



SmartBLAST: A Tool For Rapid Protein Identification

<https://blast.ncbi.nlm.nih.gov/smartblast/>

Quickly identifying close relatives of a protein query and evaluating their phylogenetic relationship

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

Introduction

SmartBLAST is a service that quickly finds closest relatives to a protein query and evaluates the phylogenetic relationship among the query and matched sequences. It does so by finding up to three best matches from the non-redundant protein database (nr) along with the two best matches from the complete proteomes of well-studied reference species (landmark dataset). If possible, the two matches from reference species will be from different organisms. In addition, SmartBLAST presents Conserved Domain Database matches to the query. A multiple sequence alignment based phylogenetic tree for the matched sequences is also part of the results presentation. Note that SmartBLAST is under active development and may change with little notice.

Workflow Behind the Service

SmartBLAST uses a combination of BLAST and a multiple sequence alignment algorithm. SmartBLAST searches the query against the nr protein database with an optimized version of BLAST targeted to find closely related sequences. It also simultaneously searches the landmark dataset (<http://bit.ly/2aOVhPO>, see “Search sets”) with BLASTP. Finally, it performs a multiple sequence alignment on the six sequences (the query plus up to five matched subject sequences) using the COBALT multiple sequence alignment tool. The multiple sequence alignment and BLASTP serve different but complementary roles: BLASTP identifies database sequences similar to the query and calculates only pairwise similarities between the query and individual database sequences. The COBALT’s multiple alignment compares all sequences to each other and produces an optimal alignment between all sequences, which is ideal for the presentation of a phylogenetic tree and shows the relationship among a set of sequences. The optimization of the procedure combined with the reduced landmark dataset make the SmartBLAST search extremely fast and ideal for initial screening of unknown protein sequences.

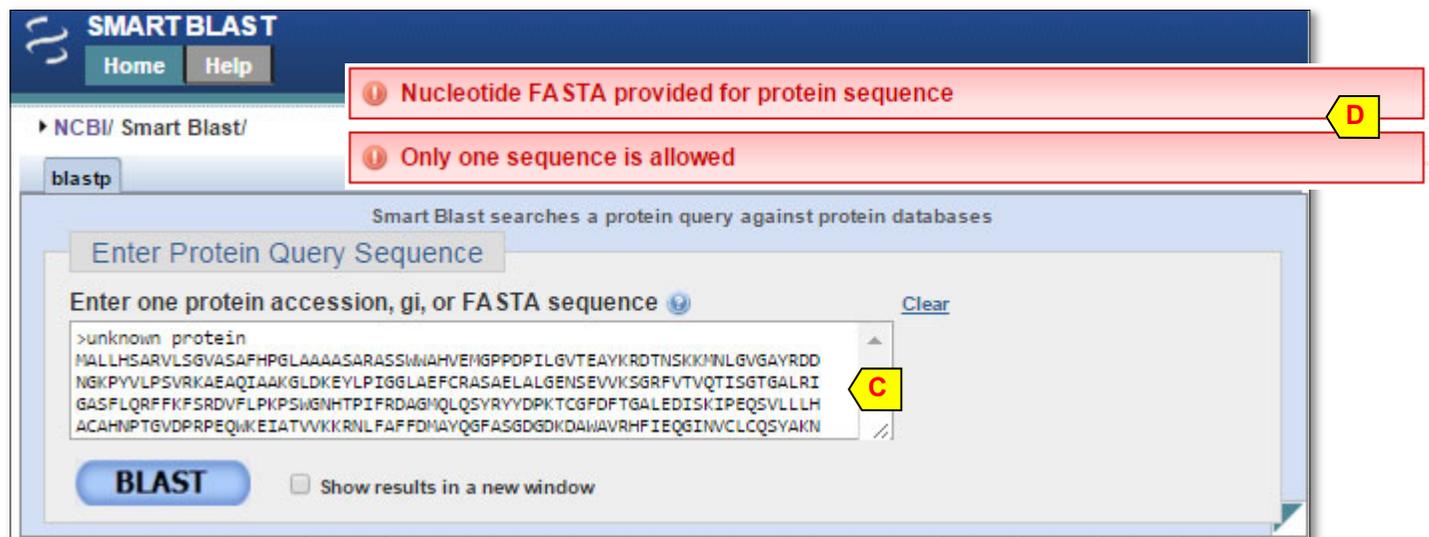
Access to the SmartBLAST Search Page

The BLAST homepage provides a convenient link to SmartBLAST (A) in the “Specialized searches” section.

The actual search page has a very simple interface (B).

Paste in a query protein sequence in a supported format, such as FASTA (C), in the search box and press “Submit” to initiate a search.

Note that SmartBLAST accepts only a single protein sequence as input and will reject input that consists of multiple protein sequences or nucleotide sequence (D).



The Results Page

SmartBLAST results contain three sections: Summary, Descriptions, and Alignments. The example below (A) has the Descriptions and Alignments sections closed (B) to emphasize the graphically enriched Summary.

Summary

The Summary presents only the top five matches from different organisms. This succinct set of best hits from a representative group of organisms allows SmartBLAST to show the relationships among the sequence, in the form of a phylogenetic tree (C), and the alignment coverage in schematic diagram form (D). This section also displays the sequence titles (E) and conserved domain(s) identified (F). For input in FASTA format, SmartBLAST will map it to existing proteins in the database. A successful mapping enables SmartBLAST to get a descriptive title from the RefSeq entry and assign taxonomic identification (G, as in this example). This section is highly interactive and it contains embedded popups in the titles, a schematic diagram depiction of alignments and CDD matches, as well as a link to the Legend.

Conserved Protein Domain Family
AAT_like

cd00609: **AAT_like**

Aspartate aminotransferase family. This family belongs to pyridoxal phosphate (PLP)-dependent aspartate aminotransferase superfamily (fold I). Pyridoxal phosphate combines with an alpha-amino acid to form a compound called a Schiff base or aldimine intermediate, which depending on the reaction, is the substrate in four kinds of reactions (1) racemization (redistribution of enantiomers), (2) decarboxylation (redistribution of enantiomers), (3) decarboxylation (redistribution of enantiomers), (4) decarboxylation (redistribution of enantiomers) reactions depending on the enzyme involved.

Click "more" to see details in the conserved domain record.

Use this link to submit your feedback to NCBI.

AAT_like: Aspartate aminotransferase family
Aspartate aminotransferase family. This family belongs to pyridoxal phosphate (PLP)-dependent aspartate aminotransferase superfamily (fold I). Pyridoxal phosphate (PLP) forms an internal aldimine bond (Schiff base linkage) with the catalytic residue lysine.

Structure: 1N1 A-L-Histidinol Phosphate Aminotransferase catalytic residue

NCBI/ Smart Blast/ Formatting Results - DRGCCSXU01T

Summary (A)

A concise summary of the three best matches in the sequence database together with the two best matches from well-studied reference species, showing phylogenetic relationships based on multiple sequence alignment and conserved protein domains.

Conserved domains for the AAT_like

house mouse
golden snub-nosed monkey
human
Bolivian squirrel monkey
rabbit (G)
American pika

aspartate aminotransferase ... (D) (F)
PREDICTED: aspartate am ...
aspartate aminotransferase ...
PREDICTED: aspartate am ...
Your query: PREDICTED: ...
PREDICTED: aspartate am ...

See full multiple alignment
Legend

Descriptions (B)

Alignments (B)

PREDICTED: aspartate aminotransferase, mitochondrial [Oryctolagus cuniculus]
430 aa protein
accession:ref|XP_002711597.1| GI:291390228
Sequence Identical proteins

A successful mapping allows SmartBLAST to replace the input FASTA with a RefSeq protein record with defined taxonomic identification and descriptive title.

Mouseover a link such as "See full multiple alignment" to see a popup. Clicking the link runs a Cobalt search to generate the full multiple alignment.

Mouseover "Legend" to see the meaning of color codings in the schematic depiction of the alignment.

Legend

- Query sequence
- Best matches
- Reference species
- Not aligned to query
- Gap

The Result Page (cont.)

Descriptions

This section lists the matches found by SmartBLAST in two separate tables. The “Best hits” table (A) contains only the five database matches displayed in the Summary section. The checkbox color (B) corresponds to the alignment cartoon to indicate the source of the hit (from the landmark dataset or NR). The “Additional BLAST Hits” table (C, shown in part) lists additional hits found in NR and the reference dataset. It supplements the “Best hits” set by expanding the taxonomic coverage and showing additional hits from the same organism that could represent isoforms, variants, or paralogs.

The tables are interactive with configurable columns and have links to specific sections of the Alignments as well as links to actual protein records.

Select: [All](#) [None](#) Selected: 1

[Alignments](#) [GenPept](#)

	Description	Max score	Total score	Query cover	E value	Ident
<input checked="" type="checkbox"/>	PREDICTED: aspartate aminotransferase, mitochondrial isoform X1 [Ochotona princeps]	875	875	100%	0.0	96%
<input type="checkbox"/>	PREDICTED: aspartate aminotransferase, mitochondrial [Rhinopithecus roxellana]	849	849	100%	0.0	95%
<input type="checkbox"/>	PREDICTED: aspartate aminotransferase, mitochondrial [Saimiri boliviensis boliviensis]	849	849	100%	0.0	95%

Available columns

- Description
- Max Score
- Total Score
- Coverage
- E-value
- IdentN
- Accession

[Restore Defaults](#) [Ok](#) [Cancel](#)

Select: [All](#) [None](#) Selected: 0

[Alignments](#) [GenPept](#)

	Description	Max score	Total score	Query cover	E value	Ident	Accession
<input type="checkbox"/>	PREDICTED: aspartate aminotransferase, mitochondrial isoform X1 [Ochotona princeps]	875	875	100%	0.0	96%	XP_004583963.1
<input type="checkbox"/>	PREDICTED: aspartate aminotransferase, mitochondrial [Rhinopithecus roxellana]	849	849	100%	0.0	95%	XP_010376502.1
<input type="checkbox"/>	PREDICTED: aspartate aminotransferase, mitochondrial [Saimiri boliviensis boliviensis]	849	849	100%	0.0	95%	XP_010329683.1
<input checked="" type="checkbox"/>	aspartate aminotransferase, mitochondrial [Mus musculus]	847	847	100%	0.0	94%	NP_034455.1
<input checked="" type="checkbox"/>	aspartate aminotransferase, mitochondrial isoform 1 precursor [Homo sapiens]	842	842	100%	0.0	94%	NP_002071.2

Select: [All](#) [None](#) Selected: 0

[Alignments](#) [GenPept](#)

	Description	Max score	Total score	Query cover	E value	Ident	Accession
<input type="checkbox"/>	PREDICTED: aspartate aminotransferase, mitochondrial [Nannospalax qalini]	847	847	100%	0.0	95%	XP_008832465.1
<input type="checkbox"/>	aspartate aminotransferase precursor [Mus musculus]	846	846	100%	0.0	94%	AAB91426.1
<input type="checkbox"/>	PREDICTED: aspartate aminotransferase, mitochondrial [Cavia porcellus]	846	846	100%	0.0	94%	XP_003472097.1
<input type="checkbox"/>	PREDICTED: aspartate aminotransferase, mitochondrial [Ursus maritimus]	845	845	100%	0.0	95%	XP_008700791.1
<input type="checkbox"/>	PREDICTED: LOW QUALITY PROTEIN: aspartate aminotransferase, mitochondrial-like [Colobus angolensis palliatus]	845	845	100%	0.0	95%	XP_011808813.1
<input type="checkbox"/>	PREDICTED: aspartate aminotransferase, mitochondrial [Felis catus]	843	843	100%	0.0	94%	XP_003998142.1
<input type="checkbox"/>	aspartate aminotransferase, mitochondrial [Canis lupus familiaris]	842	842	100%	0.0	94%	NP_001273953.1
<input type="checkbox"/>	PREDICTED: aspartate aminotransferase, mitochondrial [Ailuropoda melanoleuca]	839	839	100%	0.0	93%	XP_002912325.1
<input type="checkbox"/>	PREDICTED: aspartate aminotransferase, mitochondrial [Microcebus murinus]	838	838	100%	0.0	95%	XP_012599398.1
<input type="checkbox"/>	PREDICTED: aspartate aminotransferase, mitochondrial isoform X1 [Leptonychotes weddelli]	837	837	100%	0.0	93%	XP_006732870.1
<input type="checkbox"/>	PREDICTED: aspartate aminotransferase, mitochondrial isoform X1 [Odobenus rosmarus divergens]	836	836	100%	0.0	93%	XP_004395809.1
<input type="checkbox"/>	PREDICTED: aspartate aminotransferase, mitochondrial [Ceratotherium simum simum]	835	835	100%	0.0	95%	XP_004431730.1
<input type="checkbox"/>	mitochondrial aspartate aminotransferase [Mus musculus]	842	842	100%	0.0	94%	AAA37265.1
<input type="checkbox"/>	unnamed protein product [Mus musculus]	838	838	100%	0.0	93%	BAE39800.1

Check one or more checkboxes to activate the GenPept link for batch retrieval of the protein records from the Entrez Protein database.

Mouseover a link to see its functions (yellow popup). Here, click the title to go to its alignment in Alignments section.

Customize the column displayed in a table by clicking the gear icon, then select from the popup.

