Comparative Analysis of Local Sequence Alignment Algorithms: BLAST, FASTA, and Smith-Waterman

Sushanth Reddy Kotha, Durga Sai Surya Ram Saladi, Mayur Sai Yaram

November 9, 2024

Abstract

This report provides a comparative analysis of three sequence alignment algorithms—BLAST, FASTA, and Smith-Waterman—based on criteria such as speed, sensitivity, accuracy, and use cases. The goal is to understand each algorithm's suitability for specific types of local sequence alignments in bioinformatics.

1 Introduction

Sequence alignment plays a crucial role in bioinformatics, helping identify similar regions between DNA or protein sequences. This report compares BLAST, FASTA, and Smith-Waterman to evaluate their performance on various criteria relevant to local sequence alignment.

2 Methods

2.1 Dataset

Describe the dataset used, including the types and sources of DNA or protein sequences.

2.2 Tools

Explain each alignment tool:

- **BLAST**: Overview and basic functionality.

- **FASTA**: Description and working principles.

- **Smith-Waterman**: Details of the exact pairwise alignment algorithm.

2.3 Alignment Criteria

List the criteria used for comparison:

- Algorithm type

- Speed and efficiency

- Alignment sensitivity

- Accuracy

- Use cases

3 Results

Present your findings for each tool. Use tables or figures as needed.

4 Results

Present your findings for each tool. Use tables or figures as needed.

4.1 Score

Table 1 shows the score values for each sequence across the three alignment tools: BLAST, FASTA, and Smith-Waterman.

Sequence	BLAST Score	FASTA Score	Smith-Waterman Score
BAAFST010000044.1	1729	222.1	90
Sequence 2	78	75	88
Sequence 3	92	89	95
Sequence 4	80	78	85
Sequence 5	87	84	91

Table 1: Score values for each alignment tool.

4.2 Percentage Identity

Table 2 shows the percentage identity for each sequence across BLAST, FASTA, and Smith-Waterman.

Sequence	BLAST Identity (%)	FASTA Identity (%)	Smith-Waterman Iden
BAAFST010000044.1	78.28%	41.5%	97%
Sequence 2	89%	88%	90%
Sequence 3	92%	90%	95%
Sequence 4	85%	83%	89%
Sequence 5	91%	90%	93%

Table 2: Percentage identity for each alignment tool.

4.3 E-value

Table 3 presents the E-value for each sequence alignment across the three tools.

Sequence	BLAST E-value	FASTA E-value	Smith-Waterman E-value
BAAFST010000044.1	0.0	6.1e-55	0.0001
Sequence 2	0.005	0.008	0.002
Sequence 3	0.0005	0.002	0.0002
Sequence 4	0.007	0.010	0.005
Sequence 5	0.002	0.006	0.001

Table 3: E-value for each alignment tool.

4.4   Length of Alignment

Table 4 shows the length of alignment for each sequence across the three tools.

Sequence	BLAST Length	FASTA Length	Smith-Waterman Length
BAAFST010000044.1	4963	526	152
Sequence 2	140	138	143
Sequence 3	160	157	162
Sequence 4	130	128	135
Sequence 5	155	153	158

Table 4: Length of alignment for each alignment tool.

4.5   Execution Time

Table 5 presents the execution time for each sequence alignment across the three tools.

Sequence	BLAST Time (s)	FASTA Time (s)	Smith-Waterman Time (s)
BAAFST010000044.1	0.5	0.7	1.2
Sequence 2	0.4	0.6	1.1
Sequence 3	0.6	0.8	1.3
Sequence 4	0.5	0.7	1.0
Sequence 5	0.7	0.9	1.4

Table 5: Execution time for each alignment tool.

4.6   BLAST Analysis

Summarize results obtained from BLAST, including scores, E-values, identities, and positives.

4.7   FASTA Analysis

Summarize the findings from FASTA and compare them with BLAST results.

4.8   Smith-Waterman Analysis

Describe the results from Smith-Waterman pairwise alignments, focusing on sensitivity and accuracy.

5 Discussion

Compare the three methods based on the specified criteria. Discuss the strengths and limitations of each, highlighting when each tool is most appropriate.

6 Conclusion

Summarize the findings and recommend the appropriate use cases for each algorithm.

7 References

References

[1] S. F. Altschul et al., "Basic local alignment search tool," Journal of Molecular Biology, vol. 215, no. 3, pp. 403–410, 1990.

[2] W. R. Pearson and D. J. Lipman, "Improved tools for biological sequence comparison," Proceedings of the National Academy of Sciences, vol. 85, no. 8, pp. 2444–2448, 1988.

[3] T. F. Smith and M. S. Waterman, "Identification of common molecular subsequences," Journal of Molecular Biology, vol. 147, no. 1, pp. 195–197, 1981.