|  |  |
| --- | --- |
| Project title | **Sequence Alignment Tutor Application** |
| Team Members | *Karan Acharekar*  *Tejas Sahasranaman* |
| 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](https://en.wikipedia.org/wiki/Bioinformatics), a sequence alignment is a way of arranging the sequences of [DNA](https://en.wikipedia.org/wiki/DNA), [RNA](https://en.wikipedia.org/wiki/RNA), or [protein](https://en.wikipedia.org/wiki/Protein) to identify regions of similarity that may be a consequence of functional, [structural](https://en.wikipedia.org/wiki/Structural_biology), or [evolutionary](https://en.wikipedia.org/wiki/Evolution) relationships between the sequences. Aligned sequences of [nucleotide](https://en.wikipedia.org/wiki/Nucleotide) or [amino acid](https://en.wikipedia.org/wiki/Amino_acid) residues are typically represented as rows within a [matrix](https://en.wikipedia.org/wiki/Matrix_(mathematics)).  **Local Alignment:** The Smith–Waterman algorithm performs local [sequence alignment](https://en.wikipedia.org/wiki/Sequence_alignment); that is, for determining similar regions between two strings of [nucleic acid sequences](https://en.wikipedia.org/wiki/Nucleic_acid_sequence) or [protein sequences](https://en.wikipedia.org/wiki/Protein_sequence). Instead of looking at the [entire](https://en.wikipedia.org/wiki/Needleman%E2%80%93Wunsch_algorithm) sequence, the Smith–Waterman algorithm compares segments of all possible lengths and [optimizes](https://en.wikipedia.org/wiki/Mathematical_optimization) the [similarity measure](https://en.wikipedia.org/wiki/Similarity_measure).  **Global Alignment:** The Needleman–Wunsch algorithm is an [algorithm](https://en.wikipedia.org/wiki/Algorithm) used in [bioinformatics](https://en.wikipedia.org/wiki/Bioinformatics) to [align](https://en.wikipedia.org/wiki/Sequence_alignment) [protein](https://en.wikipedia.org/wiki/Protein) or [nucleotide](https://en.wikipedia.org/wiki/Nucleotide) sequences. It was one of the first applications of [dynamic programming](https://en.wikipedia.org/wiki/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 | **If a person has no idea about sequence alignment and all the alignment techniques, this application provides a tutorial section. This section will help the individual to gain a deeper understanding of the concept before moving forward with practicing the alignment techniques.**  **Once a user is confident that he/she understands the algorithm well, they can move forward to the input page. This page requires the user to enter 2 sequences which they wishe to align and the alignment technique they wish to use. They also need to select a scoring matrix. This application also provides the user with freedom to provide his own scoring matrix. He also needs to mention the gap penalty. Only after providing the above-mentioned details can he move forward with the alignment.**  **The alignment stage provides the user with a dynamic 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 a hint. After entering the correct value, the cell color changes to green. After filling all the cells of the matrix, the user is prompted to indicate the backtracking. After each correct selection aligned sequences appears on the screen.** |
| TECHNOLOGY STACK | MongoDB, ExpressJs, AngularJs, NodeJs |
| Individual CONTRIBUTION | *Karan Acharekar:*  *Tejas Sahasranaman:*  tgttagg |
| SNapshots |  |
| 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](http://www.sciencedirect.com/science/bookseries/00766879). 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> |
|  |  |