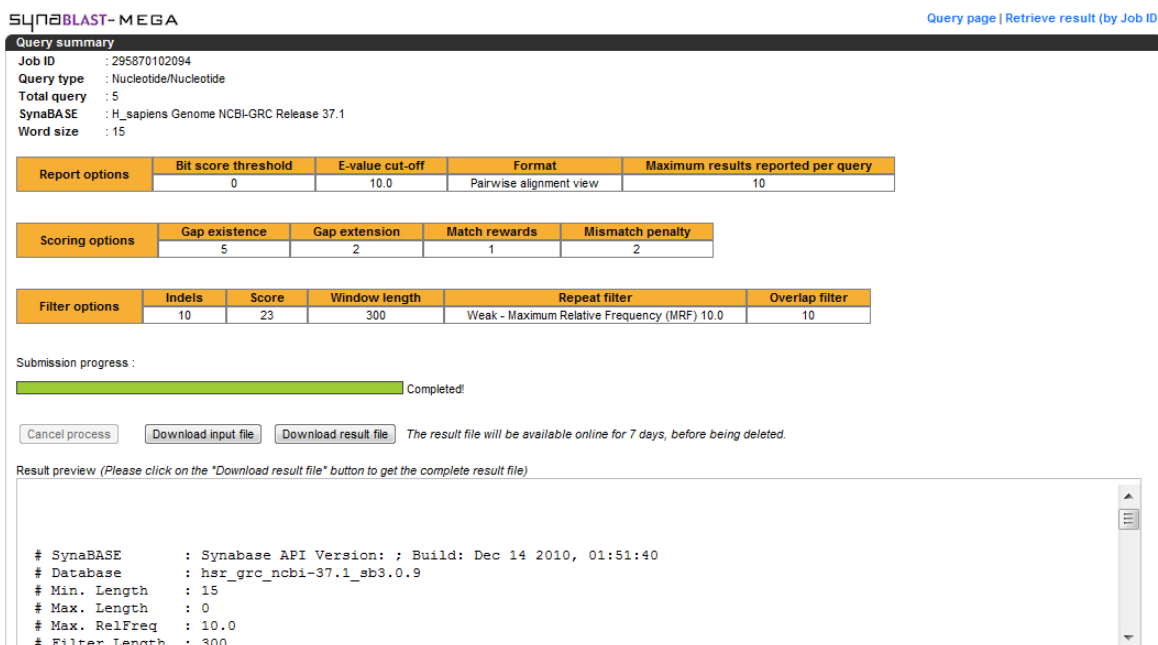
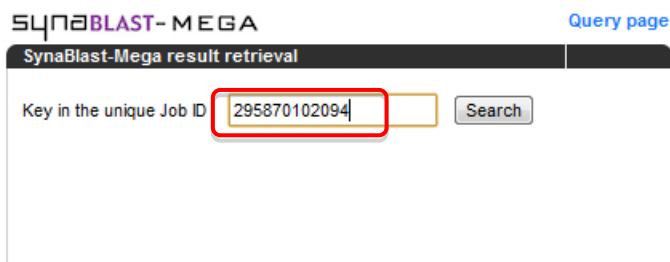


The screenshot displays the SYNABLAST-MEGA web interface. At the top left, the logo 'SYNABLAST-MEGA' is visible. Below it, the 'Query' section contains several input fields and buttons: 'Query type' (set to 'Nucleotide - Nucleotide'), 'Query sequences' (set to 'Query Sequence Input (FASTA format)'), and 'Total query : 0'. There are also buttons for 'Clear' and 'Get test sequence'. A large text area is provided for pasting query sequences. At the bottom left, the 'Job title' field contains the value '295870102094'. At the top right, a link labeled 'Retrieve result (by Job ID)' is highlighted with a red box. A red arrow originates from the 'Job title' field and points towards the 'Retrieve result (by Job ID)' link.

Result retrieval

The user is allowed to retrieve the query submission status or result by supplying a Job ID. The Job ID is an identification number that is generated once a query has been submitted. It is available from the query page.

If the web browser is closed after a query has been submitted, the user is still able to retrieve the query submission status or result via the “Retrieve Result (by Job ID)” link in the query page.



5. Output Format

The following examples illustrate the usage and output format of SynaBlast-Mega. Queries were searched against a SynaBASE containing the target sequences.

Pairwise alignment view

```
# SynaBASE      : Synabase API Version: ; Build: Dec 14 2010, 01:51:40
# Database      : sprt prtn upkb-release 2010 09 sb3.0.9
# Min. Length   : 6
# Max. Length   : 0
# Max. RelFreq  : 1.0
# Filter Length : 300
# Indels        : 5
# Score         : 15
# Fuzzy         : 0
# Overlap Filter : 10%
# Min. % Identity: no limit
# Min. BitScore : 5.0
# Max. E-value  : 10.0
# Search Mode   : PP
# NW            : on
```

```
# Scoring Matrix      : BLOSUM62
# Gap Opening Penalty : 11
# Gap Extension Penalty: 1
# Extension Cutoff    : 20
# Output Format       : pairwise
# Strand              : F+ only
# Max. Output per Query: 5
# Min. alignment length: no limit
# Major Min. identity: no limit
SynaSearch 1.4.1 [2010/02/08]

Reference: Synamatix Sdn Bhd Application Development,
"SynaSearch, a database search and alignment tool using SynaBASE",
All rights reserved.
Database: Swiss-Prot UPKB Release 2010 09
        519,348 sequences; 183,273,162 total letters

Query= sp|O34788|BDHA_BACSU (R,R)-butanediol dehydrogenase OS=Bacillus subtilis GN=bdhA
PE=3 SV=1 |kw|Complete proteome;Metal-binding;NAD;Oxidoreductase;Zinc;
(346 letters)

Sequences producing significant alignments:
Score      E
(Bits)    Value

sp|O34788|BDHA_BACSU (R,R)-butanediol dehydrogenase OS=Bacillu... 687 0E0
sp|Q65JE7|TDH_BACLD L-threonine 3-dehydrogenase OS=Bacillus li... 119 2.245E-26

>sp|O34788|BDHA_BACSU (R,R)-butanediol dehydrogenase OS=Bacillus subtilis GN=bdhA PE=3 SV=1
|kw|Complete proteome;Metal-binding;NAD;Oxidoreductase;Zinc;
Length = 346

Score = 687 bits (1772), Expect = 0E0
Identities = 346/346 (100%), Positives = 346/346 (100%)

Query: 1  MKAARWHNQKDIRIEHIEEPKTEPGKVKIKVVKWCGICGSDLHEYLGGPFIPIFVDKPHPLT 60
Sbjct: 1  MKAARWHNQKDIRIEHIEEPKTEPGKVKIKVVKWCGICGSDLHEYLGGPFIPIFVDKPHPLT 60

Query: 61  NETAPVTMGHEFSGEVVEVGEVGVENYKVGDRVVVEPIFATHGHQGAYNLDEQMGFLGLAG 120
Sbjct: 61  NETAPVTMGHEFSGEVVEVGEVGVENYKVGDRVVVEPIFATHGHQGAYNLDEQMGFLGLAG 120

Query: 121 GGGGFSEYVSVDEELLFKLPDELSYEQGALVEPSAVALYAVRSSKLGKAGDKAAVFGCGPI 180
Sbjct: 121 GGGGFSEYVSVDEELLFKLPDELSYEQGALVEPSAVALYAVRSSKLGKAGDKAAVFGCGPI 180

Query: 181 GLLVIEALKAAGATDIYAVELSPERQQKAEELGAIIVDPSKTDDVVAEIAERTGGGV DVA 240
Sbjct: 181 GLLVIEALKAAGATDIYAVELSPERQQKAEELGAIIVDPSKTDDVVAEIAERTGGGV DVA 240

Query: 241 FEVTGVPVVLQAIQSTTIAGETVIVSIWEKGAEIHPNDIVIKERTVKGIIIGYRDIFPAV 300
Sbjct: 241 FEVTGVPVVLQAIQSTTIAGETVIVSIWEKGAEIHPNDIVIKERTVKGIIIGYRDIFPAV 300

Query: 301 LSLMKEGYFSADKLVTKKIVLDDLIEEGFGALIKEKSQVKILVRPN 346
Sbjct: 301 LSLMKEGYFSADKLVTKKIVLDDLIEEGFGALIKEKSQVKILVRPN 346

>sp|Q65JE7|TDH_BACLD L-threonine 3-dehydrogenase OS=Bacillus licheniformis (strain DSM 13 /
ATCC 14580) GN=tdh PE=3 SV=1 |kw|Complete proteome;Cytoplasm;Metal-
binding;NAD;Oxidoreductase;Zinc;
Length = 346

Score = 119 bits (298), Expect = 2.245E-26
Identities = 86/287 (29%), Positives = 147/287 (51%), Gaps = 11/287 (3%)

Query: 65  PVTMGHEFSGEVVEVGEVGVENYKVGDRVVVEPIFATHGH----QGAYNLDEQMGFLGLAG 120
P  GHEFSGEVV+VGE V K G+ V E G ++ ++ LG+
Sbjct: 62  PYVFGHEFSGEVVQVGENVTTVKEGEYVSAETHIVCGKCLPCLTGKEHVCKKTLILGV-D 120

Query: 121 GGGGFSEYVSVDEELLFKLPDELSYEQGALVEPSAVALYAVRSSKLGKAGDKAAVFGCGPI 180
G F+EYV + ++K P + + ++ EP A++ V + + AG K AV GCGPI
Sbjct: 121 TDGCF AEYVKMPAANIWKNPAGMPEDLASIQEPLGNAVHTVLTG-MTAGVKVAVVCGCGPI 179
```

```

Query: 181 GLLVIEALKAAAGATDIYAVELSPERQQAEEELGAIIVDPSKTDDVVAEIAERTGG-GVDV 239
      GL+ + KA+GA + A++ + R A ++GA + + +D + ++ T G G D+
Sbjct: 180 GLMAVAVAKASGAAQVIAIDKNEYRLDLALQMGATDIISVEKEDPLKNVSALTNGEGADL 239

Query: 240 AFEVTGVPVVLRLQAIQSTTIAGETVIVSIWEKGAEIH-PNDIVIKERTVKGIIIGYR--DI 296
      E++G P +RQ+++ G ++S+ E I NDIV K TV+GI G + +
Sbjct: 240 VCEMSGHPTAIRQSLKMAANGGRVHVLVSLPEHPVCIDMTNDIVFKGLTVQGITGRKMFET 299

Query: 297 FPAVLSLMKEGYFSADKLVTKKIVLDDLIIEGFGALIKEKSQVKILV 343
      + V L++ G ++T + +++ E+GF + K + +L+
Sbjct: 300 WRQVSGLLQSGTIQIKPVITHRFPMEEF-EKGFELMRKGQCGKVVLI 345

```

```

Database: Swiss-Prot UPKB Release 2010_09
Number of letters in database: 183273162
Number of sequences in database: 519348

```

Gapped

```

Lambda      K      H
0.267      0.0410  0.140

```

Matrix: BLOSUM62

```

Gap Penalties: Existence: 11, Extension: 1
Number of Sequences: 519348
Length of database: 183273162

```

```

# Total reads      = 6
# Unique hit       = 0, 0.0%
# Multiple hits    = 6, 100.0%
# Repeat hits      = 0, 0.0%
# Missed hits      = 0, 0.0%
#
# Elapsed time     : 130 milliseconds

```

Hit table alignment view

```

# SynaBASE      : Synabase API Version: ; Build: Dec 14 2010, 01:51:40
# Database      : sprt_prtn_upkb-release_2010_09_sb3.0.9
# Min. Length   : 6
# Max. Length   : 0
# Max. RelFreq  : 10.0
# Filter Length : 300
# Indels        : 5
# Score         : 9
# Fuzzy         : 0
# Overlap Filter : 10%
# Min. % Identity: no limit
# Min. BitScore : 0.0
# Max. E-value  : 10.0
# Search Mode   : PP
# NW            : on
# Scoring Matrix : BLOSUM62
# Gap Opening Penalty : 11
# Gap Extension Penalty: 1
# Extension Cutoff : 20
# Output Format   : hit table
# Strand         : F+ only
# Max. Output per Query: 10
# Min. alignment length: no limit
# Major Min. identity: no limit

# SynaSearch 1.4.1
# Query: >sp|O34788|BDHA_BACSU (R,R)-butanediol dehydrogenase OS=Bacillus subtilis GN=bdhA
PE=3 SV=1 |kw|Complete proteome;Metal-binding;NAD;Oxidoreductase;Zinc;
# Database: sprt_prtn_upkb-release_2010_09_sb3.0.9
# Fields: Query id, Target id, % identity, alignment length, mismatches, gap openings, q.
start, q. end, s. start, s. end, e-value, bit score
sp|O34788|BDHA_BACSU      sp|O34788|BDHA_BACSU      100.00  346      0      0

```

```

1          346          1          346          0E0          687
sp|O34788|BDHA_BACSU          sp|Q06004|DHSO_BACSU          36.06          355          151          7
1          346          9          352          6.073E-48 191
sp|O34788|BDHA_BACSU          sp|P77280|YDJJ_ECOLI          32.14          336          147          9
12         339          14         338          1.148E-38 160
sp|O34788|BDHA_BACSU          sp|Q1PSI9|IDND_VITVI          29.28          345          159          5
10         342          29         363          1.658E-37 156
sp|O34788|BDHA_BACSU          sp|Q58D31|DHSO_BOVIN          30.09          329          148          5
7          326          14         332          6.299E-37 154
sp|O34788|BDHA_BACSU          sp|P07846|DHSO_SHEEP          29.36          327          151          3
7          326          13         330          6.299E-37 154
sp|O34788|BDHA_BACSU          sp|Q07993|XYL2_YEAST          33.61          241          101          3
64         293          63         301          3.954E-31 135

# SynaSearch 1.4.1-SNAPSHOT
# Query: >sp|P82974|AMPD CITFR 1,6-anhydro-N-acetylmuramyl-L-alanine amidase AmpD
OS=Citrobacter freundii GN=ampD PE=1 SV=1 |kw|3D-structure;Cell wall
biogenesis/degradation;Cytoplasm;Hydrolase;Metal-binding;Zinc;
# Database: sprt_prtn_upkb-release_2010_09_sb3.0.9
# Fields: Query id, Target id, % identity, alignment length, mismatches, gap openings, q.
start, q. end, s. start, s. end, e-value, bit score
sp|P82974|AMPD CITFR          sp|P82974|AMPD CITFR          100.00          187          0          0
1          187          1          187          6.392E-111          397
sp|P82974|AMPD CITFR          sp|P30013|AMPD_SALTY          89.84          187          12          0
1          187          1          187          1.131E-99 360
sp|P82974|AMPD CITFR          sp|P13016|AMPD_ECOLI          88.83          179          12          0
1          179          1          179          2.44E-94 342
sp|P82974|AMPD CITFR          sp|P82973|AMPD_ENTCL          80.21          187          18          0
1          187          1          187          3.647E-90 328

# Total reads          = 6
# Unique hit          = 0,          0.0%
# Multiple hits       = 5,          83.3%
# Repeat hits         = 1,          16.7%
# Missed hits         = 0,          0.0%
#
# Elapsed time        : 233 milliseconds

```

Blat PSL format

```

# SynaBASE          : Synabase API Version: ; Build: Dec 14 2010, 01:51:40
# Database          : sprt_prtn_upkb-release_2010_09_sb3.0.9
# Min. Length      : 6
# Max. Length      : 0
# Max. RelFreq     : 10.0
# Filter Length    : 300
# Indels           : 5
# Score            : 9
# Fuzzy           : 0
# Overlap Filter   : 10%
# Min. % Identity  : no limit
# Min. BitScore    : 0.0
# Max. E-value     : 10.0
# Search Mode      : PP
# NW               : on
# Scoring Matrix   : BLOSUM62
# Gap Opening Penalty : 11
# Gap Extension Penalty: 1
# Extension Cutoff : 20
# Output Format     : psl
# Overlap Chaining : 5%
# Strand           : F+ only
# Max. Output per Query: 10
# Min. alignment length: no limit
# Major Min. identity: no limit

match      mis-      rep.      N's      Q gap      Q gap      T gap      T gap      strand      Q
Q          Q          Q          T          T          T          T          block      blockSizes
qStarts   tStarts
size      match      match      name      count      bases      count      bases      name
          start      end

```

```

-----
346      0      0      0      0      0      0      0      +
        sp|O34788|BDHA_BACSU      346      1      346      sp|O34788|BDHA_BACSU
346      1      346      1      346,      1,      1,
128     151      0      0      0      0      0      0      +
        sp|O34788|BDHA_BACSU      346      1      346      sp|Q06004|DHSD_BACSU
353      9      352      1      355,      1,      9,
108     147      0      0      0      0      0      0      +
        sp|O34788|BDHA_BACSU      346      12      339      sp|P77280|YDJJ_ECOLI
347      14      338      1      336,      12,      14,
101     159      0      0      0      0      0      0      +
        sp|O34788|BDHA_BACSU      346      10      342      sp|Q1PSI9|IDND_VITVI
366      29      363      1      345,      10,      29,
99      148      0      0      0      0      0      0      +
        sp|O34788|BDHA_BACSU      346      7      326      sp|Q58D31|DHSD_BOVIN
356      14      332      1      329,      7,      14,
98      133      0      0      0      0      0      0      +
        sp|O34788|BDHA_BACSU      346      11      293      sp|P22144|XYL2_PICST
363      14      301      1      298,      11,      14,
97      132      0      0      0      0      0      0      +
        sp|O34788|BDHA_BACSU      346      64      345      sp|Q9Z9U1|DHSD_BACHD
343      54      341      1      293,      64,      54,
96      151      0      0      0      0      0      0      +
        sp|O34788|BDHA_BACSU      346      7      326      sp|P07846|DHSD_SHEEP
354      13      330      1      327,      7,      13,
93      130      0      0      0      0      0      0      +
        sp|O34788|BDHA_BACSU      346      65      343      sp|A7Z4X0|TDH_BACA2347
62      346      1      287,      65,      62,
89      113      0      0      0      0      0      0      +
        sp|O34788|BDHA_BACSU      346      65      307      sp|Q8R7K0|TDH_THETN347
62      310      1      251,      65,      62,

# Total reads      = 6
# Unique hit      = 0,      0.0%
# Multiple hits   = 5,      83.3%
# Repeat hits     = 1,      16.7%
# Missed hits     = 0,      0.0%
#
# Elapsed time    : 253 milliseconds

```

6. Nucleotide-nucleotide

Nucleotide-nucleotide searches are used to search and identify similar sequences in databases, and to find the location of sequences within a genome. The suggested parameter settings for general nucleotide-nucleotide searches are shown below.

Recommended parameters

Query type	: Nucleotide-nucleotide
Query sequence	: Nucleotide sequence in FASTA format
Target SynaBASE	: Nucleotide SynaBASE
Word size	: Default value according to selected database

Report options

Bit score threshold	: 0 (Default)
E-value cut-off	: 10.0 (Default)

Format : Pairwise alignment view (Default)
Max result reported : 10 (Default)

Advanced options

Gap cost : Existence: 5 Extension: 2 (Default)
Match mismatch score : 1, -2 (Default)
InDels : 10 (Default)
Score : Default value according to selected database
Window length : 300 (Default)
Repeat filter : Weak (Default)
Overlap filter : 10 (Default)

7. Protein-protein

In SynaBlast-Mega, protein-protein alignments are used to search and identify similar protein sequences in a SynaBASE.

Recommended parameters

Query type : Protein-protein
Query sequence : Protein sequence in FASTA format
Target SynaBASE : Protein SynaBASE
Word size : Default value according to selected database

Report options

Bit score threshold : 0 (Default)
E-value cut-off : 10.0 (Default)
Format : Pairwise alignment view (Default)
Max result reported : 10 (Default)

Advanced options

Gap cost : Existence: 11 Extension: 1 (Default)
Matrix : BLOSUM62
InDels : 5 (Default)
Score : Default value according to selected database
Window length : 300 (Default)

Repeat filter : Weak (Default)
Overlap filter : 10 (Default)

8. Nucleotide-protein

Nucleotide-protein alignments are used to search and identify similar sequences by translating nucleotide queries into amino acid sequences in both the forward and reverse directions in all three reading frames using the standard translation table. Each amino acid sequence is then aligned with sequences stored in a protein SynaBASE.

Recommended parameters

Query type : Nucleotide-protein
Query sequence : Nucleotide sequence in FASTA format
Target SynaBASE : Protein SynaBASE
Word size : Default value according to selected database

Report options

Bit score threshold : 0 (Default)
E-value cut-off : 10.0 (Default)
Format : Pairwise alignment view (Default)
Max result reported : 10 (Default)

Advanced options

Gap cost : Existence: 11 Extension: 1 (Default)
Match mismatch score : 1, -2 (Default)
InDels : 10 (Default)
Score : Default value according to selected database
Window length : 300 (Default)
Repeat filter : Weak (Default)
Overlap filter : 10 (Default)