Your Name: Student ID:

When we do sequence alignment, what is the assumption about the aligned sites? What different processes can led to sequence divergence? Since we can not observe the real history of sequence evolution, what strategy can we use to align the sequences, and how to justify the strategy?

1.2

(a) Do a LCS (longest common sequence) alignment of the following sequences using the table provided below. Fill the table and write down the aligned sequences.

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
|  |  | C | C | G | A | T | G |
|  |  |  |  |  |  |  |  |
| A |  |  |  |  |  |  |  |
| C |  |  |  |  |  |  |  |
| G |  |  |  |  |  |  |  |
| A |  |  |  |  |  |  |  |
| G |  |  |  |  |  |  |  |

(b) Now if we change the scoring matrix to mismatch=-2, gap=-3, match=1, do the alignment again using Smith-Waterman algorithm.

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
|  |  | C | C | G | A | T | G |
|  |  |  |  |  |  |  |  |
| A |  |  |  |  |  |  |  |
| C |  |  |  |  |  |  |  |
| G |  |  |  |  |  |  |  |
| A |  |  |  |  |  |  |  |
| G |  |  |  |  |  |  |  |

(c) Now use the Needleman-Wunsch algorithm to repeat the sequence alignment.

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
|  |  | C | C | G | A | T | G |
|  |  |  |  |  |  |  |  |
| A |  |  |  |  |  |  |  |
| C |  |  |  |  |  |  |  |
| G |  |  |  |  |  |  |  |
| A |  |  |  |  |  |  |  |
| G |  |  |  |  |  |  |  |

1.3

Explain what **are** the trick**s** designed in the BLAST algorithm that can save much computational time than aligning the query sequence to each of the subject sequence of the database?