

Sequence Alignment Algorithm Visualization

User Guide

By Hannah Coleman

Table of Contents

| | |
|----------------------------|----------|
| Algorithms | 2 |
| Longest Common Subsequence | 2 |
| Global Alignment | 3 |
| Local Alignment | 3 |
| View | 3 |
| Current Cell | 3 |
| Traced Path | 3 |
| Controls | 3 |
| Input Text Fields | 3 |
| Navigation Buttons | 3 |

Algorithms

The program uses a grid to work through the algorithm. Sequence 1 is set vertically along the left edge of the grid; sequence 2 is set horizontally across the top edge. There is a starting column and a starting row in the grid, adjacent to the input sequences. There are two phases the algorithm. In the first phase, the grid is filled, cell by cell, with scores and path directions. In the second phase, a path is traced back to the upper left corner to find a solution.

Longest Common Subsequence

For this algorithm, the scores in the starting column and row are all 0. Scores are filled in left to right and top to bottom. The score for the remaining cells is determined as follows:

If the residues in the current row and column match, the score in the current cell is set as the highest of

- the horizontal score (from the left)
- the vertical score (from above)
- the the diagonal score (from the upper left) + 1

In the event of a tie, the diagonal score has the highest priority. In this program, if the horizontal and vertical scores are tied, the vertical score has priority.

If the residues in the current row and column do not match, the score in the current cell is set as the highest of

- the horizontal score
- the vertical score

Again, in this program, if the horizontal and vertical scores are tied, the vertical score has priority.

An arrow is placed in the cell showing the direction the score came from.

Once all the cells are filled, the algorithm starts in the lower right corner of the grid and follows the arrows back to the upper left corner. A diagonal arrow indicates a match. A horizontal arrow indicates an insertion (a residue in sequence 2 not in sequence 1). A vertical arrow indicates a deletion (a residue in sequence 1 not in sequence 2).

| | | C | A | T | C | G | T |
|---|---|---|---|---|---|---|---|
| | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| A | 0 | | | | | | |
| T | 0 | | | | | | |
| G | 0 | | | | | | |
| T | 0 | | | | | | |
| T | 0 | | | | | | |

| | | C | A | T | C | G | T |
|---|---|---|---|---|---|---|---|
| | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| A | 0 | 0 | 1 | 1 | 1 | | |
| T | 0 | | | | | | |
| G | 0 | | | | | | |
| T | 0 | | | | | | |
| T | 0 | | | | | | |

| | | C | A | T | C | G | T |
|---|---|---|---|---|---|---|---|
| | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| A | 0 | 0 | 1 | 1 | 1 | 1 | 1 |
| T | 0 | 0 | 1 | 2 | 2 | 2 | 2 |
| G | 0 | 0 | 1 | 2 | 2 | 3 | 3 |
| T | 0 | 0 | 1 | 2 | 2 | 3 | 4 |
| T | 0 | 0 | 1 | 2 | 2 | 3 | 4 |

Global Alignment

(To be filled out when this algorithm has been implemented)

Local Alignment

(To be filled out when this algorithm has been implemented)

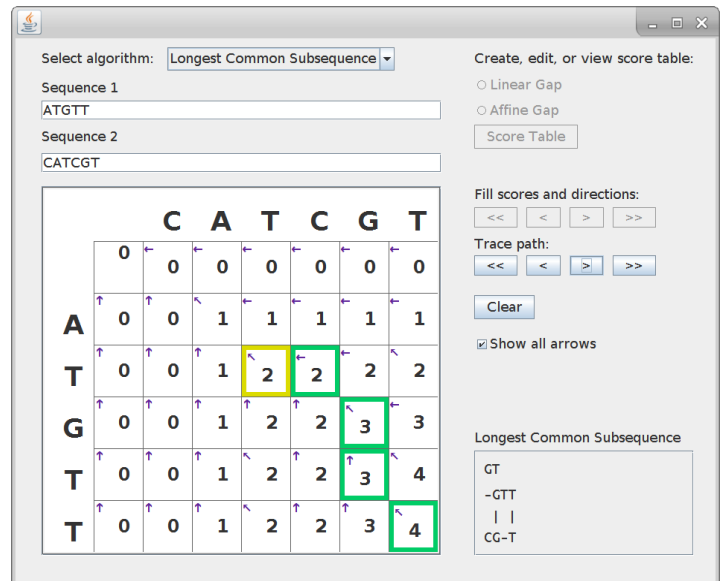
View

Current Cell

The current cell is marked with a dark yellow border and represents the cell that is about to be acted on—either filled or traced.

Traced Path

The traced path, used to find the solution, is marked with green borders. All cells marked with green borders are accounted for in the solution shown in the lower right section of the main window.



Controls

Input Text Fields

The text fields for sequence 1 and sequence 2 receive input. To update the grid with the input, click outside the text fields.

Navigation Buttons

There are nine buttons for navigating the grid. The first four are for the filling phase, the next four are for the tracing phase, and the last one clears (resets) the grid.

- The << buttons jump to the beginning of the phase.
- The < buttons go backward one step in the phase.
- The > buttons go forward one step in the phase.
- The >> buttons jump to the end of the phase.
- The **Clear** button resets the grid.

