

Algorithms for Bioinformatics

Francesco Penasa

March 2, 2020

2020 - 03 - 02

Needleman-Wunsch

$$s_{ij} \begin{cases} \emptyset & \text{if } i = 1, j = 1 \\ s_{i,j-1} + g & \text{if } i = 1, j > 1 \\ s_{i-1,j} + g & \text{if } i > 1, j = 1 \\ \max(s_{i-1,j-1} + M_{i,j}, s_{i-1,j} + g, s_{i,j-1} + g) & \end{cases}$$

$g = \text{gap_value}$

$m = \text{match(or mismatch_value)}$

Example example1

$\text{seq1} = a \ c \ g \ t \ c$

$\text{seq2} = t \ g \ a \ c$

$\text{score} = -2$

acgtc

t_gac

If we prepare a gap at the beginning (or end)

We should put at $M[0][j] = 0$ $M[i][0] = 0$, so we don't pay gap at the extremes of the sequences.

1 Local alignment - Smith-Waterman algorithm

It's a problem that answers to a different need. With local alignment i don't care at all about the gaps at the beginning or at the end. So the matrix is initialized all at 0 values.

https://it.wikipedia.org/wiki/Algoritmo_di_Smith-Waterman

The only difference is that we build a

$$H = matrix$$

$$H_{i,0} = 0 \quad 0 \leq i \leq n$$

$$H_{0,j} = 0 \quad 0 \leq j \leq m$$

$$H(i,j) = \max \begin{cases} 0 \\ H_{i-1,j-1} + w(a_i, b_j) & Match/Mismatch \\ H_{i-1,j} + w(a_i, -) & Insertion \\ H_{i,j-1} + w(-, b_j) & Deletion \end{cases}$$

$$w = gap_weight$$

In the original paper

$$w = 1 + \frac{1}{3} - k$$

example

- *match* = 1
- *mismatch* = -1

-	-	a	c	g	t	c
-	0	0	0	0	0	0
t	0	0	0	0	↖ 1	↖ 0
g	0	0	0	↖ 1	↖ 0	0
a	0	↖ 1	0	0	↖ 0	0
c	0	0	↖ 2	← 0.66	← 0.33	↖ 1

find H.index(max(H)) and rebuild the sequence through the arrows.

```
score = 2
- _acgtc
tgac
```

```
score = 1
acgtc
- _tgac
```

```
score = 1
acgtc
- - _tgac
```