



Practical Assignment

Module name: Sequence alignment theory and applications

Session name: BLAST algorithm

Trainer: Sonal Henson

Participant: <write your name here>

Date: <write today's date here>

BLAST to answer biological questions

Introduction

In this practical assignment you will use the BLAST programs to answer two biological questions:

- 1) *Detect frameshift mutations using sequence alignment.*
- 2) *Find structurally similar proteins that have very little sequence similarity.*

For each task, follow the instructions given and type your answers to the questions asked after the question.

Tools used in this session

NCBI [BLAST](#)

Please note

- **Hand-in information** If you are formally enrolled in the IBT course, please upload your completed assignment to the Vula 'Assignments' tab. Take note of the final hand-in date for each assignment, which will be indicated on Vula.

Task 1: Detecting Frameshift mutations

Task 1: instructions

1. *Search the sample sequence given below against the protein database (nr) using the BLASTX program. Answer the questions that follow.*

>sample_sequence

```
AGAAGAAGACATAGTAATTAGATCTGAAAATTTACGAACAATGCTAAAACCATAATAGTACAGCTGAAG
GAATCTATAAAAATTAATTGTACAAGACCCAACAACAATACAAGAAAAAGTATACCTATAGCAACGGGGG
GAGCAATTTATGCAACAGGAGACATAATAGGAGATATAAGACAAGCACATTGTAACCTTAGTAGAGACCA
ATGGGATAACACTTTAAGCCAGCTAGTTACAAAATAAGAGAACAATTTGGGAATAAACAATAGCCTTT
```



H3ABioNet

Pan African Bioinformatics Network for H3Africa

Introduction to Bioinformatics online course: IBT

```
AATCAATCCTCAGGAGGGGACCCAGAAATTGTAATGCACAGTTTTAATTGTGGAGGAGAATTTTTCTACT
GTAATACAACACAGCTGTTAATAGTACTTGGCCAATAAAAAAGTCTACTAACAAAACAGGAAC
TATCACACTCCCGTGCAGAATAAAACAAATTATAAACAGGTGGCAAGAAGTAGGAAAAGCAATGTATGCC
CCTCCCATCAAGGGACAAATTAGATGTTTCATCAAATATTACAGGGATATTCTTAACAAGAGATGGTGGTA
ACGCAAGCGATGAGACCGAGACCTTCAGACCTGGAGGAGGAAATA
```

2. For database hit AAL71600.1 which frame of the query sequence does alignment begin in?
3. At which nucleotide of the query sequence does the frame change?
4. Using what you learnt about navigating through the NCBI database, find the nucleotide sequence corresponding to the protein with the accession AAL71600.1. Describe the steps you took to find it.
5. Make a local alignment of the nucleotide sequence from 3. with the sample sequence. Download your alignment and paste it below.
To download your alignment or any search result, for that matter, click on 'Download' at the top of your results page. Select the format you would like to download your results in. Try the 'text' format for now. See screenshot below.

BLAST Results

[Edit and Resubmit](#) [Save Search Strategies](#) [▶ Formatting options](#) [▼ Download](#)

Download

Alignment

[Text](#) [XML](#) [ASN.1](#) [JSON](#) [Seq-align](#) [Hit Table\(text\)](#) [Hit Table\(csv\)](#) [Multiple-file XML2](#) [Single-file XML2](#) [Multiple-file JSON](#) [Single-file JSON](#) [SAM](#)

Paste your alignment below:

6. Highlight the nucleotide that is bringing about the frameshift in the alignment you have pasted in 4.
7. What other differences are there between the two nucleotide sequences? Do they also cause a frameshift? Do they change the amino acid?



Introduction to Bioinformatics online course: IBT

Task 2: Finding structurally similar proteins using PSI-BLAST

Task 2: instructions

Tyrosine tRNA ligase (TyrRS) and Tryptophan tRNA ligase (TrpRS) are structurally similar (Refs: [1](#), [2](#)). Given structural similarity you would expect to find sequence similarity. However, TyrRS and TrpRS share 13% sequence identity.

We will use PSI-BLAST to find the sequence of TyrRS.

1. Using a sequence database of your choice retrieve the protein sequence for *E. coli* Tyrosine tRNA ligase (TyrRS), alias Tyrosyl-tRNA synthetase (sp|P0AGJ9.2).
2. Open the [BLASTP](#) webpage on NCBI and paste the sequence you found in 1.
3. Submit a PSI-BLAST search against the SwissProt database, narrowing your search organism to Bacteria. Keep the PSI-BLAST threshold at 0.005.
4. In your search results the description section will be split in two tables. The top table will contain alignments with E-values below the cut-off and the lower table will show the hits above the E-value cut-off.
 - a. What range of the % identities do you see in the alignments in both the tables?
 - b. Do you get hits to proteins other than TyrRS in the first table?
 - c. Take a look at the Taxonomy report for this search. You will find a link to it just above the Graphic Summary section. Which organism has the lowest alignment score?
5. Run the 2nd iteration of PSI-BLAST using hits in the first table. This may take a while to run so please be patient. Do you get hits to TrpRS on the 2nd iteration? Using the Taxonomy report for the 2nd iteration determine:
 - a. if you get a hit to TrpRS of *E. coli*.
 - b. which is the evolutionarily closest bacterial species to *E. coli* you get a TrpRS hit for? Hint: look at the number of hits in the lineage report.
 - c. how long is the TrpRS protein for this species?



H3ABioNet

Pan African Bioinformatics Network for H3Africa

Introduction to Bioinformatics online course: IBT

6. Run a 3rd iteration this time including the TrpRS hits.
 - a. Do you get a hit to TrpRS of *E. coli*?
 - b. If yes, what is the accession number for it?
 - c. What % identity does the query sequence (*E. coli* TrpRS) align with it (*E. coli* TyrRS)?