

ABSTRACT

A lot of applications exist on the internet which implement global and local sequence alignment techniques and provide the output to the user. Our project intends to build an educational web application which helps a layman understand global and local sequence alignment techniques in an interactive and illustrative manner. The application will force the user to apply his/her mind to understand and perform the computations themselves. Our application will have a 'teaching assistant' to provide hints and annotate the screen during the computation to benefit the user.

**TOPIC
SUMMARY**

Sequence Alignment: In bioinformatics, a sequence alignment is a way of arranging the sequences of DNA, RNA, or protein to identify regions of similarity that may be a consequence of functional, structural, or evolutionary relationships between the sequences. Aligned sequences of nucleotide or amino acid residues are typically represented as rows within a matrix.

Local Alignment: The Smith–Waterman algorithm performs local sequence alignment; that is, for determining similar regions between two strings of nucleic acid sequences or protein sequences. Instead of looking at the entire sequence, the Smith–Waterman algorithm compares segments of all possible lengths and optimizes the similarity measure.

Global Alignment: The Needleman–Wunsch algorithm is an algorithm used in bioinformatics to align protein or nucleotide sequences. It was one of the first applications of dynamic programming to compare biological sequences. The Needleman–Wunsch algorithm is still widely used for optimal global alignment, particularly when the quality of the global alignment is of the utmost importance.

Dove-Tail Alignment: When the sequences are expected to share similarity extending from the beginning of the sequences to their ends they are aligned globally. This means that each residue of either sequence is part either of a residue pair or a gap. It implies that gaps at the ends are charged like any other gap. However, one can prefer to leave gaps at the ends of the sequences un-penalized. This variant is easy to implement in the dynamic programming algorithm. Two modifications are required. Firstly, the initialization of the matrix needs to reflect the gap cost of 0 in the margin of the matrix. Secondly, upon backtracking, one does not necessarily start in the corner of the matrix but much rather searches the margins for the maximum from which to start. Variants of this that penalize only particular end-gaps are easy to derive and can be used, e.g. to fit one sequence into another or to overlap the end of one sequence with the start of another.

**APPLICATION
SUMMARY**

The 'Sequence Alignment Tutor' is an educational web application which is mainly aimed at an audience with no or minimum background in Bioinformatics or Computational Geometry. The app tutors the user of the application to perform sequence alignment of two Nucleotide or Amino Acid sequences.

For ease of access, a navigation menu is provided at the top of the page to help the user navigate to 'Tutorial' and 'Algorithms' sections. If the user has no idea about basic concepts of Bioinformatics, the 'Tutorial' section will help the individual to gain a deeper understanding of the concept before practicing the alignment techniques.

Once the user is confident that he/she understands the algorithms well, he/she can access the 'Algorithms' section to practice and improve their new found understanding. This page requires

the user to enter two sequences which they wish to align, the alignment technique they wish to practice and select a scoring matrix. This application also provides the user with freedom to provide his/her own scoring matrix and input the gap penalty. Once all the required inputs have entered and validated by the page.

Once the user entered details are validated, he/she is directed to alignment page. The alignment page provides the user with a dynamic programming matrix. Initially all the cells of the matrix are red. A user get 2 chances per cell before the cell gets populated with the correct entry. After each wrong entry, the user is prompted with a hint. After entering the correct value, the cell color changes to green. This process continues till the user has filled the entire matrix with the correct values based on his knowledge of the alignment technique.

After filling all the cells of the matrix, the user is prompted to perform backtracking on the filled matrix to get a possible best alignment. There be can multiple best alignments possible depending on the inputs that were chosen, and the user has the freedom to derive one such possible alignment on his/her own. The aligned sequence appears on the screen dynamically as he/she does the backtracking.

The practice round is complete when one of the best alignment has been generated by the user. He/she can then proceed to learn and practice more algorithms if they wish to do so.

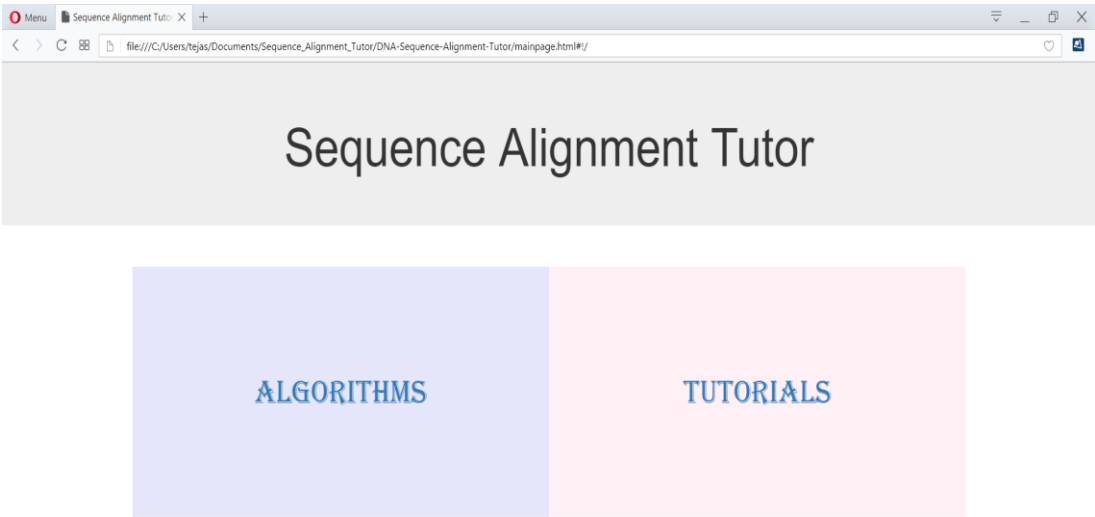
TECHNOLOGY STACK | AngularJS / Bootstrap – UI / Frontend
ExpressJS / NodeJS – Middleware
MongoDB – Backend

INDIVIDUAL CONTRIBUTION | *Karan Acharekar:*

- Developed the frontend of the landing page of the sequence alignment tutor app
- Developed the frontend to build the sequence alignment dynamic programming matrix which provides indication of correct/wrong entries along with position specific hints
- Developed the frontend of the tutorials page and embedded useful sequence alignment videos for user reference
- Wrote server script and set up ExpressJS paths to fetch and retrieve user query data from MongoDB
- Equal contribution to report creation

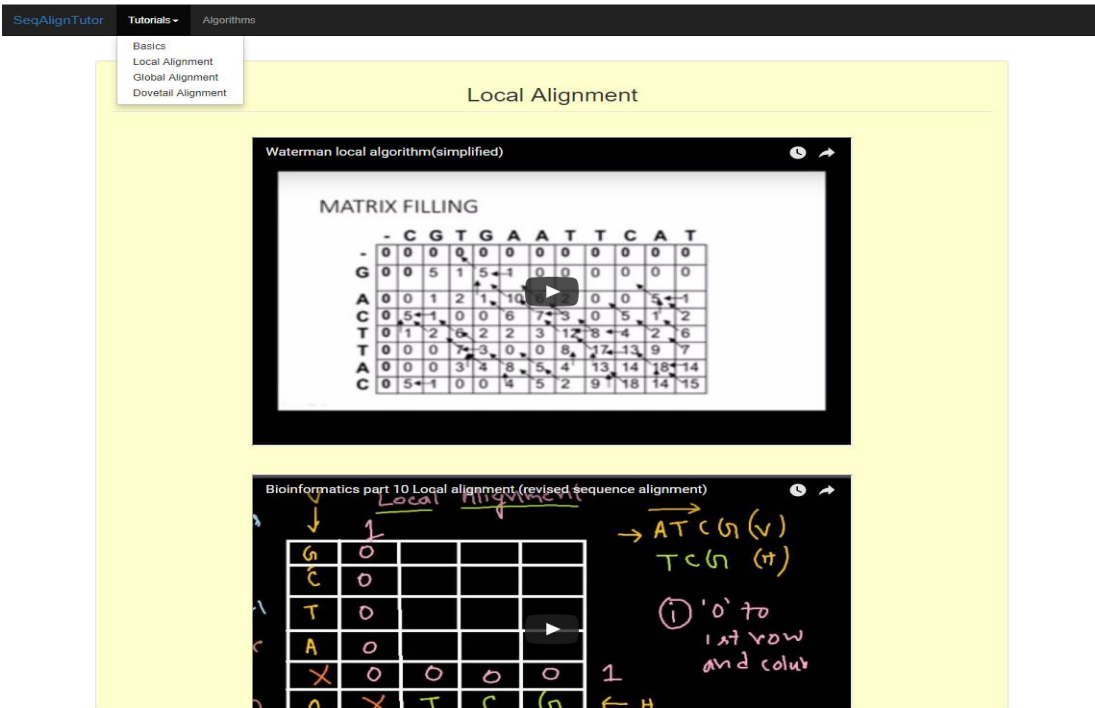
Tejas Sahasranaman:

- Developed the frontend to build the user input form and the supporting backend logic to perform sequence alignments, compute all possible best alignments and the database operations required by the form.
- Set up the MongoDB server to store and retrieve data that is required by multiple web pages of the application
- Added Bootstrap CSS framework to make the web pages viewport responsive and presentable in all possible screen sizes
- Developed the menu bar to make the app easily navigational and consistent
- Equal contribution to report creation



An educational web application which helps a layman understand global and local sequence alignment techniques in an interactive and illustrative manner

Mainpage.html



Tutorial.html

Alignment Parameters

Sequence 1 *

aactcgt

Sequence 2 *

acagata

Alignment Method *

☒ Global☐ Local☐ Dovetail

Scoring Matrix *

☒ Nucleotides☐ Amino Acids

submat2

+ Add New

Gap Penalty *

-1

Initialize

AlgoInit.html

Alignment Configuration

SEQUENCES

Query Sequence 1 :
GATTACA

Query Sequence 2 :
GCATGCT

ALIGNMENT TYPE

Global

SEQUENCE TYPE

Nucleotides

SCORING MATRIX

submat1

GAP PENALTY

-1

[Home](#)

Needleman-Wunsch

Please fill the first row and column first
Please fill the matrix in the same order as shown in the tutorials

	-	G	C	A	T	G	C	T
-		1	2	3	4	5	6	7
G	1							
A	2	1	1					
T	3	1	1					
T	4	2	2					
A	5	3	3					
C	6	4	2					
A	7	5	3					

Please click on the correct above boxes to indicate backtracking

Aligned Sequences

G - A T T A C A

| | | | | | |

G C A - T G C T

submat1

	A	C	G	T
A	1	-1	-1	-1
C	-1	1	-1	-1
G	-1	-1	1	-1
T	-1	-1	-1	1

Global Alignment Tutor Page

Alignment Configuration

SEQUENCES

Query Sequence 1 :
GATTACA

Query Sequence 2 :
TACTACAG

ALIGNMENT TYPE

Global

SEQUENCE TYPE

Nucleotides

SCORING MATRIX

Score

GAP PENALTY

2

Smith-Waterman

Please fill the first row and column first
Please fill the matrix in the same order as shown in the tutorials

-	T	G	T	T	A	C	G	G
-	0	0	0	0	0	0	0	0
G	0	3	1	0	0	0	3	3
G	0	0	0	1	0	0	0	0
T	0	3	1	0	4	2	0	1
T	0	3	1	1	0	7	5	3
G	0	1	0	0	0	6	4	0
A	0	0	4	3	5	0	0	5
C	0	0	2	1	3	0	11	0
T	0	3	1	5	4	0	10	5
A	0	1	0	0	2	0	0	7

Please click on the correct above boxes to indicate backtracking

Aligned Sequences

G T T G A C

| | | d | |

G T T - A C

xxx

A	C	G	T
A	-3	-3	-3
C	-3	-3	-3
G	-3	-3	-3
T	-3	-3	-3

Local Alignment Tutor Page

Alignment Configuration

SEQUENCES

Query Sequence 1 :
GCGTAC

Query Sequence 2 :
TTCGATG

ALIGNMENT TYPE

Local

SEQUENCE TYPE

Nucleotides

SCORING MATRIX

submat1

GAP PENALTY

-1

Dove-Tail

Please fill the first row and column first
Please fill the matrix in the same order as shown in the tutorials

-	T	T	C	C	A	G	T	G
-	0	0	0	0	0	0	0	0
C	0	0	0	0	1	0	1	0
C	0	0	0	1	1	1	0	0
A	0	0	0	0	1	0	0	0
T	0	1	1	0	0	2	2	2
G	0	0	0	0	0	1	1	2
C	0	0	0	1	1	0	1	2

Please click on the correct above boxes to indicate backtracking

Aligned Sequences

C C A - T G

| | | | | |

C C A G T G

submat1

A	C	G	T
A	1	-1	-1
C	-1	1	-1
G	-1	-1	1
T	-1	-1	-1

Dovetail Alignment Tutor Page

REFERENCES |

Research Papers and Articles:

- Needleman, S. B. and Wunsch, C. D. (1970). A general method applicable to the search for similarities in the amino acid sequence of two proteins. J. Mol. Biol. 48, 443-453.
- Murata, M. (2004). Three-way Needleman—Wunsch algorithm. Methods in Enzymology. 183, 365-375.
- Smith, T. F. and Waterman, M. (1981). Identification of common molecular subsequences. J. Mol. Biol. 147, 195-197.
- Altschul, S. F., Gish, W., Miller, W., Myers, E. W. and Lipman, D. J. (1990). A Basic Local Alignment Search Tool. J. Mol. Biol., 215, 403-410.
- https://en.wikipedia.org/wiki/Smith–Waterman_algorithm
- https://en.wikipedia.org/wiki/Needleman-Wunsch_algorithm

Related Applications:

- https://blast.ncbi.nlm.nih.gov/Blast.cgi?PAGE_TYPE=BlastSearch&PROG_DEF=blastn&BLAST_PROG_DEF=blastn&BLAST_SPEC=GlobalAln&LINK_LOC=BlastHomeLink
- <http://insilico.ehu.es/align/>
- https://embnet.vital-it.ch/software/LALIGN_form.html
- https://embnet.vital-it.ch/software/PRSS_form.html