

Gene Alignment Time Analysis and Approach Explanation

Problem Overview:

The code implements the Smith-Waterman algorithm for local sequence alignment, particularly for gene sequences. The goal is to find the highest-scoring alignment and calculate its score using a scoring matrix.

Time Complexity:

The time complexity of the Smith-Waterman algorithm is $O(n * m)$, where n and m are the lengths of the input sequences x and y , respectively. This is because the algorithm involves filling in a 2D dynamic programming matrix of size $(n + 1) * (m + 1)$.

Approach:

1. Dynamic Programming Matrix Initialization:

- The dynamic programming matrix (`dp`) is initialized based on the scoring matrix and the input sequences `x` and `y`.
- Initialization of the first row and first column considers gap penalties.

2. Dynamic Programming Matrix Filling:

- The matrix is filled iteratively, considering match, delete, and insert scores.
- The highest scoring alignment is determined during this process.

3. Backtracking to Retrieve Alignment:

- Starting from the cell with the highest score, the algorithm backtracks to reconstruct the aligned sequences (`alignmentX` and `alignmentY`).

4. Scoring the Alignment:

- The alignment is scored by summing the corresponding values from the scoring matrix.

5. BigDecimal for Precision:

- To mitigate precision issues associated with floating-point arithmetic, `BigDecimal` is used to perform precise addition when calculating the alignment score.

Code Execution:

The provided code successfully aligns the given gene sequences `x` and `y`. It outputs the aligned sequences and their score.

Conclusion:

The Smith-Waterman algorithm efficiently identifies local alignments in gene sequences. The use of dynamic programming ensures optimal substructure, and the scoring matrix allows customization of match, mismatch, and gap penalties. The `BigDecimal` class is employed to enhance precision in calculating alignment scores. The code demonstrates a robust approach for gene sequence alignment with accurate scoring.