Local Alignment tool

Prepared by
-mahmoud Yasser fathy
-mohamed abdo elsayed
-mohamed gamal abd elsamae
-ahmed kamal mohamed

prepared for
-Dr. Sara El-Metwally

Genomics Project

ABOUT THE PROJECT

DATE PROPOSED

19TH APRIL 2021

Initially , Local alignments are more useful for dissimilar sequences that are suspected to contain regions of similarity or similar sequence motifs within their larger sequence context. The Smith–Waterman algorithm is a general local alignment method based on the same dynamic programming scheme but with additional choices to start and end at any place.

Local alignment :-

- Very similar to global alignment!
- Biological usefulness: If we have two dissimilar sequences and want to see if there is a conserved gene or region between the two

Why Smith–Waterman algorithm ?...

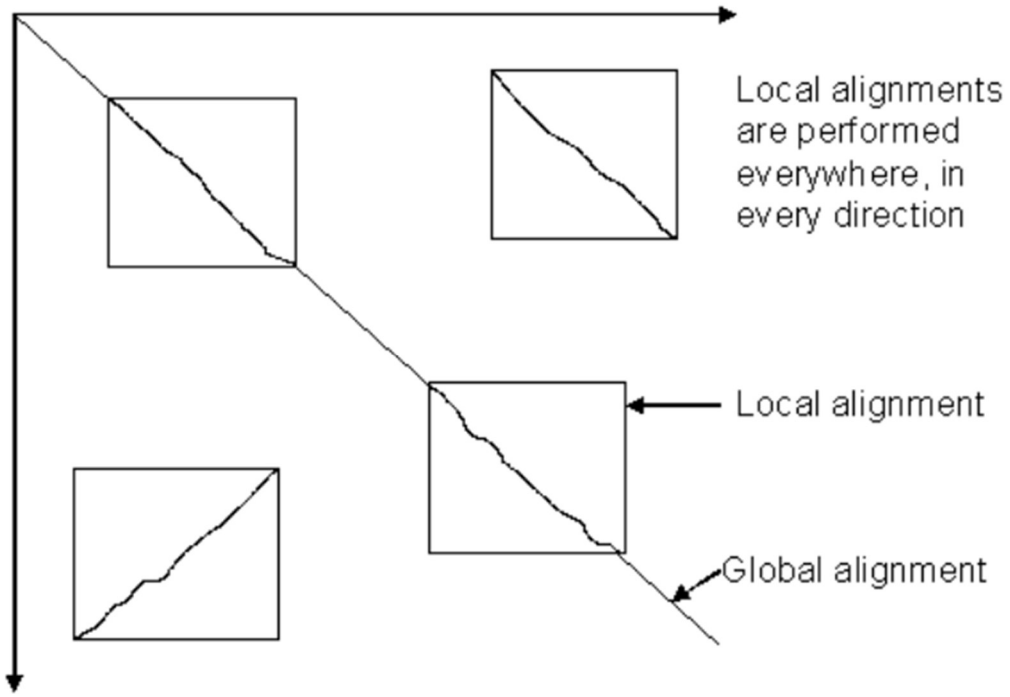
Sequence similarity searches performed using the Smith-Waterman algorithm guarantees you the optimal local alignments between query and database sequences. Thus, you are ensured the best performance on accuracy and the most precise results - aspects of significant importance when you cannot afford to miss any information gained from the similarity search as e.g. when searching for remote homology. The Smith-Waterman algorithm being the most sensitive algorithm for detection of sequence similarity has however some costs.

Note : - we use human genome as reference to align query sequence....

Smith Waterman Algorithm Pseudocode :-

1 $S[0,0] = 0$
2 for $i = 1$ to M do:
3 $\quad S[i,0] = 0$
4 for $j = 1$ to N do:
5 $\quad S[0,j] = 0$
6 $\quad$ for $i = 1$ to M do:
7 $$S[i,j] = MAX \begin{cases} 0 \\ S[i-1,j-1] + \delta(x_i, y_j) \\ S[i-1,j] + \delta(x_i, -) \\ S[i,j-1] + \delta(-, y_j) \end{cases}$$
8 return $S[M,N]$

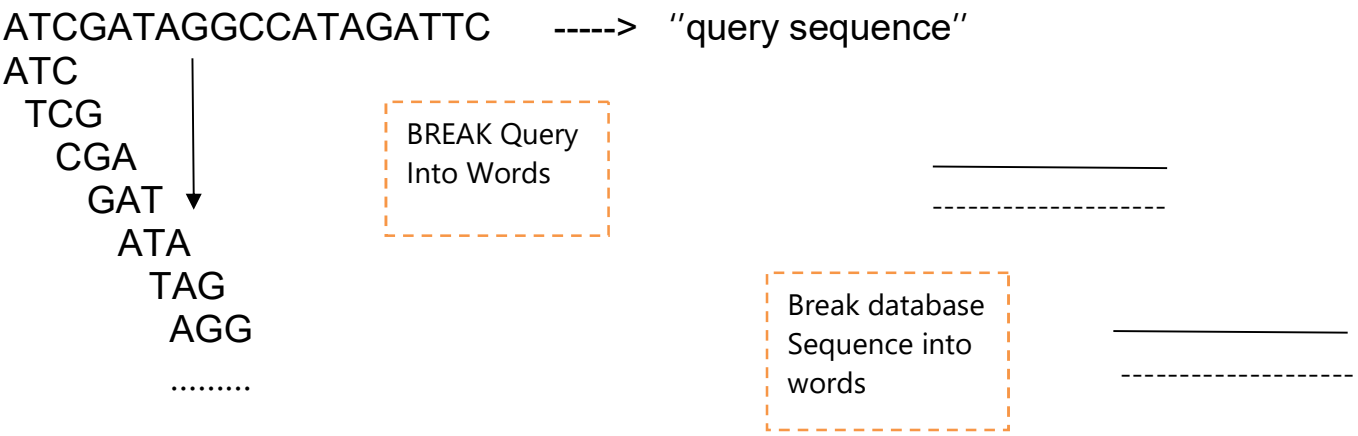
Local alignment *vs* global



Global align all query sequence with ref ,however local align substring from query sequence.

In human genome. Break whole genome sequence into **11** bases length words , then record the location of each word .

Figure1.



- Compare all locations between each 11 bases length words of query sequences. For each words, it will have many location in each chromosome, but only one of them is the right location of query sequence

Project Goals : -

Looking for query sequence " cut to words" in reference human genome" cut to words" figure.1 after this we align query Sequence with ref by "smith-waterman"