Multi-objective formulation of MSA for phylogeny estimation

(Do application-aware measures guide towards better phylogenetic tree?)

Muhammad Ali Nayeem, Md. Shamsuzzoha Bayzid, Atif Hasan Rahman, Rifat Shahriyar, and M. Sohel Rahman, *Senior Member, IEEE*

***Abstract*—Multiple sequence alignment (MSA) is a basic step in many analyses in computational biology, including predicting the structure and function of proteins, orthology prediction and estimating phylogenies. The objective of MSA is to infer the homology among the sequences of chosen species. Commonly, the MSAs are inferred by optimizing a single function or objective. The alignments estimated under one criterion may be different from the alignments generated by other criteria, inferring discordant homologies and thus leading to different evolutionary histories relating the sequences. In recent past, researchers have advocated for the multi-objective formulation of MSA, to address this issue, where multiple conflicting objective functions are being optimized simultaneously to generate a set of alignments. However, no theoretical or empirical justification with respect to a real-life application has been shown for a particular multi-objective formulation. In this study, we investigate the impact of multi-objective formulation in the context of phylogenetic tree estimation. Employing multi-objective metaheuristics, we demonstrate that trees estimated on the alignments generated by multi-objective formulation are substantially better than the trees estimated on the alignments generated by the state-of-the-art MSA tools, including PASTA, MUSCLE, CLUSTAL, MAFFT, etc. We also demonstrate that highly accurate alignments with respect to popular measures like sum-of-pair (SP) score and total-column (TC) score do not necessarily lead to highly accurate phylogenetic trees. Thus, in essence, we ask the question of whether an application-aware (in this case phylogeny-aware) metric can guide us in choosing appropriate multi-objective formulations that can result in better phylogeny estimation. And we answer the question affirmatively through carefully designed extensive empirical study. As a by-product, we also suggest a methodology for primary selection of a set of objective functions for a multi-objective formulation based on the association with the resulting phylogenetic tree. In the sequel, we end up proposing a set of nonparametric objective functions (i.e., {SimG, SimNG}) that should be used by MSA methods if the goal is to produce high-quality phylogenetic trees.**

***Index Terms*—Bioinformatics, Computational Biology, Evolutionary algorithm, Metaheuristics, Multiple sequence alignment, Multi-objective optimization, Phylogenetic tree.**

I. Introduction

In biological research, multiple sequence alignment (MSA) is a useful and/or essential task in various applications such as phylogeny estimation, prediction of the structure and function of a RNA or protein, identification of functionally important sites, orthologous gene identification etc. The MSA task seeks to arrange more than two biological sequences to infer homology, based on certain criteria such as evolutionary history, 3D structure etc. The output is a matrix in which the input sequences are the rows and each column (i.e., site) has letters (i.e., nucleotides or amino acids) which are homologous which means all those letters descend from the same letter of a common ancestor). The aligned sequences reflect historical substitution, insertion and deletion of genetic materials which are represented as gaps. Accurately recovering these properties through MSA is necessary to accomplish a biological objective such as inferring the evolutionary history relating the sequences known as phylogenetic trees. While computing MSAs, various computational methods and criteria are used to make hypotheses about homology. But the goal of MSA is entirely biological. Figure 1 illustrates this problem using a hypothetical example where four protein sequences of varying lengths need to be simultaneously aligned by inserting "appropriate" gaps to identify homology. In this research, we limit our focus on MSA in the context of phylogeny. Phylogeny estimation from molecular sequences generally operates as a two-phase approach. At first the given sequences are aligned using an MSA method, and then a tree is estimated from the resultant alignment. The quality of inferred trees heavily depends on the quality of the corresponding alignment. There is a large body of literature in the biological domain about the relationships between multiple sequence alignments and phylogenetic trees which laid the background for this study. For example, we find several studies [1]–[6] analyzing the effects of alignment errors and uncertainty on the accuracy of phylogenetic tree reconstruction. Therefore, it is important to select an MSA tool that is the "most suitable" in the phylogenetic context.

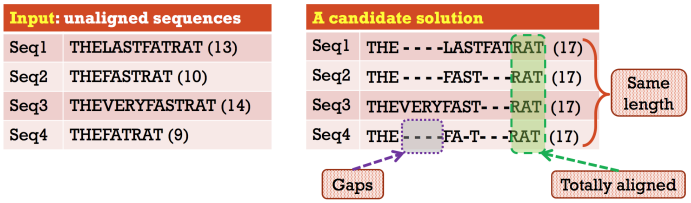


Fig. 1: A hypothetical example of an MSA problem.

In this study, we aim to design a multi-objective formulation of MSA that is more effective in phylogeny estimation. Previously several researchers ([7], [8]) made an

All the authors are with the Department of Computer Science and Engineering, Bangladesh University of Engineering and Technology, Dhaka 1205, Bangladesh. Corresponding author: M. Sohel Rahman (e-mail: msrahman@cse.buet.ac.bd).

attempt to improve the phylogeny estimation focusing on MSA computation. While they also agree with our core idea that the nature of MSA computation may influence the outputs (in a domain specific manner), they do not focus on any application-aware multi-objective formulation. Our motivation for a multi-objective formulation comes from the fact that the alignment estimated under one objective may be different to the alignments generated under other objectives, inferring discordant homologies and thus leading to different and often conflicting evolutionary histories relating the sequences under consideration. Multi-objective formulations can address this issue by optimizing multiple conflicting objectives simultaneously to generate a set of alignments. However, we are faced with the challenge of using appropriate measures/metrics to choose from among a number of objective sets to optimize. So, we ask the natural question whether the popular general purpose measures to judge the alignment quality can truly reflect the quality in the context of a particular application domain, i.e., phylogeny estimation in our case. While this question has received some shallow discussion in several studies ([9]–[11]), to the best of our knowledge no systematic investigation has been reported in the literature to this end. Therefore, in essence, we systematically investigate whether an application-aware metric can guide us better in choosing appropriate multi-objective formulation or tools capable of generating alignments that can produce better phylogenetic trees.

There are numerous tools available in the literature to compute MSAs. We can broadly divide them into three groups: progressive techniques, consistency based techniques and iterative techniques. This division is not exclusive as many tools also use a combination of these techniques. Progressive technique is the foundation of many MSA tools such as, Clustal Ω [12], PRANK [13], Kalign [14], FSA [15], RetAlign [16] etc. They compute the alignment using a guide tree by aligning pairs of sequences in a "bottom-up" manner. Among them, FSA employs an explicit statistical model to generate the alignments. It is the only method that gives an estimation of uncertainty for every column and character of the alignment. Moreover, it utilizes machine-learning techniques to estimate gap and substitution parameters at runtime for each input data.

The consistency based techniques first construct a database of local and global pairwise alignments to facilitate generating an overall accurate alignment. The representatives of this category are T-Coffee [17], ProbCons [18], MSAProb [19], ProbAlign [20] etc. On the other hand, the iterative techniques were designed to achieve reliable alignments. These techniques try to fix the effect of mistakes made during the initial phases by repeating some crucial steps until some criteria are met. We find several examples of such techniques, such as, MAFFT [21], MUSCLE [22], MUMMALS [23], ProbConsetc. In this category, we also see some "meta-methods" such as, SATé [11] and PASTA [10], which co-estimate alignment and tree using other methods. These tools achieve scalability by employing the divide-and-conquer principle and are being used widely in practice.

The performance of an MSA tool is usually evaluated by comparing its output alignment with the reference alignment (provided with the dataset as the ground truth) in terms of several measures. To this end, the most popular measures are perhaps sum-of-pair (SP) score and total-column (TC) score. SP score is the fraction of the homologies (i.e., pairs of aligned characters) in the reference alignments recovered in the estimated alignment. Similarly, TC score is the fraction of the actual aligned columns that appear in the estimated alignment.

In this post-genomic era, the MSA datasets are posing new challenges to the researchers. Usually, an MSA method is provided with a default parameter configuration for aligning any problem instance with satisfactory accuracy. But we know that these default values can not guarantee the best output throughout all kinds of datasets [24]. For instance, there is a parameter in ProbCons called the number of iterative refinement passes. Although it can vary between 0 to 1000 the default value is set to 100. We can achieve better results by tuning the parameter values which is not a straightforward task. A systematic approach, namely, parameter advertising [25], helps to choose the best parameter setting of an MSA method for each input data. Moreover, despite rigorous parameter tuning, no method can consistently outperform other methods for all datasets.

Therefore, we see the emergence of novel approaches that combine different alignment tools [26]. One such approach is metaheuristics where alignments generated from different tools are exploited to produce improved alignments without vesting any effort in parameter tuning. The success of such a metaheuristic approach depends on the selection of proper objective function that can help to select better solutions from among the alternatives and thereby guide the search process towards optimal solutions of MSA. As any single objective function alone can not be effective to tackle different challenges, it is wise to simultaneously optimize multiple objective functions. This will produce a set of competing solutions as the final output, which can be expected to contain our desired solution(s). Thus the formulation of MSA as a multi-objective optimization problem turns out to be appealing.

During the last decade, we find several studies ([27]–[32]) with multi-objective formulation for MSA have been published – proposing two to four objective functions to capture and quantify different aspects of an alignment. Among them, probably the most popular is the sum-of-pairs score and its weighted variants, where pairwise score is calculated for each pair of aligned sequences using a substitution matrix. This matrix should reflect the characteristics of the data at hand. Although we know that the same character across all rows of a column does not necessarily indicate homology, the count of such columns in an alignment is seen as a maximization objective known as totally conserved columns. Next, we find attempts to minimize total number of gaps to maintain the compactness of an alignment. Then there are different types of gap penalties that penalize each sequence for introducing gaps. Also we find two other objective functions, Entropy and Similarity, that compute column-wise scores and then sum those together. Both of them try to express the homogeneity of characters in a column using two different ways. Contrary to the performance/quality measures mentioned earlier (such as

SP score, TC score), we are not allowed to use the reference alignment while calculating these objective functions.

We notice several issues in the works advocating multi-objective formulation of MSA (in the context of different applications where the MSA will be used). First of all, in these works, there is a lack of sound theoretical or empirical justification for the choice of a particular objective function to be optimized. Secondly, we also notice the absence of a sound rationale/justification behind the two most popular performance metrics, namely, sum-of-pair score and total column score. On the contrary, it seems only natural that performance score should reflect the actual purpose of MSA. For example, if the goal is to estimate a phylogenetic tree, the performance metric to be used for evaluation should be able to accurately measure the quality and usefulness of the constructed tree. Notably, another issue, specific to the domain of phylogeny estimation, is the use of relatively smaller (number of taxa below 50) datasets in experiments.

In this article, we attempt to demonstrate the effectiveness of multi-objective MSA by addressing the above mentioned limitations in the context of its intended application domain (i.e., phylogeny estimation). To make a fair comparison with nine state-of-the-art MSA tools, we conduct comprehensive experimentations on both simulated and biological datasets using tree as well as alignment quality measures. Our study represents the only known work on devising a phylogeny-aware multi-objective formulation for MSA. In particular, this article makes the following key contributions:

- To the best of our knowledge, this is the first attempt to investigate whether a domain specific measure (as opposed to generic alignment measures) can guide us better in choosing an appropriate multi-objective formulation or tool for MSA when the goal is to infer phylogeny. Notably, although there exist some prior works that proposed different approaches in MSA computation (such as averaging MSA [8]) to improve phylogeny reconstruction, our novelty lies in providing an application-aware multi-objective formulation for MSA.
- We suggested a methodology based on multiple linear regression to judge the potential efficacy of a multi-objective formulation of MSA. Then, based on this methodology, we proposed two multi-objective formulations that had the potential to yield better phylogenetic trees.
- Finally, we demonstrated that the multi-objective formulations can consistently yield better phylogenetic trees than several state-of-the-art MSA tools. Following our methodology, we identified {SimG, SimNG} to be the best set of objective functions for computing MSAs with an aim to infer phylogeny, considering its overall accuracy, runtime and nonparametric nature. And, interestingly we found that popular alignment quality measures do not necessarily lead to highly accurate phylogenetic trees.

II. Methods

We begin this section with an overview of our experimental design. Then we introduce the objective functions that we used to compute MSA. Finally, we discuss the multi-objective metaheuristics applied to optimize those objective functions as well as the state-of-the-art tools that we utilized in this study. Because we deal with a number of objective functions, more than 10 state-of-the-art MSA tools and more than 30 instances of MSA problem, a reader is exposed to an over-preponderance of acronyms and short-cut notations. Therefore, for the sake of ease in exposition and understanding, we alphabetically list all the acronyms used in this study with their usage in Table S1 of the supplementary file.

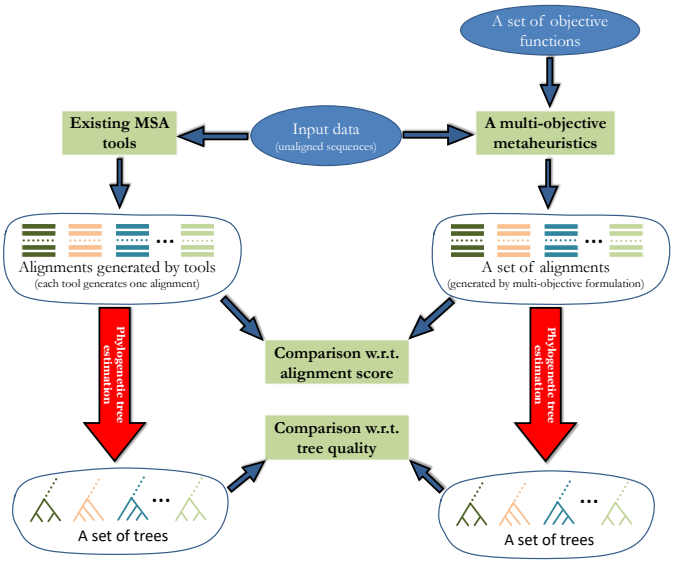


Fig. 2: Our methodology for finding the impact of a multi-objective formulation (i.e., a set of objective functions) of MSA on phylogenetic tree estimation.

A. Experimental design

Our experimental methodology is briefly described below (please see also Figure 2):

- Step 1: Following a systematic approach involving multiple linear regression applied on a simulated dataset, we first make an attempt to identify and choose two multi-objective formulations that turn out to be potentially more effective in the context of phylogeny estimation (discussed in Section S6 of the supplementary file).
- Step 2: We run a popular and effective multi-objective metaheuristics on biological datasets to optimize each set of objective functions selected in Step 1. Each run of the metaheuristics on each dataset gives us a set of alignments as the final output.
- Step 3: We also run nine state-of-the-art MSA tools (please see Table III) to generate alignments on all these datasets.
- Step 4: We evaluate the quality of each generated alignment with respect to the reference alignment using two popular measures, namely, SP score and TC score (discussed in Section S3 of the supplementary file).
- Step 5: For each of the generated alignments, we infer maximum likelihood (ML) phylogenetic tree (discussed

in Section S4 of the supplementary file). Then we measure the quality of each inferred tree with respect to the reference tree (true tree) using a commonly used measure in the literature called false negative (FN) rate [33] (discussed in Section S5 of the supplementary file).

- Step 6: Finally we compare the alignments and the corresponding ML trees generated by the multi-objective optimization with the ones generated by the state-of-the-art tools.

B. Objective functions

Most real-world optimization problems naturally work towards achieving several objectives. Some of these objectives are conflicting to each other. However, these problems can be transformed into single-objective ones using various simplifying techniques to avoid complexities [34]. On the contrary, a multi-objective formulation defines the problem using a set of objective functions and subsequently specialized methods can be applied to optimize all the objectives simultaneously. In this study, we have selected the following three multi-objective formulations of MSA from the literature based on their simplicity as well as performance as reported in the literature.

- {SOP, TC}: Maximize the sum of pairs (SOP) and the number of totally aligned columns (TC) [27].
- {Gap, SOP}: Maximize the sum of pairs (SOP) and minimize the number of gaps (Gap) [30].
- {wSOP, TC}: Maximize the weighted sum of pairs with affine gap penalties (wSOP) and the number totally aligned columns (TC) [31].

We describe these objective functions along with several existing ones in Section S1 of the supplementary file. Now, we propose four new objective functions that quantify different aspects of an MSA. Unlike the existing objective functions in the literature, we avoid combining multiple aspects of an MSA into a single objective. We introduce them as follows:

- **Minimize entropy (Entropy)**: We modify the usual definition by considering only non-gap column for the calculation of entropy.
- **Maximize similarity based on gap containing columns (SimG)**: Here we calculate similarity only for those columns that contain at least one gap.
- **Maximize similarity based on non-gap columns (SimNG)**: We consider only non-gap column while measuring similarity.
- **Maximize concentration of gaps (GapCon)**: We find that a widely used objective function, namely, affine gap penalty [35], combines two aspects of an aligned sequence, number of gaps and concentration of gaps, into a single one using weighted sum. We need to tune the weight values based on the dataset. To avoid this tuning we decide to decouple the two components. We have already considered the number of gaps as an objective function. Now we define the concentration of gaps as an independent objective which as calculated as follows. For each sequence, we count the number of consecutive gaps and take the mean of these counts. Finally, we average the resultant means for all sequences.

In this study, we used the terms shown in Table I to refer to these objective functions.

TABLE I: Terms used to denote the objective functions.

#	Objective function	Term
1	Maximize no. of totally aligned columns	TC
2	Minimize no. of gaps	Gap
3	Maximize sum of pairs	SOP
4	Maximize weighted sum of pairs with affine gap penalties	wSOP
5	Minimize entropy	Entropy
6	Maximize similarity based on gap containing columns	SimG
7	Maximize similarity based on non-gap columns	SimNG
8	Maximize concentration of gaps	GapCon

We need a substitution matrix to calculate SOP and wSOP. The values of this substitution matrix depend on the trait of a particular dataset. In this study, we used NUC4.4 (supplied by NCBI at ftp://ftp.ncbi.nih.gov/blast/matrices/NUC.4.4) for nucleotide sequences and BLOSUM62 [36] for protein sequences. On the contrary, the four objective functions that we proposed are nonparametric which are independent of the dataset.

C. Multi-objective metaheuristics

To simultaneously optimize multiple objective functions, we ran two popular multi-objective metaheuristics: NSGA-II [37] and NSGA-III [38]. They belong to the class of multi-objective evolutionary algorithms. They start from a set of candidate solutions (termed as population) and then uses mechanisms inspired by biological evolution (such as mutation, crossover, selection, etc.) to evolve the population towards the optimal solutions. Unlike single-objective optimization methods, they output a set of solutions (i.e. members of the final population) which represents the best possible compromise of all objectives under consideration. Two studies ([28], [39]) demonstrated the strength of NSGA-II for solving MSA. NSGA-II works best when the number of objectives is upto three while NSGA-III is specially designed for handling more than three objectives. Hence, we applied these algorithms according to Table II. We discuss these methods along with their vital components and parameters in Section S2 of the supplementary file. We implemented them using jMetalMSA [40] which is a Java metaheuristic framework for MSA. Our implementation is publicly available at https://github.com/ali-nayeem/MSA.

TABLE II: Our selected algorithms and corresponding objective set.

Algorithm	Objective set
NSGA-II	{Gap, SOP} {SOP, TC} {wSOP, TC} {SimG, SimNG}
NSGA-III	{Gap, SOP, wSOP, TC} {Entropy, TC, Gap, SimG, SimNG, GapCon}

D. State-of-the-art MSA tools

We used the alignments generated by nine representative state-of-the-art MSA tools (shown in Table III) to compare with our approach. We run each of them with its default parameter configuration. Moreover, we initialize the multi-objective metaheuristics with a set of alignments generated by randomly mixing and modifying those nine alignments. Notably, this approach, known as the seeded initial population generation, is quite common in the metaheuristics literature specially for multi-objective optimization.

TABLE III: List of state-of-the-art MSA tools that we used in this study.

For nucleotide sequences		For protein sequences	
Tool	Version	Tool	Version
FSA [15]	1.15.9	FSA	1.15.9
PASTA [10]	1.7.8	PASTA	1.7.8
T-Coffee [17]	11.00	T-Coffee	11.00
MAFFT [21]	7.31	MAFFT	7.245
Clustal W [41]	2.1	Clustal W	2.1
Clustal Ω [12]	1.2.4	RetAlign [16]	1.0
MUSCLE [22]	3.8.31	MUSCLE	3.8.31
PRANK [13]	0.170427	ProbCons [18]	1.12
Kalign [14]	2.03	Kalign	2.04

III. Results

We conducted extensive experiments with both simulated and biological datasets. We begin by carefully and systematically selecting two multi-objective formulations which are potentially useful for phylogenetic tree estimation employing NSGA-III and multiple linear regression. Next, we generate alignments through running NSGA-II as well as nine state-of-the-art MSA tools. Then we compare those alignments with respect to both generic and domain specific quality measures. In this section, we discuss our obtained results after introducing our chosen datasets. In what follows, unless otherwise specified, when we discuss the (best) results of a tool, we mean one of the above-mentioned nine tools.

A. Datasets

We studied a simulated dataset (100-taxon simulated dataset [11]) as well as two biological datasets (biological rRNA datasets [11] and BAliBASE 3.0 benchmark [42]). As the simulated dataset comes with the true phylogenetic tree, we use this dataset to examine whether a multi-objective formulation of MSA is potentially application-aware and in the sequel, we select two such formulations somewhat similar to the training phase of a machine learning approach. Afterwards, we validate the effectiveness of the selected formulations against the state-of-the-art MSA tools based on biological datasets.

From the 100-taxon simulated dataset, we randomly selected five replicates. And among the biological datasets, we chose two challenging ribosomal RNA datasets along with 27 random instances of the widely used BAliBASE 3.0 benchmark. Section S7.1 of the supplementary file provides a detailed description of these datasets.

B. Selection of appropriate multi-objective formulations

As has been mentioned above, we have used 100-taxon simulated dataset to select one or more multi-objective formulations that have the potential to be "application-aware". We first conduct extensive experiments to choose a formulation (i.e., a set of objective functions) of MSA from among the existing popular ones from the literature (Section III-B1); subsequently, we also suggest a new promising formulation (Section III-B2).

1) Selection from among the existing formulations: To reduce the computational effort, we pre-select three multi-objective formulations of MSA and limit our investigation thereon. Thus we choose one of the formulations from among {Gap, SOP} [30], {SOP, TC} [27] and {wSOP, TC} [31] (please see Section II Table I). We experiment with five randomly selected replicates (R0, R4, R9, R14, R19) and then judge based on two criteria: firstly, we used multiple linear regression analysis to examine the association between individual objective function and FN rate; secondly, we assess the alignments generated through the optimization of each set of objective functions in terms of resultant ML trees.

We need to consider the relationship between each pair of objective functions to properly interpret the result of multiple linear regression. We perform this by running an appropriate multi-objective metaheuristic (i.e., NSGA-III [38]) for 25 times which simultaneously optimizes all the objective functions (i.e., {Gap, SOP, wSOP, TC}) and thus we obtain a large collection of diverse alignments be merging the sets of solutions output by each run. A visualization of the interrelations among the objective values of those solutions is presented in Figure S2 of the supplementary file. From these experiments, we have the following two key observations. (a)

1) In all the cases, SOP is totally correlated with wSOP. So we do not need to optimize both of them. Moreover, this high correlation creates a serious problem in multiple regression analysis called multicollinearity [43]. Therefore, we should not keep these two objective functions together in our regression analysis. Also, it is redundant to consider both of them in the multi-objective formulation.
2) SOP is clearly in conflict with Gap across all the replicates. Therefore, if we optimize them simultaneously, we can generate many diverse solutions which represent the compromise between these two objective functions [34]. This diverse collection is likely to contain the desired alignment for any kind of dataset.

As the objective functions are inter-related, we need to measure the degree of association between an objective and FN rate while holding the remaining objectives constant to avoid getting any spurious result [43]. Therefore, we perform multiple linear regression by employing the following model:

$$\begin{aligned}\text{FN rate} = \beta_0 + \beta_1 \times \text{TC} + \beta_2 \times \text{Gap} + \\ \beta_3 \times \text{SOP (or wSOP)} + \epsilon\end{aligned} \tag{1}$$

Each coefficient (β_1, β_2 and β_3) represents the expected change in the FN rate per unit change in the corresponding objective function when all the remaining objective functions

are held constant. For this reason, they (β_i) are called partial regression coefficients. ϵ is the random error component which is assumed to follow a Gaussian distribution with mean zero and some fixed standard deviation. We fit this model to the solutions generated by optimizing the set {Gap, SOP, wSOP, TC}. For each of those solutions, we estimate the ML tree and evaluate its quality in terms of FN rate. We estimate these coefficients using the least-squares method (an illustration is presented in Figure S3 of the supplementary file). We apply t-test on individual regression coefficient (i.e., slope) β_i (with null hypothesis $\beta_i = 0$) to test the significance of that association. We can note the following two interesting points from these results. (a)

1) In the majority of the cases (R0, R4 and R14), Gap, SOP and wSOP exhibit a good degree of association with FN rate (i.e positive slope) with high confidence (p-value close to 0) compared to other objective functions. So, we can expect them to be good optimization objectives for MSA.
2) For replicate R4 and R19, none of the objective exhibit good association. This shows that an objective function might not perform well across all problem instances.

Now we measure the strength of each objective set based on the FN rate achieved by the members of the generated solution set. To accomplish this, for each set of objective functions, we run a suitable multi-objective metaheuristics (NSGA-II [37]) for 20 times following the standard practice of operations research (OR) literature (due to the stochastic nature of metaheuristics). Each run generates a set of solutions that represents the trade-offs in satisfying all objectives. Afterwards, we inferred the ML tree for each of the generated alignment. We collected the best FN rates from each of the 20 solution sets and examine the distribution of these FN rates (a visualization of these distributions using boxplots is presented in Figure S4 of the supplementary file). Here we have the following key observations:

- For most of the cases, the combined set {TC, Gap, SOP, wSOP} achieves better results than the other sets. This indicates that adding suitable objective functions increase the chance of achieving the best FN rate. However, this can increase the overall complexity of the multi-objective metaheuristic. So in this study, we keep the size of the objective set as small as possible.
- Among our three pre-selected objective sets, {Gap, SOP} achieves relatively lower FN rates. This is consistent with the regression results discussed earlier.
- Both {TC, Gap, SOP, wSOP} and {Gap, SOP} persistently yield better FN rates than the state-of-the-art tools.

Based on our findings discussed so far, we consider {Gap, SOP} to be the most suitable candidate to conduct our study among all the formulations considered above.

2) Selection of a new formulation: The results reported in the last subsection suggest that the objective functions that exhibit a good association with FN rate should be more effective than the other objective functions for estimating phylogenetic trees. Based on this we make an attempt to form a new objective set as follows. We first propose four new objective functions that quantify different aspects of MSA: Entropy, SimG, SimNG and GapCon (the details are presented in Section II-B). We combine these with TC and Gap and run NSGA-III to optimize the objective set {Entropy, TC, Gap, SimG, SimNG, GapCon} for 40 times to generate numerous diverse alignments. We used those to examine the association of our proposed objective functions with FN rate using multiple linear regression analysis (please refer to Figure S5 of the supplementary file for a visualization of the relationship between the relevant pairs of objective functions within the set). The key observations of this analysis are summarized as follows:

- Entropy has a strong correlation with SimG which is problematic for multiple regression analysis. So, we should not keep these two objectives at the same time in our regression model as well as in the multi-objective formulation.
- SimG and SimNG are in conflict with each other. So by optimizing them simultaneously, a multi-objective metaheuristic can generate a large number of diverse alignments.

Now we express the relationship between FN rate and the proposed objective functions using the following model:

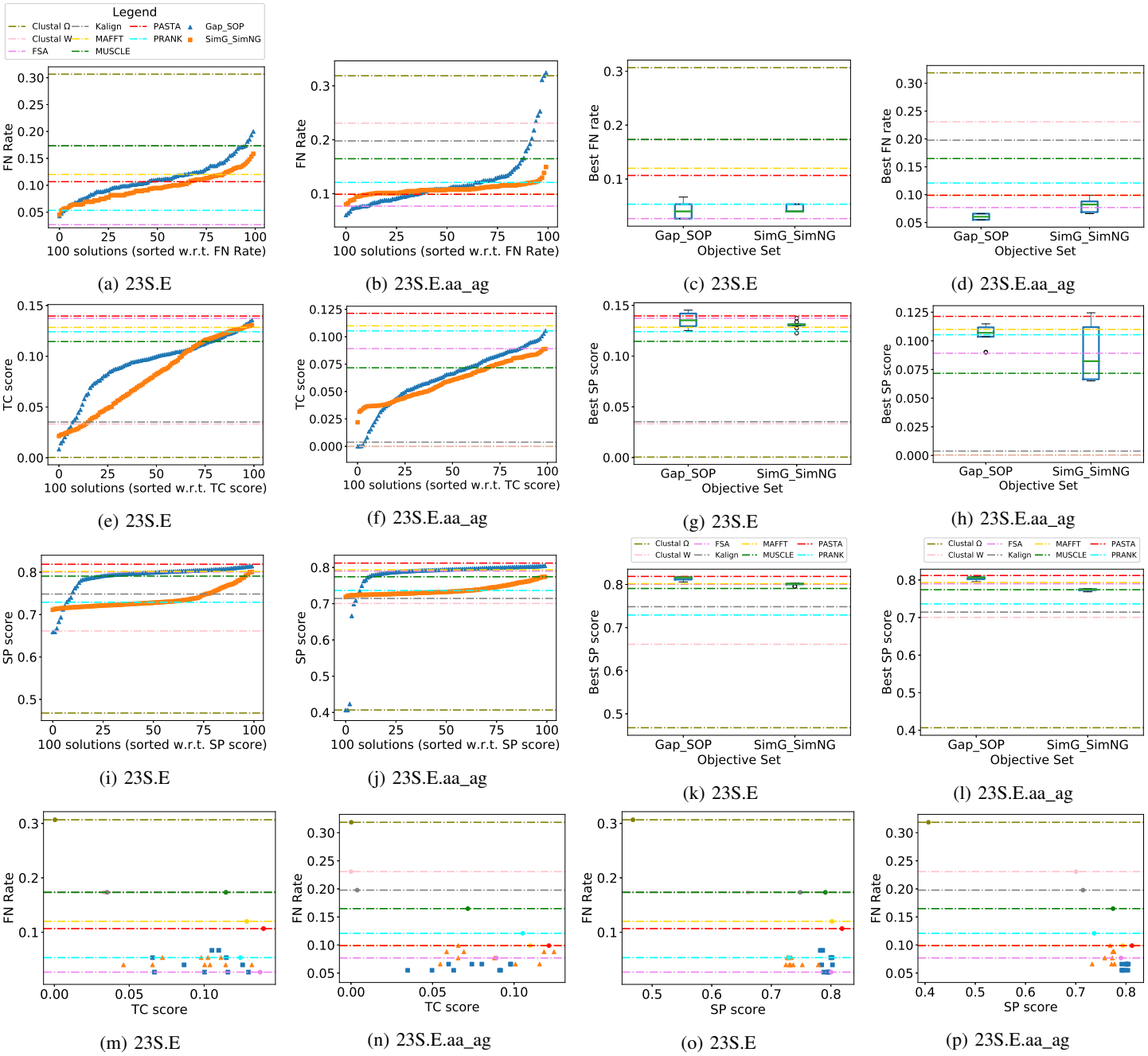
$$\begin{aligned}\text{FN rate} = \beta_0 + \beta_1 \times \text{SimNG} + \beta_2 \times \text{GapCon} + \\ \beta_3 \times \text{SimG (or Entropy)} + \epsilon\end{aligned} \tag{2}$$

We estimate the regression coefficients by fitting the above model to the solutions generated by optimizing the objective set {Entropy, TC, Gap, SimG, SimNG, GapCon}. Here we find that, in each case, both SimG and SimNG exhibit a positive correlation with FN rate. So we choose {SimG, SimNG} as our new objective set. For a visualization of the results using partial regression plots please refer to Figure S6 of the supplementary file.

C. Validation of the selected multi-objective formulations

Now we examine the effectiveness of our chosen formulations (i.e. {Gap, SOP} and {SimG, SimNG}) based on two biological datasets: two biological rRNA datasets and 27 instances of the BAliBASE 3.0 benchmark. To accomplish this we conduct several independent runs of NSGA-II on each dataset considering its stochastic nature according to the standard practice of OR literature. Then we analyze the generated solutions based on the quality of the generated alignments as well as the resultant trees.

1) Results on biological rRNA datasets: We conducted 10 runs of NSGA-II with our two selected objective sets for two datasets, namely, 23S.E and 23S.E.aa_ag. We compare the performance of the multi-objective formulations with respect to FN rate against the nine state-of-the-art tools from two perspectives in the top panel of Figure 3. Here, part (a) and (b) show the averaged FN rate of 100 solutions over 10 runs. Since each run generates 100 solutions, we make the average meaningful by sorting the 100 FN rates per run. Then we average the best FN rates across all the runs. The same applies to the second best ones and so on. These figures (part (a) and (b)) demonstrate a promising aspect of multi-objective



(a) 23S.E (b) 23S.E.aa_ag (c) 23S.E (d) 23S.E.aa_ag

(e) 23S.E (f) 23S.E.aa_ag (g) 23S.E (h) 23S.E.aa_ag

(i) 23S.E (j) 23S.E.aa_ag (k) 23S.E (l) 23S.E.aa_ag

(m) 23S.E (n) 23S.E.aa_ag (o) 23S.E (p) 23S.E.aa_ag

Fig. 3: Biological rRNA datasets: **Panel 1 (Top Panel):** part (a) and part (b) show the averaged FN rate of 100 solutions over 10 independent runs. part (c) and part (d) show the variation of the best FN rates obtained across 10 runs using boxplots. **Panel 2 (Panel 3):** part (e) and part (f) (part (i) and part (j)) show the TC score (SP score) of 100 solutions averaged over 10 runs. part (g) and part (h) (part (k) and part (l)) show the distribution of the best TC scores (SP scores) collected from all runs. **Panel 4 (Bottom panel):** part (m) and part (n) show the relationship between FN rate and TC score for different alignments. part (o) and part (p) show the relationship between FN rate and SP score. In all panels except for the bottom one, we show the performance of nine state-of-the-art tools using dashed horizontal lines; the horizontal lines at the bottom panel mark the FN rates achieved by those tools.

approach that for each data it can generate a substantial number of solutions that are better than the outputs of state-of-the-art tools. However, a practitioner would be interested in the best solutions. Therefore, we summarize the variation of the best FN rate (among 100 values) across 10 runs in part (c) and (d). We see that, for both of the datasets, FSA yielded the best performance among the nine tools, followed by PRANK for 23.S.E and PASTA for 23S.E.aa_ag. The two multi-objective formulations helped to achieve better FN rates than FSA for 23S.E.aa_ag (part (b) and (d)). Here, on average, {Gap ,SOP} generates around 10% solutions that are better than FSA and 40% solutions that are better than PASTA as shown in part (b). On the other hand, on average {SimG, SimNG} produces very few solutions that are better than FSA but around 40% solutions that are better than PASTA. Part (d) shows that {Gap, SOP} consistently outperforms the best tool (FSA) whereas {SimG, SimNG} outperforms FSA in nearly 40% of the total runs. Now let us see the results for 23S.E (part (a) and (c)), where both of the objective sets remain between the best (FSA) and second best (PRANK) tool. Both of them generates around 5% solutions better than PRANK.

We perform similar analysis based on the widely used two alignment quality measures, namely, TC score and SP score and report the results in Panel 2 and 3 of Figure 3 respectively. We notice that, according to these two popular measures, for both the datasets, the alignments generated by multi-objective formulations failed to beat the best performing tool, PASTA. The clear disagreement between FN rate and TC score (part (m) and (n)) as well as between FN rate and SP score (part (o) and (p)) has been illustrated in Panel 4 of Figure 3. To summarize, from the analysis presented in Panel 4 (bottom panel), we realize that the tools/approaches achieving better performance than our multi-objective formulations in terms of the popular measures, namely, TC score and SP score fail to achieve better FN rates than our multi-objective formulations. To elaborate, according to TC score, PASTA is the best performer among the nine tools, and our objective sets have generated several alignments having worse (lower) TC score than PASTA (and FSA). However, those alignments can produce phylogenetic trees with better FN rates than those tools. Even from among the tools, there is disagreement between TC score and FN rate: PASTA is in fact behind FSA in terms of the latter. Similarly in part (o) and (p) of Panel 4 which is dedicated to the comparison between FN rate and SP score, we find several alignments generated by the multi-objective formulations that are worse than PASTA in terms of SP score, but, achieve better FN rates than that tool.

2) Results on BAliBASE datasets: For each of the selected BAliBASE datasets under six groups (RV11, RV12, RV20, RV30, RV40 and RV50), we conducted 20 independent runs of NSGA-II. Once again we analyze the generated solutions based on the quality of alignments and resultant trees. We witnessed that the alignments that are better according to widely accepted alignment scores, not necessarily generate better phylogenetic trees. Here we discuss our key observations on the selected four datasets (BB11005, BB11018, BB11020 and BB11033) under the group RV11. For the remaining groups (RV12, RV20, RV30, RV40 and RV50), our obtained results are similar. For the sake of brevity, we present those results in Section S7 of the supplementary file. Figures 4 and 5 present the results of our experiments on the datasets of group RV11. According to FN rate (part (a) - (h) of Figure 4), at least one of the two objective sets generates better or equivalent solutions than the best tool throughout all the instances. For BB11020, {SimG, SimNG} can achieve 12% FN rate as opposed to 50% FN rate attained by the best tool which is a huge improvement. Considering TC score (part (a) - (h) of Figure 5), the two objective sets can outperform all the tools only for BB11020 which is contrary to the findings based on FN rate. So again we see the disagreement between FN rate and TC score which we examine graphically in part (i) - (l) of Figure 4. If we observe the results based on SP score (part (i) - (p) of Figure 5), we get similar disagreement between FN rate and SP score which is illustrated in part (m) - (p) of Figure 4. These figures provide evidence that a solution with the best TC and/or SP score may not give the best FN rate. We consistently observe this phenomenon across the remaining datasets as well which we present in Section S7 of the supplementary file.

Table IV shows a comparative summary of the 100 solutions generated by a single run of NSGA-II while optimizing {Gap, SOP} with respect to the nine state-of-the-art MSA tools based on FN rate for the 27 randomly selected BAliBASE datasets. Here we see that the multi-objective formulation has been able to generate better phylogenetic trees than all the state-of-the-art MSA tools except on a few cases (marked by cells with 0 value).

3) Statistical significance: Now we confirm the significance of the improvement achieved by the multi-objective formulations in terms of FN rate over nine MSA tools on 27 BAliBASE datasets by applying an appropriate statistical test. We form paired data by picking the FN rate achieved by each (MSA method, dataset) pair. For the metaheuristics, we take the average of the 20 best FN rates from 20 independent runs considering its stochastic nature. As our data do not satisfy the condition of normality and homoscedasticity [44], we choose a series of nonparametric tests following the recommendation of [45]. When applying these tests, we used the p-value threshold as 0.05 which is equivalent to 95% confidence level.

At first, we simultaneously compare all the methods using the Friedman test [46] which gives the relative ranking (lower is better) of all the methods and strongly suggests the existence of significant differences among the methods considered (as p-value is 0). The results have been presented in Column 2 of Table V. Here we see that the multi-objective formulations achieve the top two positions. Next, we complement the Friedman test by following Holm's post-hoc procedure [47] to contrast the difference between the multi-objective formulations and each of the nine tools. The results have been summarized in Columns 3 and 4 of Table V. Here, each cell shows the adjusted p-value which indicates the significance of the difference in performance (based on FN rate) between two methods. We notice that all the p-values are very close to 0 and the values for {SimG, SimNG} are lower than {Gap, SOP}. So we can state with high confidence that, the multi-objective formulations achieve statistically significant improvement over

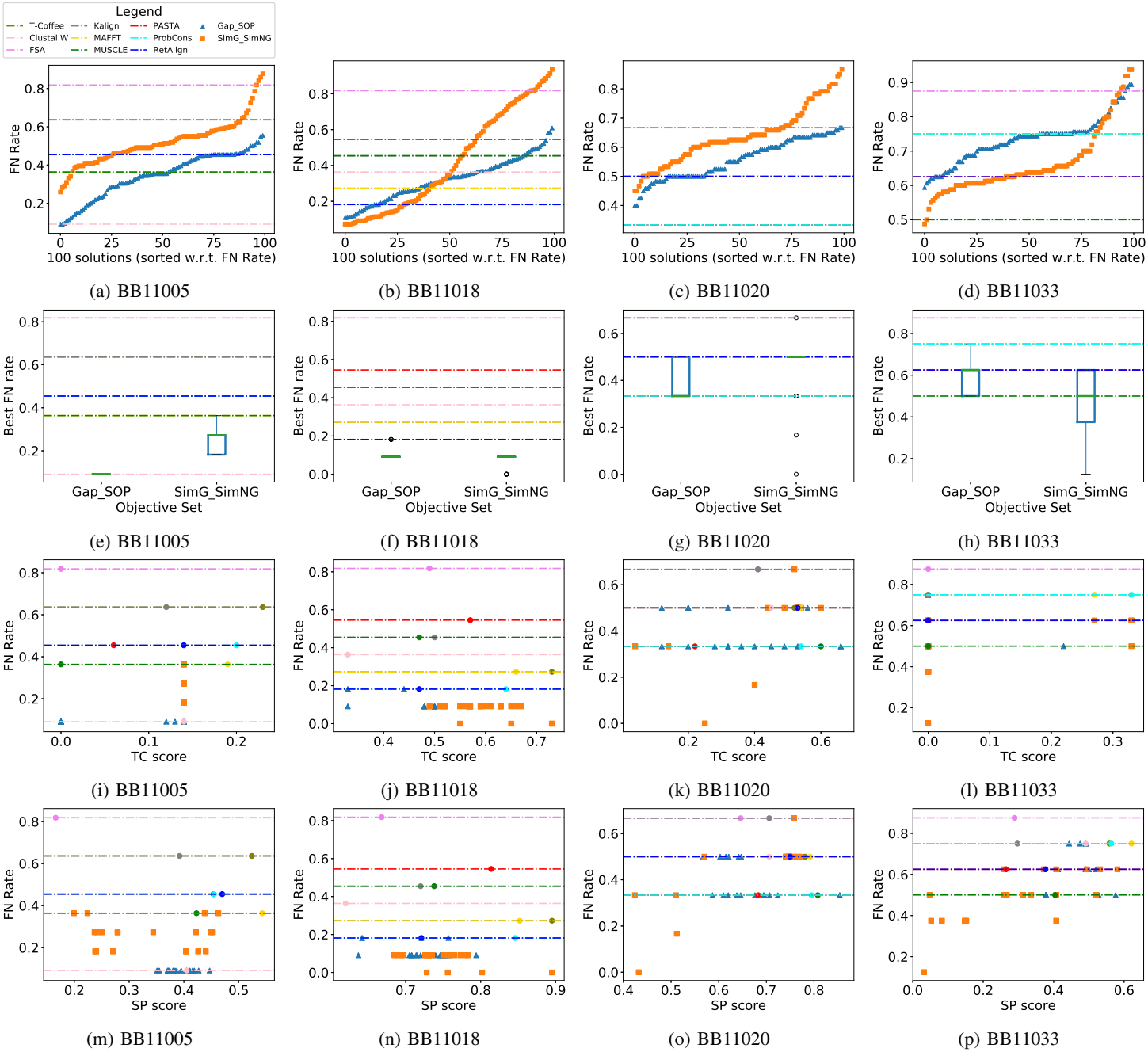


Fig. 4: RV11: **Panel 1 (Top panel):** part (a) - (d) show the FN rate of 100 solutions averaged over 20 independent runs. **Panel 2:** part (e) - (h) show the distribution of the best FN rates collected from all runs. **Panel 3 (Panel 4):** part (i) - (l) (part (m) - (p)) show the relationship between FN rate and TC score (SP score) for different alignments. In all panels, we show the FN rates achieved by the nine state-of-the-art tools using dashed horizontal lines.

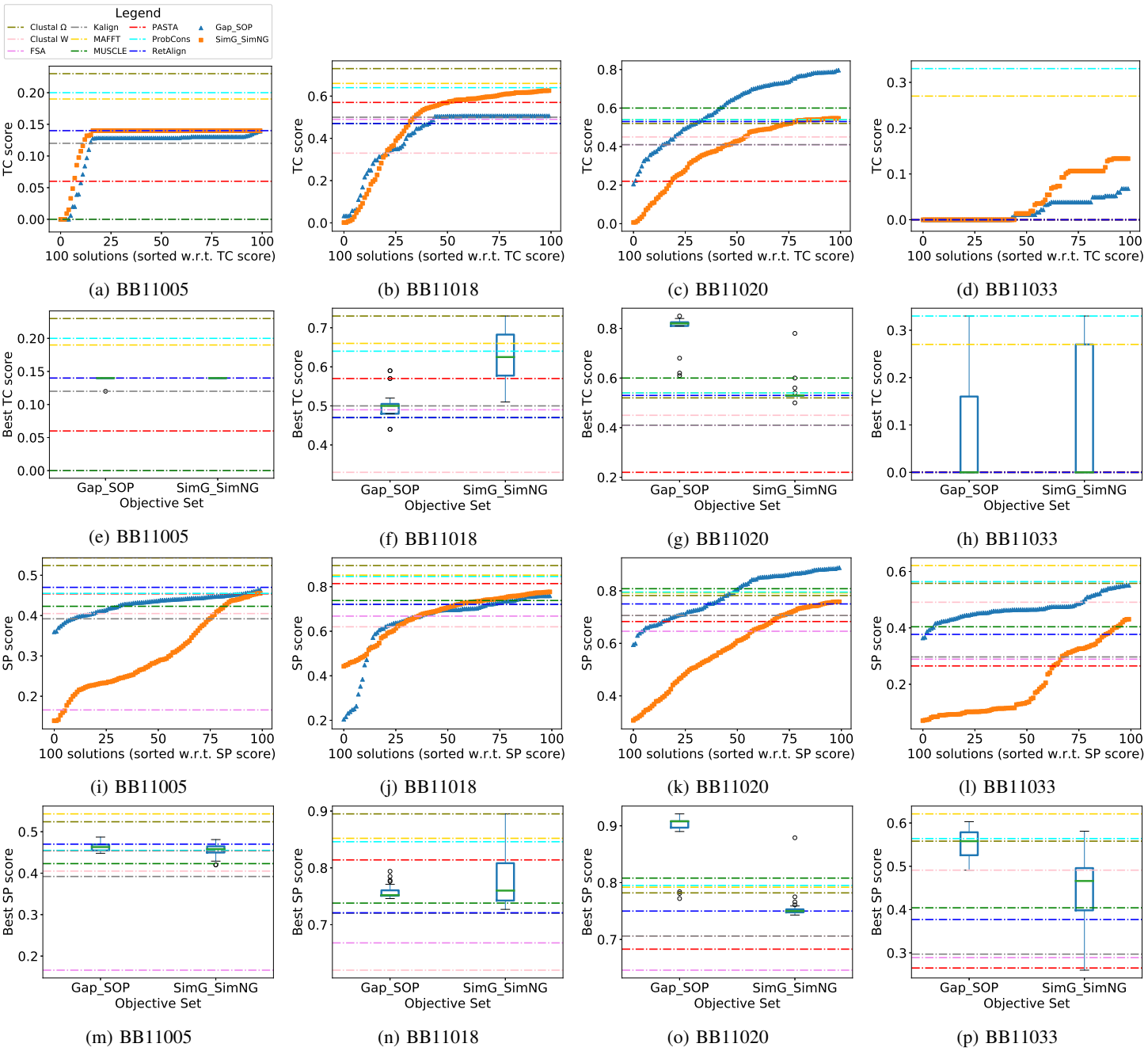


Fig. 5: RV11: **Panel 1 (Panel 3)**: part (a) - (d) (part (i) - (l)) shows the TC score (SP score) of 100 solutions averaged over 20 runs. **Panel 2 (Panel 4)**: part (e) - (h) (part (m) - (p)) shows the distribution of the best TC scores (SP scores) collected from all runs. In all panels, we show the performance of nine state-of-the-art tools using dashed horizontal lines.

the nine MSA tools.

4) Running time: As we are dealing with an offline optimization problem, the runtime is not a major concern in this study. Our multi-objective metaheuristics make an effort to generate improved MSAs for phylogeny estimation by evolving a set of candidate solutions. So depending on the size of the set of candidate solutions, our approach may exhibit higher running time than the state-of-the-art MSA tools; in fact, in our experiments, our approach does require a higher running time. Nonetheless, to put everything into context, here we report runtimes of our multi-objective approaches as well as MAFFT [21] that can generate a competitive alignment within a very reasonable time [8], keeping in mind that the former approach leverages some altered versions of the alignments output by the latter tools. Figure 6 summarizes the average runtimes for each group of BAliBASE datasets. It helps us to identify the differences in runtimes between a two objectives approach and a four objectives one, which would be informative to practitioners and method developers. From this figure, we see that the runtimes of the multi-objective approaches are at least 10 times higher than that of MAFFT. Overall, the set of nonparametric objectives {SimG, SimNG} exhibits

TABLE IV: Comparative summary of the 100 solutions generated by a single run of NSGA-II while optimizing {Gap, SOP} with respect to the nine state-of-the-art MSA tools based on FN rate.

Group	Dataset	Avg. no. of solutions (out of 100) generated by a single run of NSGA-II which are **better or equivalent** to an MSA tool								
		T-Coffee	Clustal W	FSA	Kalign	MAFFT	MUSCLE	PASTA	ProbCons	RetAlign
RV11	BB11005	100	2	100	100	54	54	86	86	86
	BB11018	36	64	100	86	36	86	97	17	17
	BB11033	71	71	97	71	71	0	9	71	9
	BB11020	34	34	100	100	34	0	0	0	34
RV12	BB12001	55	92	92	55	99	92	92	55	25
	BB12013	9	100	9	9	9	9	9	9	9
	BB12022	12	12	12	12	97	12	12	12	97
	BB12035	80	100	11	100	29	90	80	20	73
	BB12044	23	23	23	100	23	23	88	23	88
RV20	BB20001	92	0	1	0	0	92	0	0	15
	BB20010	23	99	1	7	77	23	77	23	7
	BB20022	82	59	82	100	82	82	59	59	0
	BB20033	96	5	96	19	82	29	68	58	88
	BB20041	57	71	38	22	30	83	65	51	71
RV30	BB30002	0	85	45	7	45	0	7	7	7
	BB30008	53	53	98	25	98	98	100	38	90
	BB30015	61	93	93	88	88	61	100	88	88
	BB30022	64	19	47	0	2	47	19	5	84
RV40	BB40001	60	86	60	41	75	41	97	60	97
	BB40013	51	39	39	45	26	45	45	62	62
	BB40025	0	0	0	0	0	49	49	49	49
	BB40038	26	90	66	15	15	66	0	2	66
	BB40048	69	89	69	89	69	89	69	69	69
RV50	BB50001	10	85	62	94	10	100	62	85	85
	BB50005	0	93	0	93	93	93	93	0	93
	BB50010	0	0	0	0	0	28	0	0	96
	BB50016	66	96	0	78	66	78	54	96	78

the lowest runtime among the multi-objective approaches. In several cases (such as, RV12, RV20, RV30, RV40), {SimG, SimNG} runs more than 1.5 times faster than {Gap, SOP}. The calculation of Gap takes a longer period compared to other objectives due to the additional effort of reading the substitution table values continuously. The evaluation of objective functions have been shown in the literature [39] to be the main computational bottleneck for computing MSAs by multi-objective metaheuristics. Therefore, the inclusion of Gap as an objective can heavily affect the overall running time of any algorithm. Moreover, by comparing the runtimes of {SimG, SimNG} and {Gap, SOP, SimG, SimNG}, we find that the runtimes of multi-objective metaheuristics increase linearly in the number of objectives. And the increase in runtime of the four objectives approach is mostly due to Gap. This can encourage more research effort in this direction as adding appropriate objective would definitely increase the accuracy of a multi-objective approach.

IV. Discussion & Conclusion

In this study, we have introduced an application-aware multi-objective formulation to compute MSAs with an ultimate goal to infer the phylogenetic tree from the resultant alignments. To optimize MSA, we proposed two simple objective

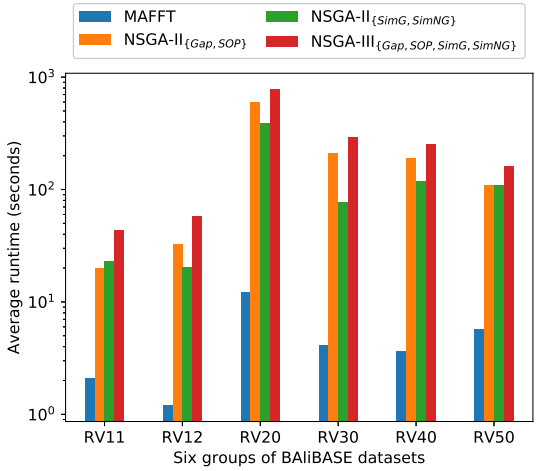


Fig. 6: Average runtimes of multi-objective approaches and MAFFT for each group of BAliBASE datasets.

functions in addition to the existing ones. We judged the potential capability of each objective function to yield better trees by employing domain knowledge as well as by applying statistical approaches. We employed multiple linear regression to measure the degree of association between the individual objective functions and the quality of the inferred phylogenetic tree (i.e., FN rate). Thus, we provide empirical justification to choose two multi-objective formulations to move forward. Afterwards, we performed extensive experimentation with both simulated and biological datasets to demonstrate the benefit of our approach. We showed that the simultaneous optimization of a set of application-aware objective functions can lead to phylogenetic trees with improved accuracy than that of the state-of-the-art MSA tools. From this finding, we would like to hypothesize that, the use of domain specific measures can aid MSA methods in other application domain as well. In the sequel we identified {SimG, SimNG} to be the best set of objective functions for computing MSAs with an aim to infer phylogeny, considering its overall accuracy, runtime and nonparametric nature.

MSAs are computed to serve various biological purposes including phylogeny estimation and protein structure prediction. The definition of what constitutes a true alignment can depend partly on the purpose of MSAs [33]. Nevertheless, regardless of the purpose, the sites within the true alignment define the “homologies”. Therefore, homology can be based on structural features or evolutionary histories, leading to the opposing concepts of “structural homology” and “evolutionary homology” [33]. While structural alignments are expected to be close to the true (evolutionary) alignment, convergent evolution may create conditions where the best structural alignment puts nucleotides or amino acids in the same site (thus implying homologies), even though these specific homologies are not present in the true evolutionary alignment [48]. In other words, structural homology may not be identical to evolutionary homology [49]. In such a situation, generic metrics such as TC/SP score might not be adequate to assess the correctness of the estimated MSAs. Therefore, using more informative

TABLE V: Friedman test (Column 2): The Average Friedman's ranking (lower is better) achieved by the MSA methods over 27 BAliBASE datasets. We performed the Friedman test based on FN rate achieved by the tools. Holm's post-hoc procedure (Columns 3 and 4): Comparison between the metaheuristics and the MSA tool using the Holm's post-hoc procedures (as a complement of the Friedman test) over 27 BAliBASE datasets.

1	2	3	4
Method	Friedman's Rank*	Holm's adjusted p-value	
		NSGA-II$_{\{\text{SimG, SimNG}\}}$	NSGA-II$_{\{\text{Gap, SOP}\}}$
NSGA-II$_{\{\text{SimG, SimNG}\}}$	2.2037	-	0.46018
NSGA-II$_{\{\text{Gap, SOP}\}}$	2.8704	0.46018	-
ProbCons	5.7963	0.00014	0.00238
Clustal Ω	6.2963	0.00002	0.00044
MAFFT	6.4074	0.00001	0.00036
Kalign	6.7037	0.00000	0.00011
PASTA	6.8148	0.00000	0.00007
FSA	6.9444	0.00000	0.00004
MUSCLE	7.1482	0.00000	0.00002
Clustal W	7.3519	0.00000	0.00001
RetAlign	7.4630	0.00000	0.00000
*Statistic	10.5911	N/A	
*p-value	0.00000		

metrics (e.g. phylogeny as done in the study) to tailor adequate multi-objective formulation of this problem seems a promising endeavor.

Standard criteria (SP score, TC score, etc.) for assessing alignment quality are usually based on shared homology pairs (SP score) or identical columns (TC score), and do not explicitly consider a particular application domain. Mistakes in alignments that are not important with respect to an application domain may not impact the ultimate accuracy of that particular inference. For example, not all sites are significant with respect to protein structure and function prediction, and hence multiple alignments with different accuracy may lead to the same predictions [9]. Similarly, in the context of phylogeny estimation, alignments with substantially different SP scores may lead to trees with the same accuracy [11]. In this study, we systematically investigate the impact of evaluation criteria of an alignment on phylogenetic tree inference problem. Our results suggest that it could be possible to develop improved MSA methods for phylogenetic analysis by carefully choosing appropriate objective functions. Moreover, in almost all existing studies on MSA, we find the researchers evaluating the effectiveness of MSA methods using some generic alignment quality measures (i.e., TC score, SP score). Contrastingly, our results revealed that optimizing those widely used measures do not necessarily lead us to the best phylogenetic tree. This finding could be an eye opener for the researchers who need to use MSA methods to address a particular application.

Our findings and proposed multi-objective formulation can be particularly beneficial for iterative methods like SATé and PASTA that iteratively co-estimate both alignment and tree. These methods obtain an initial alignment and a tree that guide each other to improved estimates in an iterative fashion. They make an effort to exploit the close association between the accuracy of an MSA and the corresponding tree in finding the output through multiple iterations from both directions. Therefore, carefully choosing an evaluation metric for an MSA with a better correlation to the tree accuracy seems likely to improve the results of these co-estimation techniques. Thus, our methodology, if adopted, may potentially have a profound positive impact on the accuracy of these iterative co-estimation techniques. Moreover, multiple "good" alignments from the output of the multi-objective approach can be served as alternative MSAs for several methods (such as [8]) which would then utilize all of them to infer phylogeny with better accuracy.

This study will encourage the scientific community to investigate various application-aware measures for computing and evaluating MSAs. This will potentially prompt more experimental studies addressing specific application domains; and ultimately will propel our understanding of MSAs and their impact in various domains in computational biology, i.e, phylogeny estimation, protein structure and function prediction, orthology prediction etc. This study will also encourage the researchers to develop new scalable MSA tools by simultaneously optimizing multiple appropriate optimization criteria. Thus, we believe that this study will pioneer new models and optimization criteria for computing MSAs – laying a firm, broad foundation for application specific multi-objective formulation for estimating multiple sequence alignment.

We performed an extensive experimental study comprising 29 datasets of varying sizes and complexities, and our findings are consistent throughout all the datasets. Still, we acknowledge the possibility of facing a few unforeseen circumstances as follows. There might be some datasets on which our approach might not exhibit satisfactory performance. Besides, currently we did not pay any effort to improve the running time of our approach which is higher as compared to top MSA tools. However, sufficient speedup could be achieved by leveraging modern computing architectures (computer cluster, GPU, etc.).

Formulating application-aware multi-objective formulation (application specific evaluation criteria in general) cannot be developed entirely in one study; it should evolve in response to scientific findings and systematists' feedback. This requires the active involvement of evolutionary biologists, computer scientists, systematists, and others – leading to improved understandings of alignments and how they are related to various fields in comparative genomics.

REFERENCES

[1] G. Jordan and N. Goldman, "The effects of alignment error and alignment filtering on the sitewise detection of positive selection," *Molecular biology and evolution*, vol. 29, no. 4, pp. 1125–1139, 2011.

[2] J.-M. Chang, P. Di Tommaso, and C. Notredame, "Tcs: a new multiple sequence alignment reliability measure to estimate alignment accuracy and improve phylogenetic tree reconstruction," *Molecular Biology and Evolution*, vol. 31, no. 6, pp. 1625–1637, 2014.

[3] J. A. Lake, "The order of sequence alignment can bias the selection of tree topology." *Molecular Biology and Evolution*, vol. 8, no. 3, pp. 378–385, 1991.

[4] D. G. Croan, D. A. Morrison, and J. T. Ellis, "Evolution of the genus leishmania revealed by comparison of dna and rna polymerase gene sequences," *Molecular and biochemical parasitology*, vol. 89, no. 2, pp. 149–159, 1997.

[5] T. H. Ogden and M. S. Rosenberg, "Multiple sequence alignment accuracy and phylogenetic inference," *Systematic biology*, vol. 55, no. 2, pp. 314–328, 2006.

[6] M. Wu, S. Chatterji, and J. A. Eisen, "Accounting for alignment uncertainty in phylogenomics," *PloS one*, vol. 7, no. 1, p. e30288, 2012.

[7] B. D. Redelings and M. A. Suchard, "Joint bayesian estimation of alignment and phylogeny," *Systematic biology*, vol. 54, no. 3, pp. 401–418, 2005.

[8] H. Ashkenazy, I. Sela, E. Levy Karin, G. Landan, and T. Pupko, "Multiple sequence alignment averaging improves phylogeny reconstruction," *Systematic biology*, vol. 68, no. 1, pp. 117–130, 2018.

[9] T. Warnow, "Large-scale multiple sequence alignment and phylogeny estimation," in *Models and algorithms for genome evolution*. Springer, 2013, pp. 85–146.

[10] S. Mirarab, N. Nguyen, S. Guo, L.-S. Wang, J. Kim, and T. Warnow, "Pasta: ultra-large multiple sequence alignment for nucleotide and amino-acid sequences," *Journal of Computational Biology*, vol. 22, no. 5, pp. 377–386, 2015.

[11] K. Liu, S. Raghavan, S. Nelesen, C. R. Linder, and T. Warnow, "Rapid and accurate large-scale coestimation of sequence alignments and phylogenetic trees," *Science*, vol. 324, no. 5934, pp. 1561–1564, 2009.

[12] F. Sievers, A. Wilm, D. Dineen, T. J. Gibson, K. Karplus, W. Li, R. Lopez, H. McWilliam, M. Remmert, J. Söding *et al.*, "Fast, scalable generation of high-quality protein multiple sequence alignments using clustal omega," *Molecular systems biology*, vol. 7, no. 1, p. 539, 2011.

[13] A. Löytynoja and N. Goldman, "An algorithm for progressive multiple alignment of sequences with insertions," *Proceedings of the National academy of sciences of the United States of America*, vol. 102, no. 30, pp. 10 557–10 562, 2005.

[14] T. Lassmann, O. Frings, and E. L. Sonnhammer, "Kalign2: high-performance multiple alignment of protein and nucleotide sequences allowing external features," *Nucleic acids research*, vol. 37, no. 3, pp. 858–865, 2008.

[15] R. K. Bradley, A. Roberts, M. Smoot, S. Juvekar, J. Do, C. Dewey, I. Holmes, and L. Pachter, "Fast statistical alignment," *PLoS computational biology*, vol. 5, no. 5, p. e1000392, 2009.

[16] A. Szabó, Á. Novák, I. Miklós, and J. Hein, "Reticular alignment: A progressive corner-cutting method for multiple sequence alignment," *BMC bioinformatics*, vol. 11, no. 1, p. 570, 2010.

[17] C. Notredame, D. G. Higgins, and J. Heringa, "T-coffee: a novel method for fast and accurate multiple sequence alignment1," *Journal of molecular biology*, vol. 302, no. 1, pp. 205–217, 2000.

[18] C. B. Do, M. S. Mahabhashyam, M. Brudno, and S. Batzoglou, "Probcons: Probabilistic consistency-based multiple sequence alignment," *Genome research*, vol. 15, no. 2, pp. 330–340, 2005.

[19] Y. Liu, B. Schmidt, and D. L. Maskell, "Msaprobs: multiple sequence alignment based on pair hidden markov models and partition function posterior probabilities," *Bioinformatics*, vol. 26, no. 16, pp. 1958–1964, 2010.

[20] U. Roshan and D. R. Livesay, "Probalign: multiple sequence alignment using partition function posterior probabilities," *Bioinformatics*, vol. 22, no. 22, pp. 2715–2721, 2006.

[21] K. Katoh, K. Misawa, K.-i. Kuma, and T. Miyata, "Mafft: a novel method for rapid multiple sequence alignment based on fast fourier transform," *Nucleic acids research*, vol. 30, no. 14, pp. 3059–3066, 2002.

[22] R. C. Edgar, "Muscle: multiple sequence alignment with high accuracy and high throughput," *Nucleic acids research*, vol. 32, no. 5, pp. 1792–1797, 2004.

[23] J. Pei and N. V. Grishin, "Mummals: multiple sequence alignment improved by using hidden markov models with local structural information," *Nucleic acids research*, vol. 34, no. 16, pp. 4364–4374, 2006.

[24] Á. Rubio-Largo, L. Vanneschi, M. Castelli, and M. A. Vega-Rodríguez, "A characteristic-based framework for multiple sequence aligners," *IEEE transactions on cybernetics*, vol. 48, no. 1, pp. 41–51, 2018.

[25] D. DeBlasio and J. Kececioglu, "Parameter advising for multiple sequence alignment," in *BMC bioinformatics*, vol. 16, no. 2. BioMed Central, 2015, p. A3.

[26] J. D. Thompson, B. Linard, O. Lecompte, and O. Poch, "A comprehensive benchmark study of multiple sequence alignment methods: current challenges and future perspectives," *PloS one*, vol. 6, no. 3, p. e18093, 2011.

[27] F. J. M. da Silva, J. M. S. Pérez, J. A. G. Pulido, and M. A. V. Rodríguez, "Alineaga—a genetic algorithm with local search optimization for multiple sequence alignment," *Applied Intelligence*, vol. 32, no. 2, pp. 164–172, 2010.

[28] F. M. Ortuño, O. Valenzuela, F. Rojas, H. Pomares, J. P. Florido, J. M. Urquiza, and I. Rojas, "Optimizing multiple sequence alignments using a genetic algorithm based on three objectives: structural information, non-gaps percentage and totally conserved columns," *Bioinformatics*, vol. 29, no. 17, pp. 2112–2121, 2013.

[29] W. Soto and D. Becerra, "A multi-objective evolutionary algorithm for improving multiple sequence alignments," in *Brazilian Symposium on Bioinformatics*. Springer, 2014, pp. 73–82.

[30] M. Abbasi, L. Paquete, and F. B. Pereira, "Local search for multiobjective multiple sequence alignment," in *International Conference on Bioinformatics and Biomedical Engineering*. Springer, 2015, pp. 175–182.

[31] Á. Rubio-Largo, M. A. Vega-Rodríguez, and D. L. González-Álvarez, "A hybrid multiobjective memetic metaheuristic for multiple sequence alignment," *IEEE Transactions on Evolutionary Computation*, vol. 20, no. 4, pp. 499–514, 2016.

[32] C. Zambrano-Vega, A. J. Nebro, J. García-Nieto, and J. F. Aldana-Montes, "Comparing multi-objective metaheuristics for solving a three-objective formulation of multiple sequence alignment," *Progress in Artificial Intelligence*, pp. 1–16, 2017.

[33] T. Warnow, *Computational phylogenetics: an introduction to designing methods for phylogeny estimation*. Cambridge University Press, 2017.

[34] D. Kalyanmoy, *Multi objective optimization using evolutionary algorithms*. John Wiley and Sons, 2001.

[35] R. R. Rani and D. Ramyachitra, "Multiple sequence alignment using multi-objective based bacterial foraging optimization algorithm," *Biosystems*, vol. 150, pp. 177–189, 2016.

[36] S. Henikoff and J. G. Henikoff, "Amino acid substitution matrices from protein blocks," *Proceedings of the National Academy of Sciences*, vol. 89, no. 22, pp. 10 915–10 919, 1992.

[37] K. Deb, A. Pratap, S. Agarwal, and T. Meyarivan, "A fast and elitist multiobjective genetic algorithm: Nsga-ii," *IEEE transactions on evolutionary computation*, vol. 6, no. 2, pp. 182–197, 2002.

[38] K. Deb and H. Jain, "An evolutionary many-objective optimization algorithm using reference-point-based nondominated sorting approach, part i: Solving problems with box constraints." *IEEE Trans. Evolutionary Computation*, vol. 18, no. 4, pp. 577–601, 2014.

[39] C. Zambrano-Vega, A. J. Nebro, J. García-Nieto, and J. F. Aldana-Montes, "M2align: parallel multiple sequence alignment with a multi-objective metaheuristic," *Bioinformatics*, vol. 33, no. 19, pp. 3011–3017, 2017.

[40] ——, "A multi-objective optimization framework for multiple sequence alignment with metaheuristics," in *International Conference on Bioinformatics and Biomedical Engineering*. Springer, 2017, pp. 245–256.

[41] 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, 1994.

[42] J. D. Thompson, P. Koehl, R. Ripp, and O. Poch, "Balibase 3.0: latest developments of the multiple sequence alignment benchmark," *Proteins: Structure, Function, and Bioinformatics*, vol. 61, no. 1, pp. 127–136, 2005.

[43] D. C. Montgomery, E. A. Peck, and G. G. Vining, *Introduction to linear regression analysis*. John Wiley & Sons, 2012, vol. 821.

[44] D. J. Sheskin, *Handbook of parametric and nonparametric statistical procedures*. crc Press, 2003.

[45] J. Derrac, S. García, D. Molina, and F. Herrera, "A practical tutorial on the use of nonparametric statistical tests as a methodology for comparing evolutionary and swarm intelligence algorithms," *Swarm and Evolutionary Computation*, vol. 1, no. 1, pp. 3–18, 2011.

[46] M. Friedman, "The use of ranks to avoid the assumption of normality implicit in the analysis of variance," *Journal of the american statistical association*, vol. 32, no. 200, pp. 675–701, 1937.
[47] S. Holm, "A simple sequentially rejective multiple test procedure," *Scandinavian journal of statistics*, pp. 65–70, 1979.
[48] S. Iantorno, K. Gori, N. Goldman, M. Gil, and C. Dessimoz, "Who watches the watchmen? an appraisal of benchmarks for multiple sequence alignment," in *Multiple Sequence Alignment Methods*. Springer, 2014, pp. 59–73.
[49] G. R. Reeck, C. De Haen, D. C. Teller, R. F. Doolittle, W. M. Fitch, R. E. Dickerson, P. Chambon, A. D. McLachlan, E. Margoliash, T. H. Jukes *et al.*, ""homology" in proteins and nucleic acids: a terminology muddle and a way out of it," *Cell*, vol. 50, no. 5, p. 667, 1987.

Muhammad Ali Nayeem received his B.Sc. and M.Sc. degrees in Computer Science and Engineering from the Department of Computer Science and Engineering (CSE), Bangladesh University of Engineering and Technology (BUET), Dhaka, Bangladesh in 2013 and 2016, respectively.

Md. Shamsuzzoha Bayzid is an Assistant Professor in the Department of Computer Science and Engineering (CSE) at Bangladesh University of Engineering and Technology (BUET). He received his Ph.D. in Computer Science from the University of Texas at Austin (UT Austin) in 2016, and was supervised by Prof. Tandy Warnow. He completed the B. Sc. Engg. and M. Sc. Engg. degrees in Computer Science from BUET in 2008 and 2010, respectively.

Atif Hasan Rahman is an Assistant Professor in the Department of Computer Science and Engineering (CSE) at the Bangladesh University of Engineering and Technology (BUET). He completed his Ph.D. from the University of California, Berkeley in 2015 under the supervision of Lior Pachter. His research area is bioinformatics and computational biology and is currently focusing on statistical methods for genome assembly and analysis.

Rifat Shahriyar is working as an Assistant Professor at Department of Computer Science and Engineering (CSE) of Bangladesh University of Engineering and Technology (BUET). He has completed Ph.D. in April 2015 from Research School of Computer Science in Australian National University (ANU). His Ph.D. under the supervision of Steve Blackburn and Kathryn McKinley changed the way people think about reference counting and conservative garbage collection.

M. Sohel Rahman is a Professor of the CSE department of BUET. He had previously worked as a Visiting Senior Research Fellow of King's College London. He is a Senior Member of both IEEE and ACM and a member of American Mathematical Society (AMS) and London Mathematical Society (LMS). He is also a Peer-review Associate College Member of EPSRC, UK. Dr. Rahman's research interest includes, but is not limited to, theory and algorithms, stringology, metaheuristics, applied informatics, computational biology and bioinformatics.