Analysis of Objective Functions for Ribonucleic Acid Multiple Sequence Alignment Fusion Based on Harmony Search Algorithm

Mubarak Saif [1,*], Rosni Abdullah[2], Mohd. Adib Hj. Omar[2], Abdulghani Ali Ahmed[3], Nurul Aswa Omar[1] , Salama A. Mostafa[1]

[1]Faculty of Computer Science and Information Technology, Universiti Tun Hussein Onn Malaysia, 86400, Malaysia
[2]School of Computer Sciences, Universiti Sains Malaysia, Penang, 11800, Malaysia
[3]School of Computer Science and Informatics, De Montfort University, The Gateway, Leicester LE1 9BH, UK
Emails: mubaraksm@uthm.edu.my; rosni@usm.my; adib@usm.my; aa.ahmed@dmu.ac.uk; nurulaswa@uthm.edu.my; salama@uthm.edu.my

Abstract

Four kinds of smaller molecules known as ribonucleotide bases-adenine (A), cytosine (C), guanine (G), and uracil (U) combine to form the linear molecule known as ribonucleic acid (RNA). Aligning multiple sequences is a fundamental task in bioinformatics. This paper studies the correlation of different objective functions applying to RNA multiple sequence alignment (MSA) fusion generated by the Harmony search-based method. Experiments are performed on the BRAliBase dataset containing different numbers of test groups. The correlation of the alignment score and the quality obtained is compared against coffee, sum-of-pairs (SP), weight sum-of-pairs (WSP), NorMD, and MstatX. The results indicate that COFFEE and SP objective functions achieved a correlation coefficient (R^2) of 0.96 and 0.92, respectively, when compared to the reference alignments, demonstrating their effectiveness in producing high-quality alignments. In addition, the sum-of-pairs takes less time than the COFFEE objective function for the same number of iterations on the same RNA benchmark.

Keywords: Objective function; Harmony search; Ribonucleic acid (RNA); Multiple sequence alignment (MSA)

1. Introduction

For most alignment techniques, protein research has been exhaustive. RNA alignment algorithms have thus far been applied to nucleic acid sequences in a limited number of instances. There has been a recent surge in the acquisition of knowledge regarding the biological functions of RNA. Furthermore, the notion that RNA merely serves as a conduit for genetic information from DNA to protein manufacturers has been disproven; its function scope has been broadened. As an example, ribosomal RNA carries genetic information, tRNA transfers genetic code, and mRNA carries genetic information, all of which are crucial functions of RNA in molecular biology. Additionally, it facilitates protein localisation, regulates gene expression, and catalyses chemical reactions [1], among other functions. It also directs the site-specific modification of RNA nucleotides.

MSA fusion has become widely used in many different areas of bioinformatics. Multiple alignments are presented in most of the computational methods to help find sequence families, predict the gene regulation and polymerase chain reaction primer design [2], predict functions, and predict patients' diseases. Recent innovations have leveraged growing sequence data and computational power. Many nature-inspired algorithms, furthermore, implemented a few algorithms such as Genetic Algorithms (GA), Simulated Annealing (SA), Particle Swarm Optimization (PSO), Artificial Bee Colony Optimization (ABC), Biography-Based Optimization (BBO), and Chemical Reaction Optimization (CRO), which are used to solve MSA problems [3]. A random heuristic technique for solving MSA problems was recently discovered in PSO [4].

DOI: https://doi.org/10.54216/FPA.170201
Received: January 11, 2024 Revised: April 10, 2024 Accepted: September 14, 2024

Parallel methods to run traditional MSA programs have also emerged, achieving orders-of-magnitude speedup [5, 6]. GPU method [7] has been used to reduce the amount of computation. MSA's optimisation problem is NP-complete [14], which motivates researchers to use heuristics [8]. Optimisation procedures necessitate using an objective function (OF) to determine the optimal alignment. Furthermore, OF is a critical component in optimisation algorithms where the alignment score and quality are interdependent. The choice of the objective function is important to get high-quality alignments [9].

This paper aims to study and examine the comparative analysis of different objective functions using a standard dataset of RNA. This paper is organized as follows: Section 2 reviews the commonly used objective functions. Section 3 designs and explains the methodology used in this research. Section 4 explains the results and comparative used to assess the comparison. Lastly, Section 5 provides the conclusion of the paper and future work.

2. Related Work

An objective function (OF) is a mathematical function capable of measuring the biological quality of an alignment and defining its expected properties and correctness. The mathematically optimal alignment for a given ideal function will also be biologically optimal [10]. There are different OFs used to score the alignment, namely sum-of-pairs [11], weighted sum-of-pairs [12], NorMD [13], MstatX [14], and COFFEE [15]. These OFs are usually used in optimising and iterative alignment methods to improve the alignment by seeking the best value of the objective function [16]. A comprehensive review of all objective functions is not given here, but only those commonly used.

A. Sum-of-Pairs

The most common way to score the MSA is using Sum-of-Pairs (SP) and the most widely used [17]. Carrillo and Lipman [11] are the first to propose SP. The statistical independence of the columns of the alignment matrix in the SP is assumed by disregarding the phylogenetic tree. The total of the N (N-1)/2 pairwise alignment scores yield the score of a multiple alignment of N sequences according to the SP. The reference alignment is usually unknown before the SP is applied [14]. The formula for the standard version of SP, which has m columns, is found in Equation 1:

$$OF = \sum_{i=1}^{l} \{S_n(m_i) - G_n(m_i)\}, \quad (1)$$

The objective function is represented by OF, l is the sequence length, $G_n(m_i)$ denotes the column m_i gap penalty, and $S_n(m_i)$ represents the similarity score of column m_i. The similarity score of the column m_i is calculated as follows:

$$S_n(m_i) = \sum_{j=1}^{n-1} \sum_{k=j+1}^{n} s(m_i^j, m_i^k), \quad (2)$$

where the j^{th} row in the i^{th} column is denoted by m_i^j. Using the substitution matrix s(x,y), the similarity score for aligning two residues, x and y, is determined.

B. Weighted Sum-of-Pairs

The MSA program [18] uses the weighted sum-of-pairs (WSP) score produced by Altschul et al.[12]. Moreover, Gotoh [19] created an iterative multiple-sequence alignment method which optimises a WSP score. The WSP score is an expansion of the SP score in which the contribution to the overall score of each pairwise alignment is distinct. The fundamental idea is to provide a unique value (the substitution cost) to each pair of aligned residues in each alignment column and a different value (the gap cost) to each gap in the alignment. Therefore, the WSP score computes the overall score by combining all the sequences' weighted pairwise scores. The mathematical version of the WSP function is represented by the following equation:

$$WSP(A) = \sum_{0<i,j<n} W_{ij} \sum_{0<k<l} s(a_{ijk}, b_{ijk}), \quad (3)$$

where n is the number of sequences, k is the column position of the alignment, l is the length of the alignment A, and $W_{i,j}$ is the weight assigned to a pair of sequences; in contrast, the similarity cost of two symbol sequences, a_{ijk} and b_{ijk} is represented by $s(a_{ijk}, b_{ijk})$. Gap-open and gap-extend penalties are included in the cost function. The WPS likely arrange sequences so that the scores are proportionate to the information included in the sequence. These weights aim to reduce the impact of repetitive information from closely connected sequences [20]. This weight is equivalent to a percentage identity (PID) generated for each pair of matched sequences [20] by ignoring gaps as the following equation:

$$PID = \frac{matches}{matched + mismatched} \quad \textbf{Error! No text of specified style in document.}(4)$$

DOI: https://doi.org/10.54216/FPA.170201
Received: January 11, 2024 Revised: April 10, 2024 Accepted: September 14, 2024

C. Normalized Mean Distance

The mean distance (MD) score is adjusted into a normalised form known as adjusted Mean Distance (NorMD) [13]. The conservation-based score calculates the average distance between the similarities of residue pairs in each alignment column. Many techniques use NorMD, including RASCAL [21] and AQUA [22]. The continuous sequence space concept, given by [23], calculates a score for each column in the alignment. Next, the scores of each column are summed over the whole length of the alignment. NorMD is formally defined as follows:

$$NorMD = \frac{MD-GAPCOST}{MaxMD-LQRID} \quad (5)$$

where MD is the mean distance, GAPCOST is the affine gap cost, MaxMD is the maximum obtainable MD score, and LQRID (the lower quartile range of the pairwise hash score) is the similarity measure of sequences based on a hash score which is obtained from dot plots of pairs of sequences.

D. Consistency Score

Gotoh [24] has introduced the concept of consistency, which is refined by Vingron et al. [25]. The consistency-based approach improves match scores during the early alignment stage using pairwise alignment information. Later, this issue is reformulated as a maximum weight trace (MWT) problem by Kececioglu [26]. The consistency-based objective function COFFEE [15] represents the consistency between a given MSA and its library of pairwise alignments. Two things are needed for the COFFEE algorithm: an objective function that assesses the consistency of a multiple alignment against the library of pairwise alignments and a collection of pairwise alignments built using any method.

The T-Coffee [27], MAFFT [28], and Align-m [29] methods utilise consistency-based scoring, which involves using a tree-based consistency objective function for alignment evaluation. The COFFEE score is computed in the following formula:

$$COFFEE = \frac{\sum_{i=1}^{n-1}\sum_{j=i+1}^{n} w_{ij} * score(A_{ij})}{\sum_{i=1}^{n-1}\sum_{j=i+1}^{n} w_{ij} * len} \quad (6)$$

where len represents the length of the MSA, $W_{i,j}$ denotes the percentage identity of the two matched sequences S_i and S_j, $A_{i,j}$ indicates the sequences of S_i and S_j pairwise from the MSA, and the score $(A_{i,j})$ denotes the count of residue pairs common to $A_{i,j}$ and pairwise.

E. MstatX

Usually, the existence of the residue in MSA and its homologue may be used to determine its relevance. An often millions-of-year narrative of evolutionary pressure, mutation, recombination, and genetic drift is revealed by the patterns of residue variability in the data columns. MstatX [14] computes scores for the alignment of several sequences. Thus, it calculates comprehensive statistics for the alignment or conservation of each column. The statistical analysis in MstatX is conducted on individual columns rather than pairs of columns. The statistical metrics utilised in Mstatx are derived from sources [30] and [31]. MstatX performs several statistical calculations, including trident analysis, Shannon entropy measurement, frequency, and gap counts.

F. Objective Functions: A Summary

Several objective functions are included in other objective functions. Every objective function has benefits and drawbacks of its own. For instance, the Coffee has inherent limitations [32, 33]. One primary limitation is that it requires significant memory [9]. The second limitation is that it treats all pairs of residues the same. Even though substitution cost is thought about when the paired library is made, that does not mean that every pair is equally important where there may be parts of the alignment with higher similarity than others. The third limitation is that COFFEE does not care about position, and because it gives each pairwise alignment a weight, they are not all the same in terms of their importance. The fourth limitation is that the COFFEE weighting method gives both conserved and non-conserved parts the same weight. As a result, it cannot handle the noise that comes from both distant sequences and sequences that are very homologous to it. The fifth limitation is that COFFEE does not use extra gap penalties, so it does not care about the substitution scores of amino acids. Finally, COFFEE does not have a useful local function. On the other hand, many tools use some objective functions because they are accurate and good enough. One example is SP, the most common way to score because it is fast and reliable.

G. Harmony Search Algorithm

This research aims to optimise the score function using the harmony search algorithm (HS). HS is created by Geem [34]. HS is emulating a collective of musicians working together to achieve optimal harmony. Each participant

DOI: https://doi.org/10.54216/FPA.170201
Received: January 11, 2024 Revised: April 10, 2024 Accepted: September 14, 2024

generates a sound by selecting one of three options: memory consideration, pitch adjustment, or random selection. This is similar to finding the most efficient solution in the optimisation process. The HS algorithm has been applied to various optimization problems [35] that include Real-world applications, computer science problems, electrical engineering problems, civil engineering problems, mechanical engineering problems, and bio & medical applications.

HS comprises three distinct processes. The first is harmonic memory (HM), which stores high-quality harmonies. A harmony is randomly picked from the harmony memory using a parameter called harmony memory considering rate (HMCR), which ranges from 0 to 1. The usual range for HMCR is 0.7 to 0.95, according to Yang [36]. The second step involves pitch adjustment, similar to a local search. It generates a different solution from the HM, depending on the pitch-adjusting rate (PAR). PAR controls the degree of the adjustment by the pitch bandwidth (range) value. Generally, PAR values between 0.1 and 0.5 are widely used in most applications. The third process is random selection when a new harmony is created randomly to increase the diversity of the solutions. The pitch adjustment probability is determined by multiplying the HMCR (Harmony Memory Consideration Rate) with the PAR (Pitch Adjustment Rate). On the other hand, the probability of randomization is equal to 1 - HMCR [35]. The pseudo-code of the adapted fusion of HS for MSA (BHS-MSA) with these three components is summarised in the following algorithm.

Algorithm 1: Pseudo-Code of the BHS-MSA Method

```
Begin
Read the input sequences
Initial HS parameters: HMS, HMCR, and PAR
Define Objective function f(x)
For i=0 to HMS
  Generate feasible alignment randomly
  Evaluate the alignment by OF
  Find the best and worst alignment according to the OF
End for
While (t<max number of iterations)
   While (i<=number of variables)
        If (rand<HMCR)
             Choose a value from HM for the variable
            If (rand<PAR)
                Adjust the value by adding a certain amount
            End if
           Else choose a random value
        End if
   End while
  Accept the new harmony if it is better than the worst
End while
Find the current best solution
End
```

The pseudocode of utilising the HS algorithm to resolve the MSA issue is illustrated in Figure 1. This approach is based on the optimisation of the scoring function. HM, score function, termination condition, and improvised new solution are the primary components of HS [36].

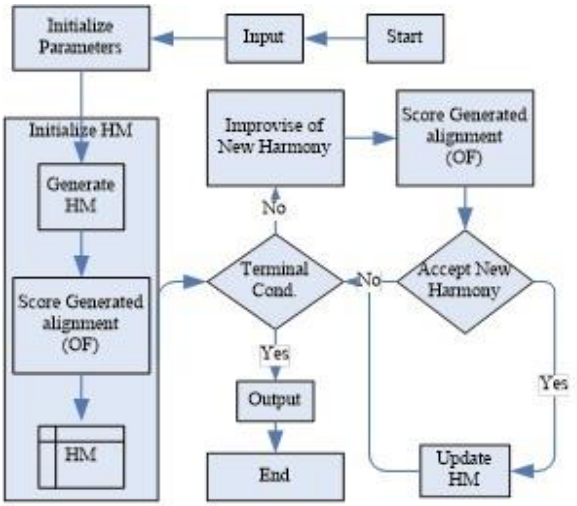

Figure 1. The HS-MSA Method

This algorithm iteratively improves solutions by creating new harmonies, adjusting them based on historical information stored in the harmony memory, and accepting the improved solutions.

DOI: https://doi.org/10.54216/FPA.170201
Received: January 11, 2024 Revised: April 10, 2024 Accepted: September 14, 2024

3. Experimental Methodology

An objective function must be defined to solve any problem using optimization methods. The objective function is a crucial component of iterative and optimization algorithms as it dictates the optimal move to enhance the solution quality. The objective function is crucial in guiding the development of a mature alignment in multiple sequence alignment. This analysis aims to verify if the aligned sequences are optimal and to calculate the alignment score without prior knowledge of the reference alignment.

A. Benchmark Dataset

Tables 1 and 2 show the details of the dataset and the description information about each test set. Two types of datasets are chosen based on Wang et al. [37] (i) The subset of BRAliBase that is highly variable and appropriate for local multiple sequence alignments (MSA), and (ii) LocalEXtR, an expanded version of BRAliBase 2.1 that includes large-scale test groups. The subset of BRAliBase 2.1 is chosen from the dataset with the highest variability in the suite. The individuals belong to the THI, Glycine riboswitch, and Yybp-Tkoy RNA families, consisting of 232 test datasets. LocalExtR uses the same seed alignments from Rfam that BRAliBase uses and forms large test groups. BRAliBase is labelled a test group k_i, where i is the number of sequences for each test set.

Table 1: The RNA Families Dataset

Test Group		gcvT	THI	yybp-ykoy	Total
BRALiBase (232 datasets)	k5	22	69	33	124
	k7	12	32	18	62
	k10	3	17	12	32
	k15	1	5	8	14
LocalExtR (90 datasets)	k20	10	10	10	30
	k40	10	10	5	25
	k60	10	10	0	20
	k80	5	10	0	15
Total		73	163	86	322

Table 2: Sequence Length of Each Test Group

Test Group		Sequence length		
		Avg.	Min.	Max.
BRALiBase (232 datasets)	k5	109	96	125
	k7	110	94	131
	k10	108	94	129
	k15	110	88	137
LocalExtR (90 datasets)	k20	115	90	172
	k40	114	87	180
	k60	107	81	189
	k80	106	77	204

B. Alignment Quality

Alignment quality necessitates a reference alignment obtained from the database benchmark. The quality is determined by comparing the alignment of the test and reference. Qscore [38] is a function that estimates the comparison between the test and reference alignments. The Q (Quality) score is calculated by dividing the number of successfully aligned residue pairs in the test alignment by the number of residue pairs in the reference alignment.

C. Experimental Design

In this study, different objective functions are investigated. Theoretically, the objective function should always assign a high score to the alignment with a superior quality. Therefore, a correlation test often measures the association between the alignment score and the alignment quality.Two experiments have been conducted to examine the efficacy of the respective objective functions using the modified harmony search algorithm (HS). These studies examine the strength of the link between the alignment score and alignment quality for each objective function individually. A significant correlation value indicates a strong positive relationship between the objective function and the variable being studied. Consequently, the objective functions with higher ranks are also organized according to their execution time, with a preference for shorter execution time.

DOI: https://doi.org/10.54216/FPA.170201
Received: January 11, 2024 Revised: April 10, 2024 Accepted: September 14, 2024

The first experiment uses a harmony search method to produce alignments based on the objective function of the chosen dataset. The objective function in the harmony search algorithm assesses and orders harmonies in the harmony memory based on their score values. Harmonies with high alignment scores in the harmony memory are kept. Each experiment evaluates the association between a specific objective function and the alignment quality value. For every objective function score and alignment quality score, regression coefficients (R^2) are computed, as shown in Figure 2.

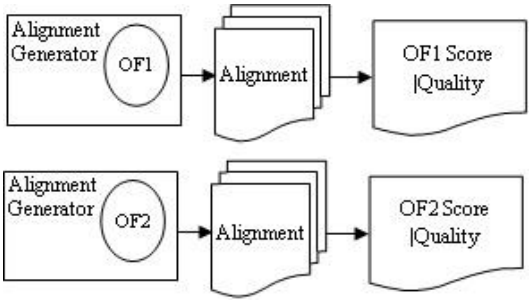

Figure 2. Experimental Design of the Objective Function Correlation.

In the second experiment, alignment is measured by different objective functions separately. The time consumed is calculated for each objective function in the same alignment. This scenario compares the consuming time of the different objective functions on the same alignment, as shown in Figure 3.

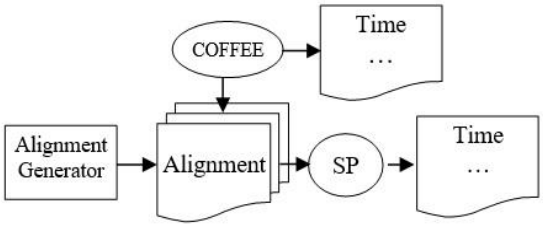

Figure 3. Experimental Design for Calculating the COFFEE and SP Time.

4. Results and discussion

The biggest challenge associated with MSA is accurately assessing the quality of computer-aligned sequences. Hence, the optimization methods need an objective function (OF). The quality of alignments is determined by selecting the objective function [13].

A. Relationship between alignment score and alignment quality

The HS-MSA algorithm utilises the SP, WSP, COFFEE, MstatX, and NorMD individually as objective functions. Regression coefficients (R^2) correlate the alignment quality and objective function values. Through comparison with the references, the evaluation function quality (Q) score is used to evaluate the final alignments. Every objective function has its default settings for the HS and MSA parameters, except for SP, which is set as follows: the substitution matrix is the muscle matrix [51], the gap open penalty is -10, and the gap extension penalty is -0.5. The dataset's average R^2 coefficient values are shown in Table 3. The R^2 values for COFFEE and SP are 0.9642 and 0.9216, respectively which are the best.

Table 3: objective functions relationships

	SP	WSP	NorMD	Mstatx	COFFEE
Regression coefficient (R^2)	**0.922**	0.728	0.761	0.826	**0.96**

Figure 4 displays the graphs of linear regression. The results demonstrate that COFFEE and SP have the most robust linear association with alignment quality compared to the other objective functions.

DOI: https://doi.org/10.54216/FPA.170201
Received: January 11, 2024 Revised: April 10, 2024 Accepted: September 14, 2024

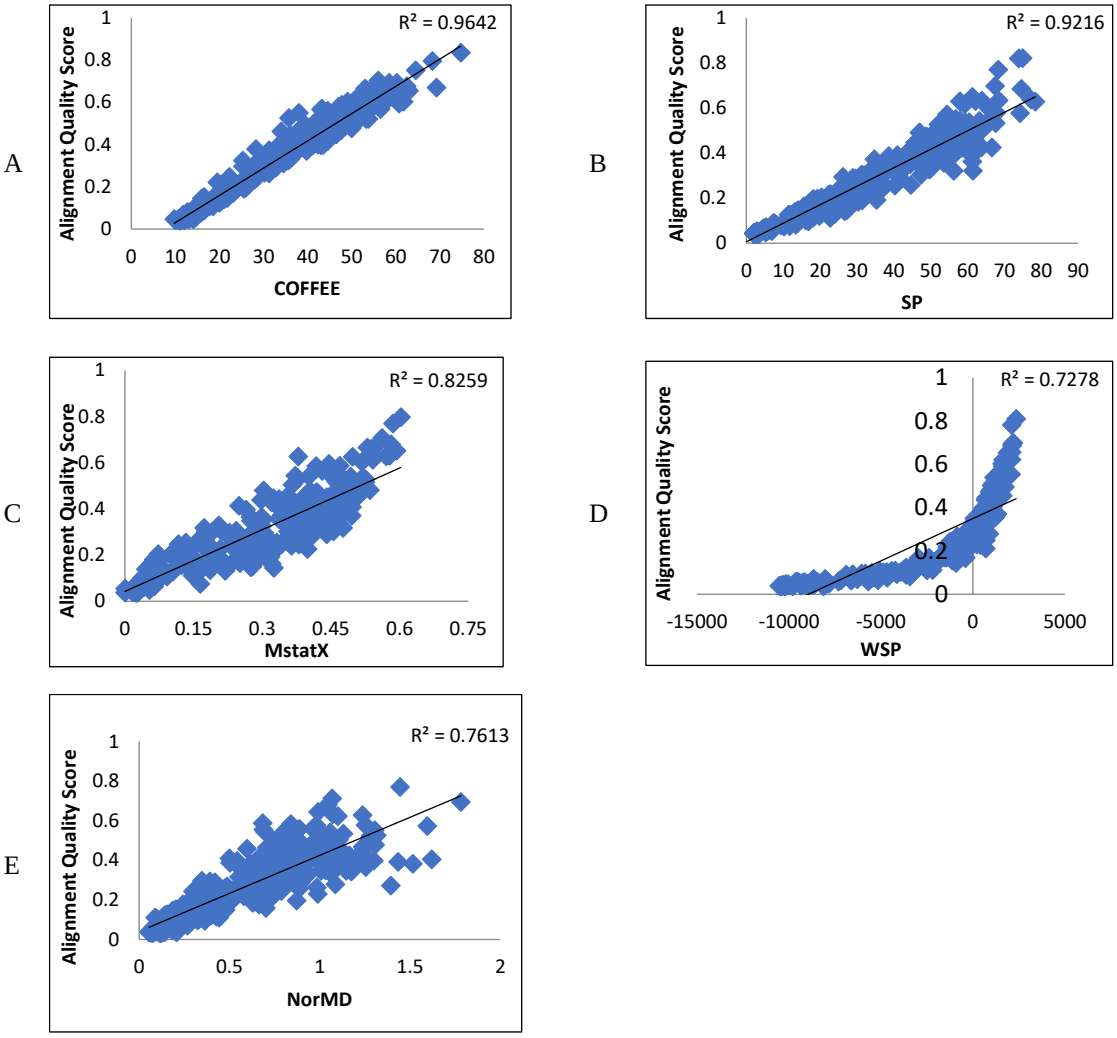

Figure 4. Regression between the Objective Function and Alignment Quality

B. An analysis of the runtime performance of COFFEE and SP

The most computationally demanding part is the scoring function in MSA. The time required to calculate alignment scores increases linearly with the number of sequences and the alignment length. The final alignment is measured separately by SP and COFFEE in this experiment, and the time for each measurement is calculated.

Table 4 and Figure 5 demonstrate that the run time of the COFFEE and SP functions increases as the number of sequences increases. The average time required to calculate the alignment score for ten trials is ascertained.

Table 4: The coffee and SP objective functions' execution time

Test Group	Sequences No.		Sequence Length			Avg. Time (MS)	
			mean	min	max	SP	COFFEE
BRALiBase (232 datasets)	k5	5	109	96	125	0.16	460.00
	k7	7	110	94	131	0.32	533.45
	k10	10	108	94	129	0.66	735.30
	k15	15	110	88	137	1.60	1242.49
	k20	20	115	90	172	3.52	1914.43
LocalExtR (90 datasets)	k40	40	114	87	180	16.96	6568.22
	k60	60	107	81	189	42.72	13799.52
	k80	80	106	77	204	88.01	25278.69

DOI: https://doi.org/10.54216/FPA.170201
Received: January 11, 2024 Revised: April 10, 2024 Accepted: September 14, 2024

Although the COFFEE function shows a slightly higher correlation than the SP function. The COFFEE required an average run time of 6316.51 Milliseconds per alignment group, while the SP averaged 19.24 Milliseconds, indicating a significant difference in computational efficiency. This may be because increasing sequence numbers complicates the objective function [14]. The SP method is mainly faster because it does not need a tree [22].

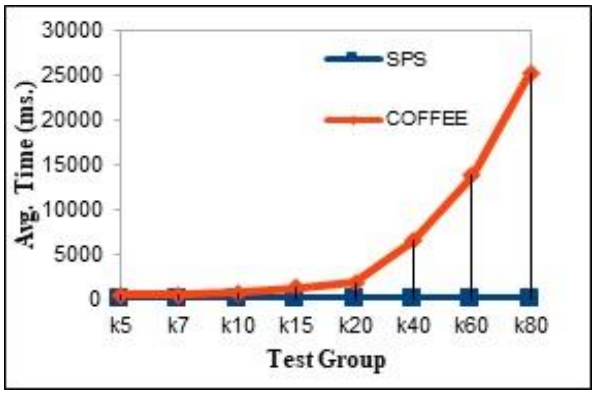

Figure 5. Run Times of COFFEE and SP Objective Functions with Number of Sequences.

5. Conclusion and Future Work

The alignment of multiple sequences remains a challenging problem due to factors such as quality, the computational complexity of aligning large datasets, and the consuming time. This paper does not discuss possible strategies to improve alignment quality; instead, it focuses on the current objective functions used to score the alignments. The relationship between the alignment score and alignment quality of different objective functions is the goal of this study, which is called correlation. Using different objective functions and comparing their correlation of score and quality is recommended to find the most suitable one. The study confirmed a strong positive relationship between the alignment scores and the quality of the alignments produced by the different objective functions. This correlation underscores the importance of selecting appropriate objective functions to achieve high-quality RNA alignments. The study evaluated several objective functions, including COFFEE, Sum-of-Pairs (SP), Weighted Sum-of-Pairs (WSP), NorMD, and MstatX, in the context of RNA multiple sequence alignment. The results indicated that COFFEE and SP exhibited the highest correlation with alignment quality, with regression coefficients (R^2) of 0.96 and 0.92, respectively. While COFFEE provided slightly better alignment scores, the SP objective function was significantly faster across all test groups. This efficiency makes SP a more practical choice for applications requiring rapid alignment processing. Based on this comparison, the SP objective function can be selected for further investigation in modifying the HS processes and hybridization of HS with other MSA approaches in future work to improve alignment quality. Moreover, future research could explore the development of hybrid objective functions that combine the strengths of multiple scoring methods to enhance alignment accuracy across a broader range of RNA types.

Funding: "The authors would like to express their thanks to University Tun Hussein Onn Malaysia (UTHM) for support. This research is supported by University Tun Hussein Onn Malaysia (UTHM) through Tier 1 (vot Q547)".

Conflicts of Interest: "The authors declare no conflict of interest."

References

[1] J. A. Doudna, and T. R. Cech, "The Chemical Repertoire of Natural Ribozymes," *Nature,* vol. 418, no. 6894, pp. 222-228, 2002.

[2] L. Anbarasu, P. Narayanasamy, and V. Sundararajan, "Multiple molecular sequence alignment by island parallel genetic algorithm," *Current Science,* vol. 78, no. 7, pp. 858-863, 2000.

[3] T. Paruchuri, G. R. Kancharla, S. Dara, R. K. Yadav, S. S. Jadav, S. Dhamercherla, and A. Vidyarthi, "Nature Inspired Algorithms for Solving Multiple Sequence Alignment Problem: A Review," *Archives of Computational Methods in Engineering,* vol. 29, no. 7, pp. 5237-5258, 2022.

[4] T. Paruchuri, G. R. Kancharla, and S. Dara, "Solving multiple sequence alignment problems by using a swarm intelligent optimization based approach," *International Journal of Electrical and Computer Engineering,* vol. 13, no. 1, pp. 1097-1104, 2023.

[5] Q. Bani Baker, R. A. Al-Hussien, and M. Al-Ayyoub, "Accelerating Multiple Sequence Alignments Using Parallel Computing," *Computation,* vol. 12, no. 2, pp. 32, 2024.

DOI: https://doi.org/10.54216/FPA.170201
Received: January 11, 2024 Revised: April 10, 2024 Accepted: September 14, 2024

[6] S. H. Almanza-Ruiz, A. Chavoya, and H. A. Duran-Limon, "Parallel protein multiple sequence alignment approaches: a systematic literature review," *Journal of Supercomputing,* vol. 79, no. 2, pp. 1201-1234, 2023.
[7] D. Sundfeld, and A. C. M. A. De Melo, "MSA-GPU: Exact multiple sequence alignment using GPU." pp. 47-58.
[8] J. D. Thompson, D. G. Higgins, and T. J. Gibson, "Clustal-W - Improving the Sensitivity of Progressive Multiple Sequence Alignment through Sequence Weighting, Position-Specific Gap Penalties and Weight Matrix Choice," *Nucleic Acids Research,* vol. 22, no. 22, pp. 4673-4680, Nov 11, 1994.
[9] T. Lassmann, and E. L. L. Sonnhammer, "Automatic assessment of alignment quality," *Nucleic Acids Research,* vol. 33, no. 22, pp. 7120-7128, 2005.
[10] R. Durbin, *Biological sequence analysis: probabilistic models of proteins and nucleic acids*, Cambridge, UK.: Cambridge university press, 1998.
[11] H. Carrillo, and D. Lipman, "The Multiple Sequence Alignment Problem in Biology," *Siam Journal on Applied Mathematics,* vol. 48, no. 5, pp. 1073-1082, Oct, 1988.
[12] S. F. Altschul, R. J. Carroll, and D. J. Lipman, "Weights for Data Related by a Tree," *Journal of Molecular Biology,* vol. 207, no. 4, pp. 647-653, Jun 20, 1989.
[13] J. D. Thompson, F. Plewniak, R. Ripp, J. C. Thierry, and O. Poch, "Towards a reliable objective function for multiple sequence alignments," *Journal of Molecular Biology,* vol. 314, no. 4, pp. 937-951, Dec 7, 2001.
[14] C. G. Gcollet, "MstatX: A Multiple Alignment Analyser. GitHub," 2010.
[15] C. Notredame, L. Holm, and D. G. Higgins, "COFFEE: An objective function for multiple sequence alignments," *Bioinformatics,* vol. 14, no. 5, pp. 407-422, 1998.
[16] J. D. Thompson, and O. Poch, "New challenges and strategies for multiple sequence alignment in the proteomics era," *The proteomics protocols handbook,* pp. 475-492, 2005.
[17] H. B. Nicholas, A. J. Ropelewski, and D. W. Deerfield, "Strategies for multiple sequence alignment," *Biotechniques,* vol. 32, no. 3, pp. 572-591, Mar, 2002.
[18] D. J. Lipman, S. F. Altschul, and J. D. Kececioglu, "A Tool for Multiple Sequence Alignment," *Proceedings of the National Academy of Sciences of the United States of America,* vol. 86, no. 12, pp. 4412-4415, Jun, 1989.
[19] O. Gotoh, "A Weighting System and Algorithm for Aligning Many Phylogenetically Related Sequences," *Computer Applications in the Biosciences,* vol. 11, no. 5, pp. 543-551, Oct, 1995.
[20] F. B. R. Zablocki, "Multiple Sequence Alignment using Particle Swarm Optimization," Department of Computer Science, University of Pretoria, 2007.
[21] J. D. Thompson, J. C. Thierry, and O. Poch, "RASCAL: rapid scanning and correction of multiple sequence alignments," *Bioinformatics,* vol. 19, no. 9, pp. 1155-1161, Jun 12, 2003.
[22] J. Muller, C. J. Creevey, J. D. Thompson, D. Arendt, and P. Bork, "AQUA: automated quality improvement for multiple sequence alignments," *Bioinformatics,* vol. 26, no. 2, pp. 263-265, Jan 15, 2010.
[23] M. Vingron, and P. R. Sibbald, "Weighting in Sequence Space - a Comparison of Methods in Terms of Generalized Sequences," *Proceedings of the National Academy of Sciences of the United States of America,* vol. 90, no. 19, pp. 8777-8781, Oct 1, 1993.
[24] O. Gotoh, "Consistency of Optimal Sequence Alignments," *Bulletin of Mathematical Biology,* vol. 52, no. 4, pp. 509-525, 1990.
[25] M. Vingron, and P. Argos, "Motif Recognition and Alignment for Many Sequences by Comparison of Dot-Matrices," *Journal of Molecular Biology,* vol. 218, no. 1, pp. 33-43, Mar 5, 1991.
[26] J. Kececioglu, "The maximum weight trace problem in multiple sequence alignment," *Combinatorial Pattern Matching*, Lecture Notes in Computer Science A. Apostolico, M. Crochemore, Z. Galil and U. Manber, eds., pp. 106-119: Springer Berlin Heidelberg, 1993.
[27] C. Notredame, D. G. Higgins, and J. Heringa, "T-Coffee: A novel method for fast and accurate multiple sequence alignment," *Journal of Molecular Biology,* vol. 302, no. 1, pp. 205-217, Sep 8, 2000.
[28] K. Katoh, and H. Toh, "Recent developments in the MAFFT multiple sequence alignment program," *Briefings in Bioinformatics,* vol. 9, no. 4, pp. 286-298, Jul, 2008.
[29] I. Van Walle, I. Lasters, and L. Wyns, "Align-m - a new algorithm for multiple alignment of highly divergent sequences," *Bioinformatics,* vol. 20, no. 9, pp. 1428-1435, Jun 12, 2004.
[30] W. S. J. Valdar, "Scoring residue conservation," *Proteins: Structure, Function, and Bioinformatics*, vol. 48, no. 2, pp. 227-241, 2002.
[31] J. Fredrik, and T. Hiroyuki, "A comparative study of conservation and variation scores," *BMC Bioinformatics,* vol. 11, pp. 388, 2010.

DOI: https://doi.org/10.54216/FPA.170201
Received: January 11, 2024 Revised: April 10, 2024 Accepted: September 14, 2024

[32] T. Riaz, Y. Wang, and L. Kuo-Bin, "A Tabu Search Algorithm for Post-Processing Multiple Sequence Alignment," *Journal of Bioinformatics & Computational Biology*, vol. 3, no. 1, pp. 145-156, 2005.
[33] Y. Wang, T. Riaz, E. ENERLY, S. ZENG, and K. B. Li, "A Position-Specific and Consistency-Based Objective Function for Iterative Multiple Sequence Alignment."
[34] Z. W. Geem, J. H. Kim, and G. V. Loganathan, "A new heuristic optimization algorithm: Harmony search," *Simulation*, vol. 76, no. 2, pp. 60-68, Feb, 2001.
[35] G. Ingram, and T. Zhang, "Overview of applications and developments in the harmony search algorithm," *Music-Inspired Harmony Search Algorithm*, pp. 15-37, 2009.
[36] X.-S. Yang, "Harmony Search as a Metaheuristic Algorithm," *Music-Inspired Harmony Search Algorithm*, Studies in Computational Intelligence Z. Geem, ed., pp. 1-14: Springer Berlin Heidelberg, 2009.
[37] S. Wang, R. R. Gutell, and D. P. Miranker, "Biclustering as a method for RNA local multiple sequence alignment," *Bioinformatics*, vol. 23, no. 24, pp. 3289-3296, Dec 15, 2007.
[38] R. C. Edgar, "MUSCLE: a multiple sequence alignment method with reduced time and space complexity," *Bmc Bioinformatics*, vol. 5, pp. 1-19, Aug 19, 2004.

DOI: https://doi.org/10.54216/FPA.170201
Received: January 11, 2024 Revised: April 10, 2024 Accepted: September 14, 2024