Pairwise and Multiple Sequence Alignment

local alignment algorithm
--Smith and Waterman (1981)

❖ **Optimal local alignment of two sequences**

❖ **Performs an exhaustive search for the optimal local alignment**

- Complexity $O(n{\times}m)$ for sequence lengths n and m

❖ **Based on the 'dynamic programming' (DP) algorithm**

- Fill the DP matrix

- Find the maximal value (score) in the matrix

- Trace back from the score until a 0 value is reached

Smith-Waterman Algorithm

$$H(i,j) = \max \begin{cases} 0 \\ H(i-1,j) - W_i \\ H(i,j-1) - W_i \\ H(i-1,j-1) + S(S1_i, S2_j) \end{cases}$$

$$H(i,0) = 0, \ 0 \le i \le m$$

$$m = length(S1)$$

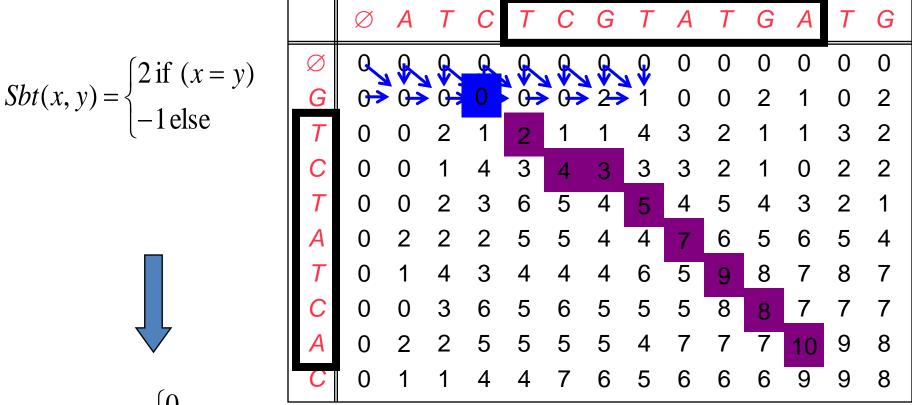$$H(0,j) = 0, \ 0 \le j \le n$$

$$n = length(S2)$$

$S(S1_i, S2_j)$ is a similarity function on the alphabet

$H(i,j)$ is the maximum Similarity-Score

W_i is the [gap-scoring](#) scheme

Smith-Waterman Algorithm (Example)

- Align $S1=ATCTCGTATGATG$ $S2=GTCTATCAC$

$$Sbt(x, y) = \begin{cases} 2 \text{ if } (x = y) \\ -1 \text{ else} \end{cases}$$

	∅	A	T	C	T	C	G	T	A	T	G	A	T	G
∅	0	0	0	0	0	0	0	0	0	0	0	0	0	0
G	0	0	0	0	0	0	2	1	0	0	2	1	0	2
T	0	0	2	1	2	1	1	4	3	2	1	1	3	2
C	0	0	1	4	3	4	3	3	3	2	1	0	2	2
T	0	0	2	3	6	5	4	5	4	5	4	3	2	1
A	0	2	2	2	5	5	4	4	7	6	5	6	5	4
T	0	1	4	3	4	4	4	6	5	9	8	7	8	7
C	0	0	3	6	5	6	5	5	5	8	8	7	7	7
A	0	2	2	5	5	5	5	4	7	7	7	10	9	8
C	0	1	1	4	4	7	6	5	6	6	6	9	9	8

$$H(i, j) = \max \begin{cases} 0 \\ H(i-1, j) - 1 \\ H(i, j-1) - 1 \\ H(i-1, j-1) + Sbt(S1_i, S2_j) \end{cases}$$

$A T C T C G T A T G A T G$

$G T C - T A T C A C$

Smith-Waterman Algorithm (Example)

- Sequence 1 = ACACACTA

- Sequence 2 = AGCACACA

$s(a, b) = +2$ if $a = b$ (match), -1 if $a \neq b$ (mismatch)

$W_i = -1$

Smith-Waterman Algorithm (Example)

- Sequence 1 = ACACACTA
- Sequence 2 = AGCACACA

$$s(a, b) = +2 \text{ if } a = b \text{ (match)}, \ -1 \text{ if } a \neq b \text{ (mismatch)}$$

$W_i = -1$

$$H = \begin{pmatrix}
 & - & A & C & A & C & A & C & T & A \\
- & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
A & 0 & 2 & 1 & 2 & 1 & 2 & 1 & 1 & 2 \\
G & 0 & 1 & 1 & 1 & 1 & 1 & 1 & 0 & 1 \\
C & 0 & 1 & 3 & 2 & 3 & 2 & 3 & 2 & 2 \\
A & 0 & 2 & 2 & 5 & 4 & 5 & 4 & 4 & 4 \\
C & 0 & 1 & 4 & 4 & 7 & 6 & 7 & 6 & 6 \\
A & 0 & 2 & 3 & 6 & 6 & 9 & 8 & 8 & 8 \\
C & 0 & 1 & 4 & 5 & 8 & 8 & 11 & 10 & 10 \\
A & 0 & 2 & 3 & 6 & 7 & 10 & 10 & 10 & 12
\end{pmatrix}^T =
\begin{pmatrix}
 & - & A & C & A & C & A & C & T & A \\
- & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
A & 0 & \nwarrow & \leftarrow & \nwarrow & \leftarrow & \nwarrow & \leftarrow & \leftarrow & \nwarrow \\
G & 0 & \uparrow & \nwarrow & \uparrow & \nwarrow & \uparrow & \nwarrow & \nwarrow & \uparrow \\
C & 0 & \uparrow & \nwarrow & \leftarrow & \nwarrow & \leftarrow & \nwarrow & \leftarrow & \leftarrow \\
A & 0 & \nwarrow & \uparrow & \nwarrow & \leftarrow & \nwarrow & \leftarrow & \leftarrow & \nwarrow \\
C & 0 & \uparrow & \nwarrow & \uparrow & \nwarrow & \leftarrow & \nwarrow & \leftarrow & \leftarrow \\
A & 0 & \nwarrow & \uparrow & \nwarrow & \uparrow & \nwarrow & \leftarrow & \leftarrow & \nwarrow \\
C & 0 & \uparrow & \nwarrow & \uparrow & \nwarrow & \uparrow & \nwarrow & \leftarrow & \leftarrow \\
A & 0 & \nwarrow & \uparrow & \nwarrow & \uparrow & \nwarrow & \uparrow & \nwarrow & \nwarrow
\end{pmatrix}$$

The results are:

Sequence 1 = A-CACACTA
Sequence 2 = AGCACAC-A