



A Novel Population-based Optimization for Multiple Sequence Alignment in Protein Sequencing

Anjali Goswami¹ and Kamlesh Kumar Dubey^{2,*}

Abstract

This paper presents a novel approach to tackle the multiple sequence alignment (MSA) problem using Population-based optimization (PBO) in conjunction with an improved migration operator. The proposed PBO algorithm is based on genetic algorithms and aims to solve problems related to protein sequences. Current PBO methods have a limitation in the lack of population diversity, which results in slow convergence and lack of population variety. The improved migration operator used in this study addresses this issue by absorbing more information from various habitat types and maintains population variation, thus exploring new possibilities. The performance of the proposed approach has been compared to that of many current techniques, including VDGA, MOMSA, and GAPAM, using publicly available benchmark datasets. The results show that the proposed approach outperforms them all, specifically the Bali base. The comparison has been performed over more than two hundred iterations and demonstrates superior performance in terms of fitness function convergence. In addition, the proposed method uses a multidimensional search space and utilizes both exploration (diversity) and exploitation capabilities, helping to prevent the trap of local minima and explore the needed options. This leads to improved alignment and sequencing. Overall, the proposed approach demonstrates superior performance compared to existing techniques in most cases. The proposed PBO algorithm is a unique method that provides a new perspective on solving MSA problems, which can have a significant impact on the field of bioinformatics and computational biology. It has been observed the suggested approach competitive and outperforms the existing techniques.

Keywords: Population-based optimization; Population-based algorithm, Population variation; Genetic algorithms.

Received: 05 May 2022; Revised: 05 November 2022; Accepted: 07 November 2022.

Article type: Research article.

1. Introduction

MSA is a term that refers to the alignment of more than three amino acid sequences or protein sequences at the same time, which is known as multiple sequence alignment. Molecular diagnostics (MSA) is the important method in the treatment of biological problems. With the assistance of MSA, present study can address a wide range of biological issues and challenges. Multiple sequence alignment (MSA) can aid in the prediction of secondary and tertiary structure in proteins by predicting their secondary and tertiary structure. The MSA can be used to reassemble phylogenetic trees, and it can also be used to predict the function of an unknown amino acid by comparing its sequence with the sequences of other amino

acids that are known to perform specific functions. Beside from determining sequence similarity, MSA can also be used to determine similarities between functions and structures, which can aid in the definition. In order for a MSA to be considered legitimate, all of the multiple alignments must originate from the same place in the genome. The objective is to increase the amount of acid that matches the target as closely as possible. This has resulted in MSA being regarded as a crucial topic in bioinformatics for the investigation of genetic and phylogenetic relationships among organisms.^[1,2,3] A variety of approaches to resolving MSA issues have been tried in the past, with varying degrees of success.

It has been demonstrated that MSA is an effective method for a wide variety of scientific endeavors, including the reconstruction of phylogenetic trees, the illuminating of functionally relevant areas, and the prediction of higher order structures of proteins and Ribonucleic acid (RNA). However, mechanically constructing a multiple alignment from a group of similar sequences is a very simple task that can be done with little effort. A wide range of approaches to resolving this

¹ Department of Mathematics and Statistics, College of Science and Theoretical Studies, Saudi Electronic University, Riyadh-13323.

² Department of Applied Science and Humanities, Invertis University, Bareilly, India-243123.

These authors contributed to this work equally.

*E-mail: kkdubey4maths@gmail.com (K. Dubey)

computationally challenging issue are discussed in this article. There are several significant applications of multiple alignment that are explored here to shed light on the linkages.

The application of dynamic programming (DP) can be used to address MSA problems and produce the best possible alignment outcomes. When it comes to DP, it employs a scoring formula that has a wide range of potential applications. Needleman and Wunsch were the first to propose the use of DP algorithms to solve when they published an article in 1970 proposing the DP algorithms for two sequence alignment.^[1-3] However, difficulty with using DP is that as with quantity and sequences increase, the complexity of the sequences increases in rapidly as well, which makes it difficult to maintain a consistent level of performance. After that, the MSA problem is considered to be NP-complete. This is due to the fact that complexity is one of the most significant constraints on the computer's capacity to complete any activity. The present study therefore must optimize the matching sequences with a short period of time or with a low degree of complexity. It's one of the major reasons researchers are increasingly turning to alternative approaches.^[4-6]

The progressive approach can also be to deal with the MSA problem in some cases.^[7] Compared to the conventional approach, the progressive strategy for tackling the MSA problem is less time-consuming and spatially complex. Progressive alignment is a method of aligning sequences that are more similar to one another in the first alignment, and then increasingly more divergent sequences or groups of sequences are aligned in the second alignment, and so on. It is widely considered that CLUSTALW^[8] is the default representation of progressive methods. As the first step in the procedure, the initial stage of this approach consists of assigning weights to each pair of sequences in a partial alignment, which serves as the starting point for the technique. In order to distinguish between sequences that are most similar to each other and sequences that are most distinct, they assign tiny weights to the sequences that are most similar. Afterwards, it creates a substitution matrix, which describes the difference in rank among two components of a protein sequence based on their similarity to each other in the sequence. A total of two different types of gaps were introduced during the third stage. The first type of gap penalty is referred to as residue-specific gap penalties, whereas the second type is referred to as locally residue gap penalties. In the fourth phase, those spots where a gap has been established at an early stage get locally reduced gap penalties in order to stimulate the opening of gaps at these locations. CLUSTALW, which is open to the public, includes these four stages into the design of its website. Overall, the progressive alignment technique beats all other methods for the MSA package in terms of accuracy and turnaround time. Although this technique has certain advantages, it also has some substantial drawbacks. However, the issue with this strategy is that it is very dependent on both the beginning alignment and the scoring mechanism that is being employed. In other words, the present study limited the alignment to

sequences that were more similar to one another in the initial phase. If the present study did not align any further similar sequences during the initial phase, the solution may become entangled in local optimality, which would be undesirable. When dealing with numerous sequence alignment difficulties, an iterative method can be utilized to solve the problem more effectively. This is because iterative techniques do not rely on initial alignment since they begin with initial alignment and continue to improve the answers with each iteration until there is no more improvement to be made. In the iterative approach, its primary objective is to improve the overall quality of a series of sequence alignments. The HMMT approach is based on a simulated annealing process.^[9-11] It is possible that the issue that underlies the solution of these techniques is trapped in a local optimal solution state. Evolutionary algorithms (EAs) are algorithms that are based on the population of a certain species or species group.^[12] Creating a random starting population for these algorithms is the first stage in the process, which is completed in the first phase. As a further stage, the present study will apply some operators to the initial population in order to change it for the future generation of the population. It repeats the process of using these operators over and again until the present study reach the global optimum. EAs for MSA are employed in two ways: first, a random starting generation is created, and then the steps of an EAs are applied to enhance the similarity between the sequences; second, the steps of an EAs are used to increase the similarity between the sequences. There have been several works where evolutionary calculations for MSA^[13-15] have been done. In addition to the techniques described above, some alternative genetic algorithm (GA)^[16] based methods for MSA have been suggested, including SAGA. It is possible to utilize a genetic algorithm to build an approach for tackling particular algorithm MSA issues (GA). In SAGA, the first generation is generated using a random number generator (RNG). According to SAGA, a total of 22 distinct operators are used to gradually enhance the fitness of MSA over a period of time. The SAGA problem, on the other hand, is caused by the increased temporal complexity generated by the repetitive application of the fitness function.^[17,18] RBT-GA is another GA-based approach that is used in conjunction with the rubber band technique (RBT) to discover the best protein sequence alignments. RBT-GA^[19,20] is a GA-based method that is used to find the optimal protein sequence alignments. RBT is an iterative method for sequence alignment that starts with a DP table as a starting point and iterates over the sequence. In [21], the authors solved 56 problems from reference sets 1, 2, 3, 4, and 5 of the benchmark Bali base 2.0 dataset and the Bali base 3.0 dataset. The authors used the benchmark Bali base 2.0 dataset as a starting point for their work. As a consequence of a lack of variety in the solutions, evolutionary techniques have disadvantage are that they create local optima as a result of their use.

A method known as the population based approach is an additional option that might be used to resolve the MSA

problem. The population based technique needs far less complexity in terms of both time and area in order to resolve the MSA problem. In the first step of the alignment method, the sequences that are most comparable to one another are brought into alignment with one another. After that, you will increasingly align the sequences in the beginning alignment that are the most distinct from one another.

Since it can handle a broad variety of common difficulties such as structure prediction, phylogenetic property, and so on, Multiple Sequence Alignment (MSA) has become the most significant instrument in the area of biology. MSA is still considered a tough topic to tackle, according to the current research. As a result, the present study are inspired to develop an improved version of PBO to address the MSA problem. These important contributions are made by this study, in contrast to the previous one.

1. The present study provided a method for improving the migration operator in PBO for the first time, and then applied it in MSA to ensure that the solutions stay diverse.
2. Experimental analysis produces more accurate results than observational analysis when timing factors are taken into consideration. As an additional point of reference, it provide a comparison table in which claim that the present methodology beats existing rival methods in terms of matching score.

In this section, the present the outline of our work. As in first section, the present study presented the introduction with preliminaries requires. In Section. 2, the methodology of the proposal has explained. In section. 3, the result and discussion followed by the conclusion in section. 4.

2. Methodology

The Population Based Optimization (PBO) method have several ways to proceeds in this study. There is the interesting observation which has been made on gravitational force (GF) action upon the masses. The gravitational force applied upon the any object is directly proportional to its mass. The greater the mass the effect of pull on object will be greater according to the gravitational force (GF). The force directly proportional to its mass is rule based on the star and satellites revolves around the galaxy. The gravity is basically the attraction force; therefore, the present study can see in the term of this law in several application of data structure and algorithms. Suppose there is several files in any input and output (i/o) system which requires the sorting. The simple PBO based iteration based on GF is that the file which is greater in size will sort first and it will follow by second largest and so on. In each iteration, the large file will be sorted fully before making the transition towards the sort of file. In our PBO, the subset of population will be considered as sub-swarm. The whole population first divide into the sub groups which individually behave like object in the universe.

A program known as the gravitational search optimization (GSO) technique mimics the motion of galaxies and stars inside an infinite universe. It is not conceivable to have a uniform distribution of stars across the universe; yet,

they are mostly concentrated in galaxies, which are likewise not evenly distributed. In the GSO algorithm, above - the method is combined with the gravity of galaxies and stars to get the desired result. First, the entire population is divided into subpopulations, which are referred to as sub-swarms in this context. According to the PBO method, each individual of sub-swarms starts their movement by iterating a pre-determined number of times before moving forward. Because of the improved fitness values, all entities within each sub-population will be drawn to the specified one. At the conclusion of the iterations, the well as on each and every subpopulation would be presented as the best individual from every single subpopulation. A second step (similar to PSO) will be used to choose the best individuals from all subpopulations that have been clustered together by new super swarms. The greatest solution provided by the PBO is also the highest scoring super swarm from the total starting population, which is another benefit of the PBO.

An algorithm that develops populations by making use of the gravitational attraction that occurs between people is referred to as a gravitational search algorithm, which is also referred to by its acronym, GSA. Although it is an effective population-based algorithm, GSA suffers from subpar search performance and early convergence. The fact that the algorithm is based on populations is the root cause of both problems. As a potential remedy to these issues, the authors of this study suggest using a multi-layered variation of GSA that they refer to as ML-GSA. The structure of the GSA, which consists of two layers, provided as inspiration for the building of four levels, which included population, iteration-best, personal-best, and global-best layers. The structure of the GSA is comprised of two layers. The implementation of hierarchical interactions among four search layers, which are dynamically implemented at various stages of the search process, can significantly benefit a population's capacity to explore new territory and exploit existing resources. Hierarchical interactions are implemented dynamically at various stages of the search process.

Following Pseudo code performs the necessary operations of Population Based Optimization (PBO)

1. Initialize PBO
2. Divide population P in subgroup SG_1, SG_2, \dots, SG_n in random manner as shown in Eq. (1)

$$SG_i \subset P: i = 1, 2, \dots, n \quad (1)$$

3. Lets initialize the population with randomness as Eq. (2)

$$SG_j^{(i)} \in SG_i: j = 1, 2, \dots, r \quad (2)$$

4. Let define the function f for the population as subgroup.
5. It define as $f(lb^{(i)})$ and $f(gb^{(i)})$ as function of local best subgroup and and globalbest sub group of swarm. When the best value of $P_j^{(i)}$ have significance less then $g^{(i)}$, the scenerio will force all the sub swarm X_i which is directly linked with globel best $g^{(i)}$ to get updated as Eq. (3)

$$f(lb^{(i)}) < f(gb^{(i)}) \quad (3)$$

6. Intilaize the state Level 1. The M represent the sub

population which will be initialize once the velocity and position has been computed. It will start the PSO procedure as described in Eqs. (4-5).

$$v_j^{(i)} \leftarrow \omega_1 v_j^{(i)} + c_1 r_1 (p_j^{(i)} - x_j^{(i)}) + c_2 r_2 (g^{(i)} - x_j^{(i)}) \quad (4)$$

$$x_j^{(i)} \leftarrow x_j^{(i)} + v_j^{(i)} \quad (5)$$

where r_1 and r_2 has been the weight assigned as initialization in Eq. (6)

$$\omega_1 = 1 - \frac{k}{L_1 + 1} \quad (6)$$

Our approach will iterate over k in the range of 0 to L_1 subject to

where $-1 < r < 1$ as Eq. (7)

$$r_1 = U(-1, 1) \quad (7)$$

End PSO

End level 1.

7. Start Level 2.

SSI phase initialization as Eqs. (8-9).

$Y^{(i)} \in Y: i = 1, 2, \dots, M$

Activate PSO.

Parameters calculation.

$$v^{(i)} \leftarrow \omega_2 v^{(i)} + c_3 r_3 (p^{(i)} - Y^{(i)}) + c_4 r_4 (g - Y^{(i)}) \quad (8)$$

$$Y^{(i)} \leftarrow Y^{(i)} + v^{(i)} \quad (9)$$

end PSO.

8. The outcome will be the fitness value $f(g)$ and appropriated position g .

9. End GSO.

The gravitational constant G in GSA will gradually and exponentially diminish over the course of time using a decay constant. This will occur over the period. Because this function does not support its own self-adaptation, including it in the method will lead the method to become dependent on yet another argument, which is something that the present study does not suggest doing. When the GSA is computing the overall force that was applied to a body, it assigns a weighting to each component (the force that was applied by each other body) based on a random integer. This allows the GSA to arrive at the total force that was applied. This is done to expand the GSA's powers of exploration, and it is being done as such right now. In the worst-case scenario, the metaphor that underlying this circumstance, the gravitational interaction, will be lost. This is because the situation will remain as it is. Even though the complexity of the algorithm has not changed, GSA still makes use of best agents to cut down on the amount of time that is spent computing. This is even though the amount of time spent computing has not changed. On the other hand, the present study makes it possible for all the masses, who are also referred to as agents, to have conversations with one another. The CSS algorithm attempts to imitate the electrostatic dynamics by assigning charge to the particles in the same way that the GSA algorithms assign mass to the bodies to optimize the performance of unimodal and multimodal functions. Therefore, the iteration will depend upon the number of element available in the best pool. In which the sequence is repeated, and domains are reversed,

addition, the time complexity can be $O(n^{\log k})$ where, k will be best available agent usually greater or equal to 1. This is done in order to maximize the efficiency of these functions. This is done to ensure that the algorithm operates as effectively as possible. These are of the opinion that the high degree of detail of the charges of the particles in CSS is unnecessary, and the parameter and is quite big for certain functions; the calculation of this parameter needs a considerable bit of study to be conducted.

Test Dataset

In this study, the proposal has been tested against a large number of datasets from the Bali base benchmark database in order to ensure that it was effective over a broad range of datasets from the Bali base benchmark database. The Bali base version 1.0 has a total of 142 reference alignments, each of which contains more than 1000 sequences that were previously preserved inside the alignments itself. Balinese basic version 2.0 is an improved version of the Bali base version 1.0 that was first published in 2007. It is a more comprehensive upgrade of the Bali base version 1.0. It is intended to replace the Bali base version 1.0, which was first published in 2007. This database, which includes more than 2100 sequences overall and is the most complete database accessible, has 167 reference alignments. Bali basic version 2.0 contains 167 reference alignments. More than 2100 sequences have been added to Bali base version 2.0, and an astonishing number of 167 reference alignments have been included, making it the most complete database accessible. Bali basic version 2.0 includes eight reference sets, each of which is detailed in more detail further down this page. Bali basic version 2.0 has a total of eight reference sets. Bali basic version 2.0 contains a lot of improvements and bug fixes over the previous version. In Bali basic version 2.0, there are a total of eight reference sets available. Each reference has a unique set of sequences that are not present in any other reference and that are exclusive to that reference alone. This is a positive development since reference set 1 only has a small number of equidistance sequences, which is a rarity in genetics.^[16-18] The second reference set, in contrast to the first reference set, contains a sequence that is completely distinct from and unrelated to the first reference set, and it is referred to as the second reference set in this context. It has two sub-families that are diametrically opposed to one another in terms of their morphological and physiological features, which are discussed in more detail later in this section. Unlike the previous reference sets, reference set 4 includes sequences that contain both long and short terminal extension sequences, as well as sequences that do not have any terminal extension sequence at all. Several substantial internal insertions and deletions have been implemented, resulting in a considerable increase in the overall length of the text as a consequence of these modifications. Lastly additionally, the following sources provide additional test case problems, which include scenarios as well as other test case issues: An algorithmic process or

algorithm is given a Bali score, which is a statistic that may be used to assess the overall quality of the algorithm.^[6-8] There are two alignment sequences used: one that has been manually aligned (which is accessible in BALI Second Gen data set). It has one that has been generated by an existing method (which is not available in BALI Second Gen data set). The findings of the two alignment sequences are used to calculate Bali scores, which are then shown on a graph. It is possible for the Bali score to be any number between 0 and 1, with the greatest value being 1. The Bali score is determined by the current circumstances as the results assess the fitness of each habitat for turnaround time can be seen in Fig. 1. Unless the manually aligned file and our output file are the same size, the present study will not get a one-point grade on this assignment. A score of 0 is assigned to files that are completely distinct from one another, such as those that were manually aligned and those that were created by us. According on how close the BALI Second Gen data set manually aligning the files with the output file. There are one another, a number between 0 and 1 is returned as a result of the comparison between the files.

3. Result and discussion

For showing IBBOMSA's superiority over current multiple sequence alignment approaches that are based on evolutionary algorithms, the present study compares it to recently developed sequence alignment methods that are based on evolutionary algorithms, such as VDGA, GAPAM, and MOMSA. It is the purpose of this section to compare and contrast the performance of IBBOMSA with the performance of a number of other well-known aligners in the area of orthodontics. This article describes the C programming language that was used to create the programme IBBOMSA, as well as how the programme was ran on a computer running the Linux operating system.

3.1 Effect of improved operator in BBO

Researchers utilised a computer to scan through a multidimensional search space to create the BBO method, which was designed to monitor species' movement and emigration across different environments. At some point, every situation has the potential to offer a solution to some

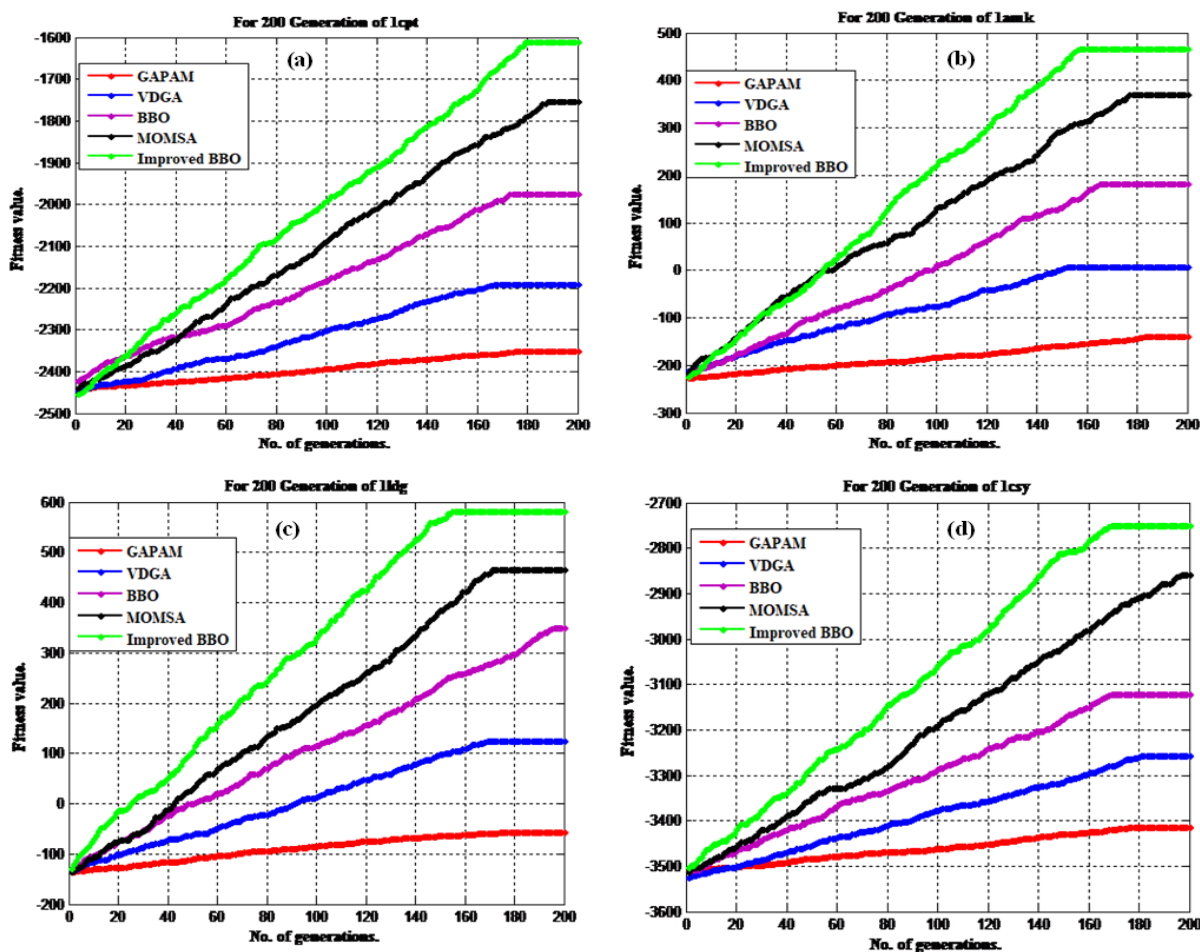


Fig. 1 Graphical comparison of fitness values for turnaround time a. Graphical comparison of turnaround time for GAPAM, VDGA, BBO, MOMSA, and improved PBO count of tasks are changing, and count of virtual machines are fixed. b. Graphical comparison of turnaround time for GAPAM, VDGA, BBO, MOMSA, and improved PBO count of tasks are fixed, and count of virtual machines are changing. C. Graphical comparison of for turnaround time for GAPAM, VDGA, BBO, MOMSA, and improved PBO count of tasks are varied, and count of virtual machines are fixed. d. Graphical comparison of for turnaround time for GAPAM, VDGA, BBO, MOMSA, and improved PBO when count of tasks are fixed, and count of virtual machines are changing.

problem. Features of BBO migration that were previously present in the superb solution are added as a new feature in the typical BBO migration in the poor solution, while remaining present in the good solution. These features are important to notice since it is likely to be present in a significant percentage of possible solutions. It is possible to improve both exploitation skills and search space while simultaneously reducing the amount of search space that may be exploited. If a subpar solution is implemented, a better migration path and updated features are made available to the user. In this scenario, it will be the responsibility of our suggested migration operator to provide you with the updated capabilities. It was decided to use a single scaling function to guarantee that the exploration (diversity) and exploitation capabilities were not jeopardised. However, in order to maintain the above-mentioned parameters, it is necessary to perform scaling operation in a suitably. The result of conventional BBO formula are identical when F is equal to zero. When the number of options in a search space falls because of a loss in

its variety, the ability to utilise the search space also drops because of this reduction in the number of possibilities.

Increasing the number of elements in F makes the search space more varied and increasing the number of ways to utilise the results of the search improves the capacity to use the results of the search as well. For these two criteria to be maintained at the same time, it has been established that the factor should be 0. An investigation into the effect of the suggested operator on the performance of the algorithms under discussion resulted in the creation of a series of five experiments, the results of which were examined. As well as many other algorithms, the GAPAM, VDGA, BBO, MOMSA, and Improved BBO were put to the test, as did a number of others. The results assess the fitness of each habitat in accordance with the fitness function provided in this document. A diagram of the Fitness Function may be seen in Fig. 2. According to the previous section, these tests were carried out

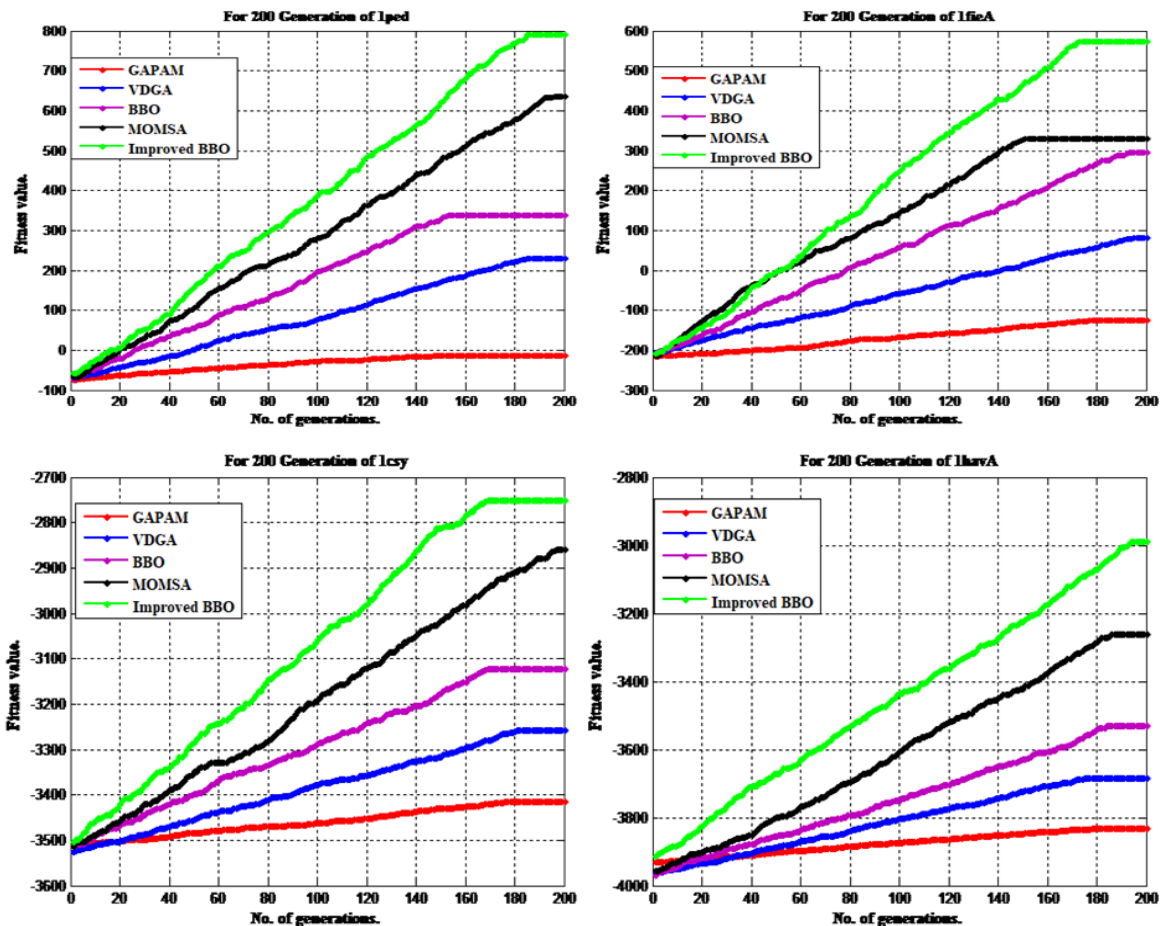


Fig. 2 Graphical comparison of fitness values for accuracy a. Graphical comparison of fitness value with respect to the number of eneration for GAPAM, VDGA, PBO, MOMSA, and improved PBO is shown when the task are changing, and count of virtual machines (computation power) are fixed. b. Graphical comparison of fitness value with respect to the number of eneration for GAPAM, VDGA, PBO, MOMSA, and improved PBO is shown when the tasks are fixed, and count of virtual machines (computation power) are changing. C. Graphical comparison of fitness value with respect to the number of eneration for GAPAM, VDGA, PBO, MOMSA, and improved PBO is shown when the tasks are varied, and number of virtual machines (computation power) are fixed. d. Graphical comparison of fitness value with respect to the number of eneration for GAPAM, VDGA, PBO, MOMSA, and improved PBO is shown when the tasks are fixed, and count of virtual machines (computation power) are changing.

using eight Bali base datasets (four from each of the reference sets 1 and 2), which were obtained from the Bali base database, as previously stated.

3.2 Experimental results and analysis

To assess the effectiveness of IBBOMSA, it is validated with to state-of-the-art current techniques such as VDGA, GAPAM, and MOMSA, which are regarded to be the best multiple sequence alignment algorithms available at the time of writing. Our proposed methodology was compared to current methods using a sample dataset collected from the MOMSA, which was selected properly. The subset of BALI Second Gen data set, precisely fifty-six test set has been considered in our experiment. The reference one to five which accumulated the 18, 23, 11, 3, and 2 respectively. The present study used the following references in their selection: Similarly, to what was previously stated in Section 3, the fitness function of multiple sequence alignment is determined. To calculate the fitness value of the matching multiple sequence alignment that has been recorded, it is required to do some calculations. Approximately 10 times, IBBOMSA is done, with the most significant findings being kept in a database. In accordance with Bali basis references 1, 2, 3, 4, and 5, the IBBOMSA, MOMSA, VDGA, and GAPAM tests were conducted as shown though several figures.

3.3 Discussion

Rubber Band (RB) has combines with genetic algorithm (GA) on a plate with many poles is analogous to biological input sequences. The GAs mimic biological evolution to solve issues. RB-GA population aligns protein sequences. The models alignment answers it as chromosomes to perform the alignment. These poles are sequence inputs and GA-based optimization improves chromosomes, yielding optimal MSA solution.^[21,22] Multiple sequence alignment is NP-complete problem, and it is one of the tricky problem. In [22], it is utilising ant colony optimization optimises multiple sequence alignment via ant colony optimization. Genetic algorithms can align to avoid local optima. The artificial bee colony (ABC) is recommended for MSA.^[23] The Bees influenced the ABC algorithm in such a way so that alignment can be perform optimally. Considering MSA's discreteness, a new ABC method is delivering the ABC algorithms which can beats GA and PSO in length and identical sequences. In [24], it has hybrid MSA combining PSO and variable neighbourhood descent (VND). The VND improves swarm leader solution quality and solves local optimum. Experiments on Bali-BASE have shown that the suggested method provides high-quality alignment. Bioinformatics uses multiple sequence alignment to achieve the highly precise and statistically significant multiple alignments which are challenging.^[25] It identifies optimal affine gap alignments in multiple sequencing data. The technique produces non-dominated trade-off alignments with opposing goals: affine gap penalty reduction and similarity and support maximisation. First three-goal effort.

Any sequence works. Accepts any similarity measure. Optimal alignments help decision-makers grasp trade-offs. The progressive before evolutionary has aligns affine-gap sequences more accurately than three methods. In study,^[26] it solves MSA with Flower Pollination Algorithm (FPA) and profiles. Meta-heuristic population-based method is able for MSA to detects genomic primers, conserved motifs, and phylogeny. Profile enhances MSA quality. It has Benchmarking its result on seven techniques and solution improved group from Balibase3.0 SP and scores over MSA problems. Hidden Markov Models (HMMs) aid NP-complete sequence alignment (MSA) to achieve the optimal solution.^[27] The PSO HMMs is for learning under MSA challenges and offered improves global search using evolution equations. Diversity-guided search improves algorithmic performance on two MSA benchmark datasets which are used to test. The Baum-Welch and other PSO algorithms misalign benchmark data sets. Bioinformatics predicts protein function using vectors contain protein sequences.^[28] The Protein datasets have complicated characteristics which can handles high-dimensional feature space. The GA and ACO enhance search simple classifiers minimise algorithm complexity. Unlike ACO and GA. It has enzymes are used. Compare maximum accuracy and lowest feature subset. Bioinformatics aligns sequences provides the gaps and improve sequence similarity.^[29] It has evolutionary method aligns DNA to optimizing uses differential evolutions. Opposition-based chromosome representation has able to provide the fitness is scored. Score accepted mutations. BALIBASE 1.0 and verify the superiority of results. Multiple Molecular sequence structure prediction using amino acid sequence has been provided in.^[30,31] In silico challenges include high conformational space dimension and rough energy surface. Energy surfaces have several minima which unchanneled surfaces feature deep non-native minima. Decoy conformations help inclusive algorithms to achieve the conformational diversity reflects nature.

4. Conclusions

As a result of this research, the present study have developed an enhanced BBO approach which application is to tackle the MSA problem efficiently. The creation of a new migration operator was necessary in order to ensure that exploration and exploitation continued to function smoothly. However, while using the scaling function, it is important to use extreme caution. With the previous BBO algorithm as a baseline, it looked at the new approach to determine which one was more effective in terms of performance. The present study provides the example to demonstrates our approach and its superiority over the existing work. In addition, it seems competitive with the existing work.

Our current method was tested against a wide variety of parameters with the existing datasets such as BALI second generation to provide the significance achievement of our work. The MOMSA is one of the tests which has been

considered for various testing were available at the time of this work completion. Consequently, in comparison to other techniques, the Bali score associated with this solution was employed because the Bali score was being used as their primary assessment tool for the quality and accuracy of the MSA at the time of study. Following the completion of the testing, it was revealed that the suggested BBO outperformed the competition in virtually all of the test scenarios. Even though the recommended BBO technique did not always offer the greatest solution for specific test conditions, it was consistently near to being the best option. According to the study's results, the enhanced migration operator included in the suggested method, which contributes to preserving the variety of the search space, beat all other approaches in terms of performance. On the basis of the outcomes of the experimental investigation, it can be concluded that the suggested approach is capable of successfully overcoming the challenges associated with multiple sequence alignment problems.

Acknowledgements

The work is independent done by authors and involved parties acknowledge in reference.

Conflict of Interest

The authors declare no conflict of interest.

Supporting Information

Not applicable.

References

- [1] D. K. Sah, T. Amgoth, Parametric survey on cross-layer designs for wireless sensor networks, *Computer Science Review*, 2018, **27**, 112-134, doi: 10.1016/j.cosrev.2017.12.002.
- [2] D. K. Sah, C. Shivalingagowda, D. P. Kumar, Optimization problems in wireless sensors networks, *Soft Computing in Wireless Sensor Networks*, Chapman and Hall/CRC, 2018, 29-50, doi: 10.1201/9780429438639-2.
- [3] B. Chowdhury, A review on multiple sequence alignment from the perspective of genetic algorithm, *Genomics*, 2017, **109**, 419-431, doi: 10.1016/j.ygeno.2017.06.007.
- [4] D. K. Sah, T. N. Nguyen, K. Cengiz, B. Dumba, V. Kumar, Load-balance scheduling for intelligent sensors deployment in industrial Internet of Things, *Cluster Computing*, 2022, **25**, 1715-1727, doi: 10.1007/s10586-021-03316-1.
- [5] N. Agrawal, S. Kumar, D. K. Sah, Prevention of DDoS attack through harmonic homogeneity difference mechanism on traffic flow, *2018 4th International Conference on Recent Advances in Information Technology (RAIT)*, IEEE, 2018, 1-6, doi: 10.1109/RAIT.2018.8389081.
- [6] E. Bernal, O. Castillo, J. Soria, F. Valdez, Optimization of fuzzy controller using galactic swarm optimization with type-2 fuzzy dynamic parameter adjustment, *Axioms*, 2019, **8**, 26, doi: 10.3390/axioms8010026.
- [7] G. Han, J. Wu, H. Wang, M. Guizani, J. A. Ansere, W. Zhang, A multicharger cooperative energy provision algorithm based on density clustering in the industrial Internet of Things, *IEEE Internet of Things Journal*, 2019, **6**, 9165-9174, doi: 10.1109/JIOT.2019.2928557.
- [8] G. F. Gomes, S. S. Cunha, A. C. Ancelotti, A sunflower optimization (SFO) algorithm applied to damage identification on laminated composite plates, *Engineering with Computers*, 2019, **35**, 619-626, doi: 10.1007/s00366-018-0620-8.
- [9] J. Wang, K. Wang, J. Niu, W. Liu, A K-medoids based clustering algorithm for wireless sensor networks, *2018 International Workshop on Advanced Image Technology (IWAIT)*. IEEE, 2018, 1-4, doi: 10.1109/IWAIT.2018.8369769.
- [10] J. Zhang, Y. Zhou, Q. Luo, Nature-inspired approach: a wind-driven water wave optimization algorithm, *Applied Intelligence*, 2019, **49**, 233-252, doi: 10.1007/s10489-018-1265-4.
- [11] J. Lin, Accurate prediction of potential druggable proteins based on genetic algorithm and Bagging-SVM ensemble classifier, *Artificial Intelligence in Medicine*, 2019, **98**, 35-47, doi: 10.1016/j.artmed.2019.07.005.
- [12] L. Cheng, A staged adaptive firefly algorithm for UAV charging planning in wireless sensor networks, *Computer Communications*, 2020, **161**, 132-141, doi: 10.1016/j.comcom.2020.07.019.
- [13] N. Gharaei, Y. D. Al-Otaibi, S. A. Butt, S. J. Malebary, S. Rahim, G. Sahar, Energy-efficient tour optimization of wireless mobile chargers for rechargeable sensor networks, *IEEE Systems Journal*, 2021, **15**, 27-36, doi: 10.1109/JSYST.2020.2968968.
- [14] V. Muthiah-Nakarajan, Galactic Swarm Optimization: a new global optimization metaheuristic inspired by galactic motion, *Applied Soft Computing*, 2016, **38**, 771-787, doi: 10.1016/j.asoc.2015.10.034.
- [15] J. Lin, H. Chen, S. Li, Y. Liu, X. Li, B. Yu, Accurate prediction of potential druggable proteins based on genetic algorithm and Bagging-SVM ensemble classifier, *Artificial Intelligence in Medicine*, 2019, **98**, 35-47, doi: 10.1016/j.artmed.2019.07.005.
- [16] A. Yang, W. Zhang, J. Wang, K. Yang, Y. Han, L. Zhang, Review on the application of machine learning algorithms in the sequence data mining of DNA, *Frontiers in Bioengineering and Biotechnology*, 2020, **8**, 1032, doi: 10.3389/fbioe.2020.01032.
- [17] B. Yu, C. Chen, X. Wang, Z. Yu, A. Ma, B. Liu, Prediction of protein-protein interactions based on elastic net and deep forest, *Expert Systems with Applications*, 2021, **176**, 114876, doi: 10.1016/j.eswa.2021.114876.
- [18] F. A. Zainuddin, Md. Fahmi Abd Samad and Durian Tunggal, A review of crossover methods and problem representation of genetic algorithm in recent engineering applications, *International Journal of Advanced Science and Technology*, 2000, **29**, 759-769.
- [19] Y. Zhang, Q. Zhang, J. Zhou, Q. Zou, A survey on the algorithm and development of multiple sequence alignment, *Briefings in Bioinformatics*, 2022, **23**, doi: 10.1093/bib/bbac069.
- [20] D. K. Sah, H. Ahmad, and A. Uniyal, Reliable ilp approach of max-rwa problem for translucent optical network, *2016*

Thirteenth International Conference on Wireless and Optical Communications Networks (WOCN), IEEE, 2016, doi: 10.1109/WOCN.2016.7759879.

[21] D. K. Sah, T. N. Nguyen, M. Kandulna, K. Cengiz, T. Amgoth, 3D localization and error minimization in underwater sensor networks, *ACM transactions on sensor networks (TOSN)*, 2022, **18**, 1-25, doi: 10.1145/3460435.

[22] J. Taheri and A. Y. Zomaya, RBT-GA: a novel metaheuristic for solving the multiple sequence alignment problem, *Bmc Genomics*, 2009, **10**, 1-11, doi: 10.1186/1471-2164-10-S1-S10.

[23] Z.-J. Lee, S. -F. Su, C. -C. Chuang, K. -H. Liu, Genetic algorithm with ant colony optimization (GA-ACO) for multiple sequence alignment, *Applied Soft Computing*, 2008, **8**, 55-78, doi: 10.1016/j.asoc.2006.10.012.

[24] X. Lei, J. Sun, X. Xu, L. Guo, Artificial bee colony algorithm for solving multiple sequence alignment, *2010 IEEE Fifth International Conference on Bio-Inspired Computing: Theories and Applications (BIC-TA)*, IEEE, 2010, 337-342, doi: 10.1109/BICTA.2010.5645304.

[25] L. Chaabane, An enhanced hybrid model for solving multiple sequence alignment problem, *International Conference on Big Data and Smart Digital Environment*, 2019, 86-91, doi: 10.1007/978-3-030-12048-1_11.

[26] M. Kaya, A. Sarhan, R. Alhajj, Multiple sequence alignment with affine gap by using multi-objective genetic algorithm, *Computer Methods and Programs in Biomedicine*, 2014, **114**, 38-49, doi: 10.1016/j.cmpb.2014.01.013.

[27] A. M. Hussein, R. Abdullah, N. Abdul Rashid, Flower pollination algorithm with profile technique for multiple sequence alignment, *2019 IEEE Jordan International Joint Conference on Electrical Engineering and Information Technology (JEEIT)*, IEEE, 2019, 571-576, doi: 10.1109/JEEIT.2019.8717513.

[28] J. Sun, V. Palade, X. Wu, W. Fang, Multiple sequence alignment with hidden Markov models learned by random drift particle swarm optimization, *IEEE/ACM Transactions on Computational Biology and Bioinformatics*, 2014, **11**, 243-257, doi: 10.1109/TCBB.2013.148.

[29] S. Nemat, A novel ACO-GA hybrid algorithm for feature selection in protein function prediction, *Expert Systems with Applications*, 2009, **36**, 12086-12094, doi: 10.1016/j.eswa.2009.04.023.

[30] S. Pramanik, S. K. Setua, An opposition based differential evolution to solve multiple sequence alignment, *Computational Intelligence, Communications, and Business Analytics: First International Conference, CICBA 2017*, 2017, 440-450, doi: 10.1007/978-981-10-6430-2_34.

[31] S. Saleh, B. Olson, A. Shehu, A population-based evolutionary search approach to the multiple minima problem in de novo protein structure prediction, *BMC Structural Biology*, 2013, **13**, 1-19, doi: 10.1186/1472-6807-13-S1-S4.

Author Information



Anjali Goswami is presently working as an Associate Professor in department of Basic Science, College of Science and Theoretical Studies, Saudi Electronic University, Riyadh, Saudi Arabia. She received her Ph.D. in year 2007 from India. Previously she worked as an Associate Professor, HOD at the Department of Mathematics, Jagannath Institute of Engineering and Technology, Jaipur, Rajasthan, India. She has more than 20-years of teaching and Research experience. She has published two books and more than 30 research papers in various referred reputed journals.



Dr. Kamlesh Kumar Dubey is presently working as a Professor and Head in Mathematics in the Department of Applied Sciences and Humanities at Invertis University Bareilly India. He got his Undergraduate and Postgraduate degree from University of Allahabad, Allahabad India and Ph.D. from D. D. U. Gorakhpur University, Gorakhpur in 2006 in the field of Differential geometry. His area of interest is Differential geometry, Mathematical Modelling and Computational Mathematics. He has numerous no of research papers in these fields.

Publisher's Note: Engineered Science Publisher remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.