

HSSP curve

Protein Prediction 2

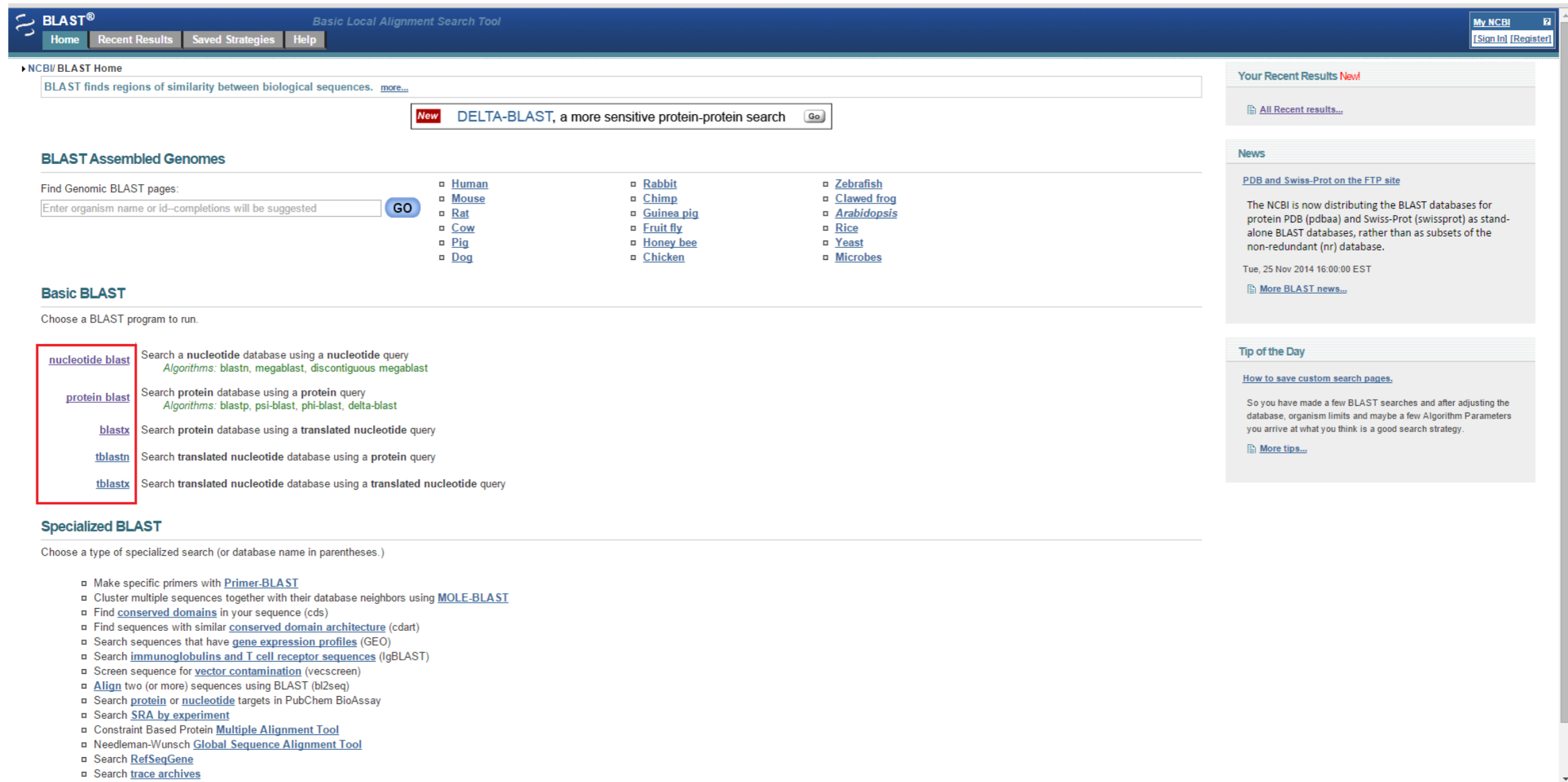
Students: Agon Lohaj, Bardh Lohaj

Mentors: PP2_CS_2014 mentors

Tool's objective

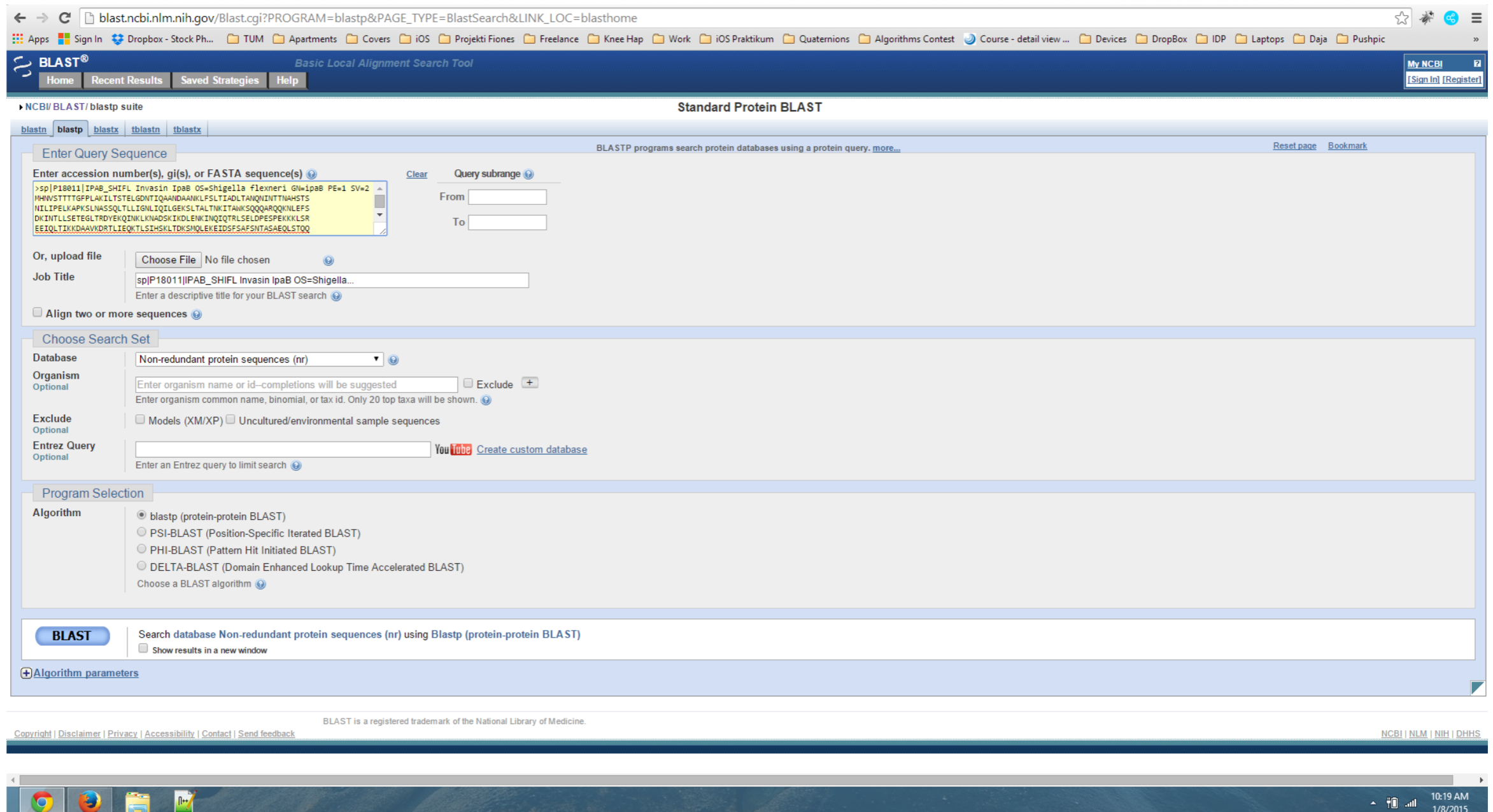
Visualize the HSSP curve and impose the BLAST-derived sequence alignments and allow the user to dynamically filter and export the data shown on the graph for better insights.

Input data (format: Blast output version BLASTP 2.2.29+)



The screenshot shows the NCBI BLAST website interface. At the top, there is a navigation bar with 'Home', 'Recent Results', 'Saved Strategies', and 'Help'. The main content area is titled 'NCBI BLAST Home' and includes a search bar with the text 'BLAST finds regions of similarity between biological sequences. more...'. Below this, there is a 'New' button and a link to 'DELTA-BLAST, a more sensitive protein-protein search'. The 'BLAST Assembled Genomes' section features a search box for organism names and a 'GO' button, with a list of organisms including Human, Mouse, Rat, Cow, Pig, Dog, Rabbit, Chimp, Guinea pig, Fruit fly, Honey bee, Chicken, Zebrafish, Clawed frog, Arabidopsis, Rice, Yeast, and Microbes. The 'Basic BLAST' section offers five options: 'nucleotide blast', 'protein blast', 'blastx', 'tblastn', and 'tblastx', each with a brief description and algorithms used. The 'Specialized BLAST' section provides a list of specialized search options such as 'Primer-BLAST', 'MOLE-BLAST', 'conserved domains', 'conserved domain architecture', 'gene expression profiles', 'immunoglobulins and T cell receptor sequences', 'vector contamination', 'Align', 'protein or nucleotide targets', 'SRA by experiment', 'Multiple Alignment Tool', 'Global Sequence Alignment Tool', 'RefSeqGene', and 'trace archives'. On the right side, there are sections for 'Your Recent Results', 'News' (with a link to 'PDB and Swiss-Prot on the FTP site'), and 'Tip of the Day' (with a link to 'How to save custom search pages').

Input data (format: Blast output version BLASTP 2.2.29+)



blast.ncbi.nlm.nih.gov/Blast.cgi?PROGRAM=blastp&PAGE_TYPE=BlastSearch&LINK_LOC=blasthome

Basic Local Alignment Search Tool

Standard Protein BLAST

Enter Query Sequence

Enter accession number(s), gi(s), or FASTA sequence(s)

```
>sp|P18011|IPAB_SHIFL Invasin IpaB OS=Shigella flexneri GN=ipaB PE=1 SV=2
HHWSTTTTGFPLAKILSTELGNTIQAANDAAKLFELTIADLTANQININTTNAHSTS
NLLIPELVKPKSLNASSQLTLIGNLIQTLGKSLTALTNITANKSQQAQQKNIKLEFS
DKINTLLSETEGLTRDYKQINKLNADSKIKDLINKINQITRLSELDPEPFEKKLKR
EEIQLTIKKDAAVKDRTLIEQKTLSTHSLKLDKSHOLEKEIDFSFAFNTASAEQLSTQQ
```

Or, upload file: Choose File (No file chosen)

Job Title: sp|P18011|IPAB_SHIFL Invasin IpaB OS=Shigella...

Choose Search Set

Database: Non-redundant protein sequences (nr)

Exclude: Models (XM/XP) Uncultured/environmental sample sequences

Program Selection

Algorithm: blastp (protein-protein BLAST)

BLAST Search database Non-redundant protein sequences (nr) using Blastp (protein-protein BLAST)

BLAST is a registered trademark of the National Library of Medicine.

Copyright | Disclaimer | Privacy | Accessibility | Contact | Send feedback

NCBI | NLM | NIH | DHHS

10:19 AM 1/8/2015

Input data (format: Blast output version BLASTP 2.2.29+)

The screenshot displays the NCBI BLAST web interface. At the top, the browser address bar shows `blast.ncbi.nlm.nih.gov/Blast.cgi`. The page title is "NCBI/BLAST/blastp suite/Formatting Results - AWJMCU8R01R".

Key search details include:

- Query ID:** lc|336404
- Description:** sp|P18011|IPAB_SHIFL Invasin IpaB OS=Shigella flexneri GN=ipaB PE=1 SV=2
- Molecule type:** amino acid
- Query Length:** 580
- Database Name:** nr
- Description:** All non-redundant GenBank CDS translations+PDB+SwissProt+PIR+PRF excluding environmental samples from WGS projects
- Program:** BLASTP 2.2.30+ [Citation](#)

The "Graphic Summary" section shows "Putative conserved domains have been detected, click on the image below for detailed results." It features a horizontal bar representing the query sequence (0-580) with a cyan bar indicating a "PRK15374" domain. Below this, a "Distribution of 100 Blast Hits on the Query Sequence" is shown as a heatmap. A color key for alignment scores is provided:

Score Range	Color
<40	Black
40-50	Blue
50-80	Green
80-200	Purple
>=200	Red

The heatmap shows numerous red bars (high scores) across the query sequence, with a concentration between positions 100 and 300.

Demo