

1. Purpose

This document describes the mechanism and process by which SwissProt hits by BlastP are generated for the proteins stored in BioHealthBase.

2. Method Description

The Basic Local Alignment Search Tool (BLAST) finds regions of local similarity between sequences. The program compares nucleotide or protein sequences to sequence databases and calculates the statistical significance of matches. BLAST can be used to infer functional and evolutionary relationships between sequences as well as help identify members of gene families. BlastP is one of a suite of algorithms offered in NCBI's BLAST software, and is specially designed for protein-protein comparisons.

BLAST software is obtained from <ftp://ftp.ncbi.nih.gov/blast>.

Version 2.2.13 or later are installed and used locally at BioHealthBase for the generation of BlastP data.

3. Input Data Preparation

Protein sequences of various organisms are extracted from BioHealthBase data warehouse and formatted appropriately.

UniProt-SwissProt is periodically downloaded from European Bioinformatics Institute at <ftp://ftp.ebi.ac.uk/pub/databases/swissprot/release/>. Formatdb from NCBI Blast suite is used to prepare the SwissProt data as Blast indices.

4. Output Data Post-Processing and Display

BlastP outputs are parsed using three scripts based on BioPerl to extract hit-level and HSP-level information and stored to BioHealthBase data warehouse.

Pre-computed BlastP results are delivered to end users both by text and by graph. Here is an example of displaying BlastP data in Gene Details page:

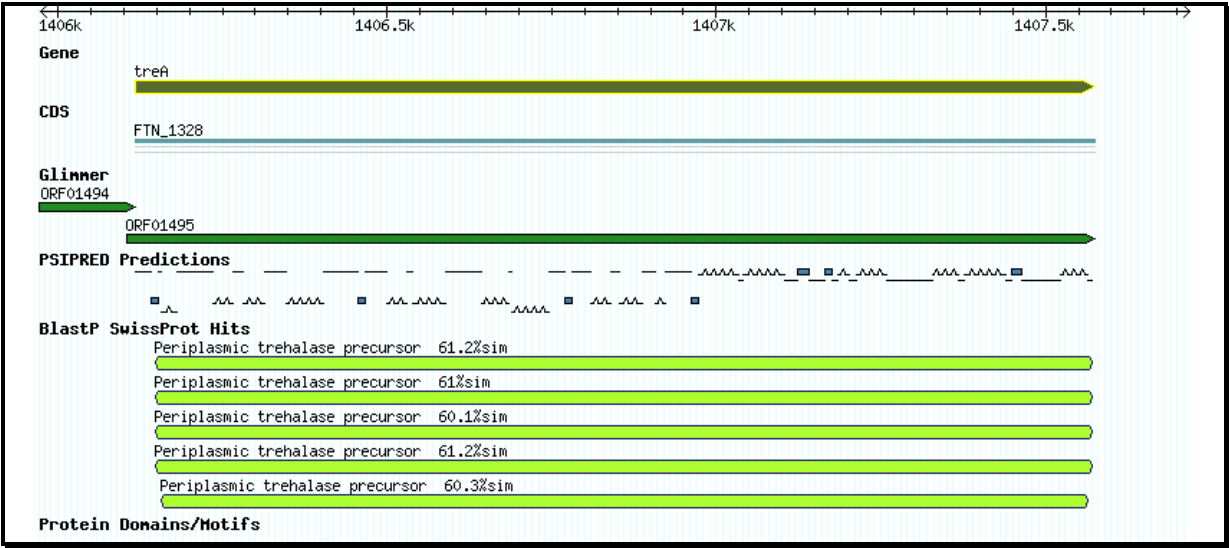
blastp: Swiss-Prot Searches

Top

View Full Report

Accession	% Sim	% S-Length	% Q-Length	Description	E-value
SP : P13482	61.18%	86.55%	97.93%	Periplasmic trehalase precursor Escherichia coli	1.0E-108
SP : P59765	60.13%	85.79%	97.93%	Periplasmic trehalase precursor Salmonella typhi	1.0E-108
SP : Q8CW46	60.97%	86.55%	97.93%	Periplasmic trehalase precursor Escherichia coli O6	1.0E-108
SP : Q8PPT1	60.34%	85.39%	96.90%	Periplasmic trehalase precursor Xanthomonas axonopodis pv. citri	1.0E-108
SP : Q83RP6	61.18%	86.55%	97.93%	Periplasmic trehalase precursor Shigella flexneri	1.0E-107
SP : Q8ZP20	59.28%	85.79%	97.93%	Periplasmic trehalase precursor Salmonella typhimurium	1.0E-107
SP : Q9I165	59.91%	89.17%	96.90%	Periplasmic trehalase precursor Pseudomonas aeruginosa	1.0E-105
SP : Q8XDH7	60.55%	86.45%	97.93%	Periplasmic trehalase precursor Escherichia coli O157:H7	1.0E-105
SP : Q8P519	59.70%	85.39%	96.90%	Periplasmic trehalase precursor Xanthomonas campestris pv. campestris	1.0E-105
SP : Q8ZLC8	61.83%	88.52%	96.90%	Cytoplasmic trehalase Salmonella typhimurium	1.0E-101

BlastP data are also displayed in the genome browser, as shown below:



5. References

- 1. Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. Basic Local Alignment Search Tool. J Mol Biol 215:403-410; 1990.**
- 2. Altschul SF, Madden TL, Schaffer AA, Zhang J, Zhang Z, Miller W, Lipman DJ. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Nucleic Acids Res 25:3389-3402; 1997.**