In bioinformatics, there is a sequence comparison problem, defined as follows. Assume to score 2 for a match, −3 for a mismatch, and −1 for an insertion or deletion; then derive the matching with the highest score. What if score −1.5 for a mismatch? Briefly describe the algorithm and illustrate your algorithm by the following two sequences, ACGCTGA and AACTGT.

Ans.

2022.4.5

假設 X = <ACGCTGA> , Y = <AACTGT>　花 39:27

令一個矩陣 M，其中 M[i,j] 存放，X 和 Y comparison 的最高分數。

推測其遞迴的表示法為

$$M[i,j] = \max \begin{cases} M[i-1,j-1]+2, & \text{if } x_i == y_j \\ M[i-1,j-1]-3, & \text{if } x_i \neq y_j \\ M[i-1,j]-1, & \text{if delete } x_i \\ M[i,j-1]-1, & \text{if insert } y_j \end{cases}$$

如果最高分是
- match，標記 $\nwarrow^m$
- dismatch，標記 $\nwarrow^x$
- delete，標記 ↑
- insert，標記 ←

由 Dynamic Programming 得最高分

	∅	A	A	C	T	G	T
∅	0	←-1	←-2	←-3	←-4	←-5	←-6
A	↑-1	$\nwarrow^m$ 2	$\nwarrow^m$ 1	←0	←-1	←-2	←-3
C	↑-2	(↑1)	←0	$\nwarrow^m$ 3	←2	←1	←0
G	↑-3	(↑0)	(←-1)	↑2	←1	$\nwarrow^m$ 4	←3
C	↑-4	↑-1	←-2	$\nwarrow^m$ 1	←0	↑3	←2
T	↑-5	↑-2	←-3	↑0	$\nwarrow^m$ 3	←2	$\nwarrow^m$ 5
G	↑-6	↑-3	←-4	↑-1	↑2	$\nwarrow^m$ 5	←4
A	↑-7	↑-4	$\nwarrow^m$ -1	←-2	↑1	(↑4)	(←3)

得最高分：3

old: A ∅ G - C T G A -
new: A - - A C T G - T

如果 mismatch 權重改成 -1.5

$$M[i,j] = \begin{cases} M[i-1,j-1] + 2 & , \text{if } x_i == y_j \\ \max \begin{cases} M[i-1,j-1] - 1.5 & , \text{if } x_i \neq y_j \\ M[i-1,j] - 1 & , \text{delete } x_i \\ M[i,j-1] - 1 & , \text{insert } y_j \end{cases} \end{cases}$$

	∅	A	A	C	T	G	T
∅	0	←-1	←-2	←-3	←-4	←-5	←-6
A	↑-1	↖m 2	↖m 2	←1	←0	←-1	←-2
C	↑-2	↑1	↑1	↖m 4	←3	←2	←1
G	↑-3	↑0	↑0	↑3	↖x 2.5	↖m 5	←4
C	↑-4	↑-1	↑-1	↖m 2	↖x 1.5	↑4	↖x 3.5
T	↑-5	↑-2	↑-2	↑1	↖m 4	↑3	↖m 6
G	↑-6	↑-3	↑-3	↑0	↑3	↖m 6	↑5
A	↑-7	↖m 4	↖m -1	↑-1	↑2	↑5	↑4

最高分數為 4

old: - A C - G ~~C~~ T ~~G~~ ~~A~~

new: A A C T G - T - -