

BLAST Method: Initialize Problem

PURPOSE: This Interactive teaches how the BLAST method determines which sequences are the best match to an input sequence. The unknown input sequence is call the QUERY. The BLAST method breaks up the QUERY into little bits, called “Words”, and then matches the little bits against other sequences, called the SUBJECTS.

The screenshot shows a web-based interface for BLAST. At the top right, a yellow box contains the text "The default scoring of matches." Below this are two columns of settings: "Length" and "Scores". Under "Length", there are three input fields: "Subject Size: 25 (25 to 50)", "Query Size: 10 (10 to 15)", and "Word Size: 5 (5 to 7)". Under "Scores", there are two input fields: "Match: 5" and "Mismatch: -4". Below these settings is a "Query:" input field containing "AGATTTATAA". A red arrow points from this field down to a "Query:" field where the sequence "AGATTTATAA" is displayed with each letter in a different colored box (A: black, G: green, A: green, T: purple, T: purple, T: purple, A: green, T: green, A: green, A: green). Below the query field are three buttons: "RANDOM QUERY", "SUBMIT", and "QUIZ MODE". At the bottom, a yellow box contains the text "In Concept Mode, click 'Create Words'", and below it is a "CREATE WORDS" button.

Set length of SUBJECT sequences

Set length of QUERY sequence

Length to break up the QUERY

The default scoring of matches.

Length

Subject Size: 25
(25 to 50)

Query Size: 10
(10 to 15)

Word Size: 5
(5 to 7)

Scores

Match: 5

Mismatch: -4

Query: AGATTTATAA

RANDOM QUERY SUBMIT QUIZ MODE

Query: A G A T T T A T A A

In Concept Mode, click "Create Words"

CREATE WORDS

BLAST Method: Create “Words” and Search

Select a word to search with

Query: **A G A T T T A T A A**

- A G A T T**
- G A T T T**
- A T T T A**
- T T T A T**
- T T A T A**
- T A T A A**

Makes all possible overlapping words of a given length (5 in this case).

DNA 1: **C A C G C A G A T T T A C G G C T C T G C C A T G**

Query: **A G A T T T A T A A**

- A G A T T**
- G A T T T**
- A T T T A**
- T T T A T**
- T T A T A**
- T A T A A**

Slide along subject sequence until a perfect match is found.

DNA 1: **C A C G C A G A T T T A C G G C T C T G C C A T G**

If no perfect match found for that particular word, select the next word. Repeat until a perfect match is found.

BLAST Method: Expand match and score it

DNA 1: **C A C G C A G A T T T A C G G C T C T G C C A T G**
 Score:

Click EXPAND once match found to expand and score.

EXPAND

Do same for the other SUBJECT sequences.

DNA 1: **T T C T T C A T G G T A A G A T T T A T A A G A G**
 Score: 5 5 5 5 5 5 5 5 5 5

EXPAND

NOTE: In the example, different words make perfect matches to DNA 2 and 3.

DNA 2: **T T C T T C A T A C A T T T A C A C G C A T G A G**
 Score: 5 -4 5 5 5 5 5 -4 5 -4

EXPAND

DNA 3: **T T A C A T T T A T A A T T T G G T G C A T G A G**
 Score: 5 -4 5 5 5 5 5 5 5 5

EXPAND

Result: **Xenopus borealis**

	Score	Rank
DNA 1:	50	1
DNA 2:	23	3
DNA 3:	41	2

The Scores for the alignments to each SUBJECT sequence are summed and ranked.

BLAST Method: Quiz Mode

Follow the same protocol as the Concept Mode, but determine the scores for each match and the relative rank of each match at the end.

Query: **T G T C C A G T C C**

T G T C C

G T C C A

T C C A G

C C A G T

C A G T C

A G T C C

DNA 1: **T T A C T T T T A T T T C C A G G C C T G A T T T**
 Score: 5 -4 5 5 5 5 5 -4 5 5
 EXPAND

DNA 2: **T T A C T T T T A A G A A T A G T C C A G G T C T**
 Score: -4 5 5 5 5 5 5 -4 -4 5
 EXPAND

DNA 3: **T T A C T T A G T C C A G T C C T A C T G A T T T**
 Score: -4 5 5 5 5 5 5 5 5 5
 EXPAND

Result:

	Score	Rank
DNA 1:	32 ✓	2 ✓
DNA 2:	23 ✓	3 ✓
DNA 3:	41 ✓	1 ✓

Of the 3 SUBJECT choices, our QUERY matched the 3rd DNA the best.