

*Original Paper*

# Statistical Comparative Study of Multiple Sequence Alignment Scores of Iterative Refinement Algorithms

DAIGO WAKATSU<sup>†1</sup> and TAKEO OKAZAKI<sup>‡2</sup>

Iterative refinement algorithm is a useful method to improve the alignment results. In this paper, we evaluated different iterative refinement algorithms statistically. There are four iterative refinement algorithms: remove first (RF), bestfirst (BF), random (RD), and tree-based (Tb) iterative refinement algorithm. And there are two scoring functions for measuring the iteration judgment step: log expectation (LE) and weighted sum-of-pairs (SP) scores. There are two sequence clustering methods: neighbor-joining (NJ) method and unweighted pair-group method with arithmetic mean (UPGMA). We performed comprehensive analyses of these alignment strategies and compared these strategies using BALiBASE SP (BSP) score. We observed the behavior of scores from the view point of cumulative frequency (CF) and other basic statistical parameters. Ultimately, we tested the statistical significance of all alignment results by using Friedman nonparametric analysis of variance (ANOVA) test for ranks and Scheffé multiple comparison test.

## 1. Introduction

Multiple sequence alignment has become an essential method in molecular biology such as phylogenetic analysis and protein structure prediction. Many different techniques have been developed. For example, ProbCons<sup>1)</sup> uses Bayesian consistency and fills the primary library using the posterior decoding of a pair hidden Markov model. MAFFT<sup>2)</sup> uses fast Fourier transform technique, and SAGA<sup>3)</sup> uses a genetic algorithm to try and optimize a multiple sequence alignment given an objective function. Progressive alignment<sup>4)</sup> is the most widely used heuristic approach for aligning a large number of sequences. Multiple sequence

alignment is performed by progressively aligning pairs of sequences followed by pairs of alignments/profiles. The guide tree determines the order in which these pairs are aligned. This technique is used in many different multiple sequence alignment programs such as ClustalW<sup>5)</sup>, T-COFFEE<sup>6)</sup>, and MUSCLE<sup>7)</sup>. However, failures occurrence in the alignment process can never be corrected in the progressive alignment technique. The iterative refinement algorithm solves this problem. By applying dynamic programming to partially aligned sequences iteratively, their alignment quality can be improved. Such an iterative strategy employs heuristic search methods to solve practical alignment problems. Many different iterative refinement techniques have been proposed; MUSCLE uses a tree-dependent restricted partitioning technique for the iterations. PRRP/PRRN<sup>8)</sup> uses a best-first iterative refinement strategy with tree-dependent partitioning. Most multiple sequence alignment programs were reviewed by Thompson, et al.<sup>9)</sup>, Notredame, et al.<sup>10),11)</sup>, Wallace<sup>12)</sup>, and Pirovano<sup>13)</sup>.

Hirosawa, et al.<sup>14)</sup> investigated the performance of different iterative refinement algorithms. They tested the effectiveness of each algorithm by using the sum-of-pairs score in order to improve the alignment results. They used a group of 30 protein kinase sequences for evaluating the effectiveness of the algorithms. Wallace, et al.<sup>15)</sup> systematically tested different iterative refinement algorithms by using HOMSTRAD, which is a database of structure-based alignments for homologous protein families<sup>16)</sup>. They showed that iterative refinement algorithms could be used to effectively improve the performance of progressive alignment by using existing alignment software programs. Iterative refinement algorithms were found to be very effective when they were directly incorporated into the progressive alignment scheme. For example, direct incorporation of remove first iterative refinement algorithm into ClustalW improved its average accuracy by 6%<sup>15)</sup>.

In this paper, we revisited important studies on iterative refinement algorithms. There are several types of iterative refinement algorithms, scoring functions for measuring the iteration steps, and sequence clustering methods. We carried out comprehensive analyses of these alignment strategies. Hirosawa, et al. did not consider the various types of scoring functions and clustering methods in their study. Wallace, et al. did not consider the sequence clustering method in

<sup>†1</sup> Information Engineering Course, Graduate School of Engineering and Science, University of the Ryukyus

<sup>‡2</sup> Department of Information Engineering, Faculty of Engineering, University of the Ryukyus

their study. Hirosawa, et al. evaluated the alignment strategies by the mean value of the alignment scores (sum-of-pairs scores) and the execution time. On the other hand, Wallace, et al. evaluated the alignment strategies by using the mean value of column score. The column score was calculated by the number of identical columns in the reference alignment and the alignment to be tested as a percentage of the number of columns in the reference. Hirosawa, et al. and Wallace, et al. did not consider the statistical evaluation enough, they only used mean value. In this study, we evaluated alignment strategies more statistically. For this purpose, we considered the distribution of scores and other statistical values and tested the statistical significance of all the alignment strategies by using the Friedman ANOVA test<sup>17)</sup>. When significant differences were found among the alignment strategies by using the Friedman test, appropriate post-hoc tests for multiple comparisons were performed. To determine the significance of specific combinations of strategies, we used a Scheffé multiple comparison test<sup>18)</sup> by all-pair comparisons. Moreover, we studied the characteristics of different types of data sets and evaluated the performance of all alignment strategies on the basis of the sequence types. We identified the best and worst alignment strategies for each type of data set on the basis of the statistical significance of the strategies.

## 2. Benchmark Data Set

BaliBASE 3.0 a benchmark alignment database<sup>19)</sup> is used to compare the performances of different alignment algorithms. BaliBASE contains 218 reference alignments and is divided into six different reference sets, each having different characteristics (**Table 1**).

Reference 1-1 provides the alignments of equi-distant, very divergent sequences

(identity: <20%) divided into 38 alignment sets. Reference 1-2 provides the alignments of equi-distant, medium to very divergent sequences divided into 44 alignment sets. Reference 2 reports families aligned with highly divergent “orphan” sequences divided into 41 alignment sets. Reference 3 reports subgroups with a residual identity of <25% between groups divided into 30 alignment sets. Reference 4 reports sequences with N/C-terminal extensions divided into 49 alignment sets. Reference 5 reports sequences with large internal insertions grouped into 16 alignment sets.

## 3. Multiple Sequence Alignment Algorithm

Seven multiple sequence alignment algorithms were selected for comparison.

The progressive alignment (PA) algorithm performs multiple sequence alignment by successively aligning pairs of sequences/profiles. The guide tree determines the order in which the sequences/profiles are to be aligned. Initially, two sequences are chosen by the guide tree and aligned by standard pairwise alignment using the Needleman-Wunsch<sup>20)</sup> algorithm. In the alignment process, a new sequence is added to an existing alignment and certain rules are used to determine the order in which the sequences are aligned.

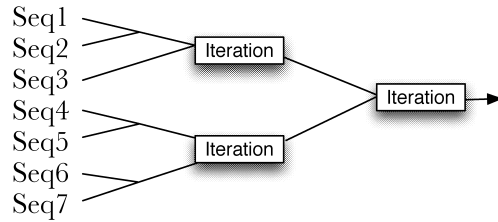
The remove first iterative refinement (RF) algorithm has a simple iterative strategy. In each iteration step, one sequence is removed from the alignment and realigned to the remaining alignment. If the alignment result is better than the previous one, it is retained and used as the input for the next iteration. The iteration cycle is terminated if the alignment score converges or when  $2N^2$  iterations are completed.  $N$  is the number of sequences.

In the random iterative refinement (RD) algorithm, the alignment is split randomly into two sets of sequences, which are then realigned. If the score improves, the alignment result is retained. The iteration cycle is terminated if the limit of  $2N^2$  splits are carried out.

In the bestfirst iterative refinement (BF) algorithm, in each iteration cycle, every sequence is removed from the alignment and realigned to the rest. The alignment with the best score is used as the input for the next iteration. The iteration cycle is terminated if the alignment score converges or if the limit of  $2N^2$  profile-profile alignment is reached.

**Table 1** BaliBASE reference alignments.

References	Sets	Contents
1-1	38	equi-distant sequence (very divergent sequences)
1-2	44	equi-distant sequence (medium to very divergent sequences)
2	41	families aligned with a highly divergent “orphan” sequence
3	30	subgroups with a residue identity of <25% between groups
4	49	sequences with N/C-terminal extensions
5	16	sequences with internal insertions



**Fig. 1** Tree-based iterative refinement algorithm.

In the tree-based iterative refinement (Tb) algorithm, the alignment improvement algorithms are incorporated into a progressive alignment strategy as shown in **Fig. 1**. Whenever two profiles are combined, the resultant alignment is refined using one of the other iterative refinement algorithms described above. Therefore, the Tb algorithm have three types, Tb using RF (TbRF), Tb using RD (TbRD), and Tb using BF (TbBF).

#### 4. Scores

Two different types of scores were used for the multiple sequence alignment and used in iterative refinement algorithms to measure the iteration judgment steps.

The sum-of-pairs (SP) score is a well-known scoring function for the multiple sequence alignment. To calculate the score of a multiple sequence alignment, the scores of each pair of rows in the multiple sequence alignment are summed to obtained the overall score. The SP score of a multiple sequence alignment  $A$  of length  $l$  constructed from  $N$  nucleotide or amino acid sequences is defined as follows:

$$\text{SP}(A) = \sum_{j=2}^N \sum_{k=1}^{j-1} S_{j,k}, \quad (1)$$

where  $S_{j,k}$  is the score associated with the pairwise alignment of the  $j$ th and  $k$ th sequences within  $A$ . In this study, we use the weighted SP score, which is used by ClustalW. When a set of weights,  $\{w_{j,k}\}$ , is given to individual pairs of sequences in  $A$ , the weighted SP score of  $A$  is analogously defined as follows:

$$\text{weighted SP}(A) = \sum_{j=2}^N \sum_{k=1}^{j-1} w_{j,k} S_{j,k} \quad (2)$$

The weights assigned to individual pairs of sequences are adjusted to compensate for biased contributions.

The log expectation (LE) scoring function is used in MUSCLE.

$$\text{LE}^{xy} = (1 - f_G^x)(1 - f_G^y) \log \sum_i \sum_j \frac{f_i^x f_j^y p_{ij}}{p_i p_j} \quad (3)$$

Here,  $p_i$  is the background probability of amino acid  $i$ ;  $p_{ij}$ , the joint probability of  $i$  and  $j$  being aligned;  $f_i^x$  the observed frequency of  $i(j)$  in column  $x$ ; and  $f_G^x$ , the observed frequency of gaps in column  $x$  in the first profile.

The weighted SP and LE scoring functions are used for multiple sequence alignment and used in iterative refinement algorithms to measure iteration judgement steps.

#### 5. Statistical Method for Comparison

We examined 28 alignment strategies, which were combinations of seven multiple sequence alignment schemes (PA, RF, RD, BF, TbRF, TbRD, and TbBF), two scoring functions (SP and LE score), and two sequence-clustering methods (NJ and UPGMA).

To compare all the strategies statistically, we used basic statistical parameters mean, maximum, minimum, median, and variance as a guide for the evaluation measure. It is preferable that the mean, maximum, and minimum values are high and that the variance is low.

In addition to using these index values, we employed another statistical view. We regarded the distribution of scores as important. In order to estimate the distribution function, we utilized the parameter cumulative frequency (CF). It is preferable that the frequency is low for low scores and that it increases rapidly for high scores. The distribution of the scores was determined by using CF.

Further, we determined the statistical significance of all alignment results. As the significance of the results could not be evaluated using the above-mentioned evaluation indices, a nonparametric ANOVA test by ranks was used. Consider-

ing the characteristics of the data, we used the Friedman ANOVA test because three or more paired groups were compared in this method and this test was a nonparametric alternative to the two-way ANOVA for ranks. Because significant differences were found among the alignment strategies by using the Friedman ANOVA test, the multiple comparison test was performed for the post-hoc tests. The significant nonparametric ANOVA results suggest that the global null hypothesis  $H_0$ : “The distributions of the ranks are identical” should be rejected. Multiple comparison procedures were then used to identify the distributions that were different from others. To determine the significance of a specific combination, we used a Scheffé multiple comparison test to perform all-pair comparison. Ultimately, we identified the best and worst alignment strategies for characteristics of different types of data sets.

## 6. Experimental Results

BioPerl (<http://www.bioperl.org>) modules was used to implement the iterative refinement algorithms. The multiple sequence alignment program MUSCLE (v 3.7) generates alignments to be used as the progressive alignment and the profile-profile alignment in the process of iterative refinement algorithms. We used the amino acid scoring matrix VTML 240<sup>21)</sup>. In the iteration judgement step, we used weighted SP (ClustalW’s sequence weighting) and LE scoring functions to determine whether the score was improving. Furthermore, we employed the NJ and UPGMA clustering methods. These scoring functions and clustering methods were implemented by using MUSCLE. The PA results was used as initial alignments in the iterative refinement algorithms.

To measure the performance of multiple alignment, we used BALiBASE SP (BSP) scores. Given a true and estimated multiple sequence alignment, the accuracy of the estimated alignment is usually computed using the BSP score. The BSP score is the ratio of the number of correctly aligned pairs in the core blocks of the test alignment to the number of aligned pairs in the reference alignment. The core block is a region in which reliable alignments are known to exist.

$$\text{BSP} = \frac{\sum_{i < j} S_{i,j}}{S_r} \quad (4)$$

**Table 2** Mean Values of BSP Scores. The highest value for each reference is highlighted in bold with underline, the lowest value is highlighted in italic with underline.

PA					RF			
Reference	NJ		UPGMA		NJ		UPGMA	
	LE	SP	LE	SP	LE	SP	LE	SP
1-1	<u>0.399</u>	0.408	0.403	0.429	0.413	0.430	0.425	0.437
1-2	<u>0.788</u>	0.794	0.792	0.821	0.805	0.823	0.800	0.826
2	<u>0.763</u>	0.773	0.783	0.798	0.776	0.787	0.788	<b>0.803</b>
3	0.680	0.674	0.680	0.665	<b>0.689</b>	0.681	<b>0.689</b>	0.679
4	0.684	0.707	0.702	0.732	0.701	0.727	0.711	<b>0.750</b>
5	0.622	0.637	0.650	0.668	0.642	0.654	0.653	0.673
RD					BF			
Reference	NJ		UPGMA		NJ		UPGMA	
	LE	SP	LE	SP	LE	SP	LE	SP
1-1	0.404	0.413	0.410	0.431	0.419	0.428	0.429	0.442
1-2	<u>0.788</u>	0.807	0.795	0.824	0.805	0.823	0.809	0.827
2	<u>0.760</u>	0.770	0.780	0.794	<u>0.773</u>	0.789	0.788	<b>0.803</b>
3	0.677	0.669	0.678	0.662	<b>0.689</b>	0.682	0.687	0.676
4	0.682	0.703	0.701	0.729	0.697	0.730	0.711	0.746
5	0.619	0.630	0.649	0.665	0.647	0.659	0.654	0.679
TbRF					TbRD			
Reference	NJ		UPGMA		NJ		UPGMA	
	LE	SP	LE	SP	LE	SP	LE	SP
1-1	0.433	0.416	0.431	0.442	0.415	0.412	0.417	0.432
1-2	0.807	0.823	0.810	0.833	0.789	0.809	0.809	0.828
2	0.780	0.786	0.793	0.800	<u>0.758</u>	0.767	0.784	0.794
3	0.680	0.677	0.685	0.676	<u>0.674</u>	0.668	0.669	<u>0.659</u>
4	0.699	0.730	0.688	0.729	<u>0.676</u>	0.693	0.689	0.716
5	0.644	0.660	0.663	0.678	<u>0.613</u>	0.636	0.650	0.675
TbBF								
Reference	NJ		UPGMA					
	LE	SP	LE	SP				
1-1	0.433	0.419	0.438	<b>0.450</b>				
1-2	0.809	0.825	0.808	<b>0.834</b>				
2	0.780	0.790	0.793	0.801				
3	0.682	0.683	0.685	0.680				
4	0.700	0.722	0.692	0.732				
5	0.641	0.667	0.655	<b>0.685</b>				

$$S_{i,j} = \sum P_{i,j}^k \quad (5)$$

$S_r$  is the total number of residual pairs in a core block of the test alignment. If the pair of residues  $i$  and  $j$  in column  $k$  of the test alignment exist in a core block of the reference alignment,  $P_{i,j}^k$  becomes 1, otherwise 0.

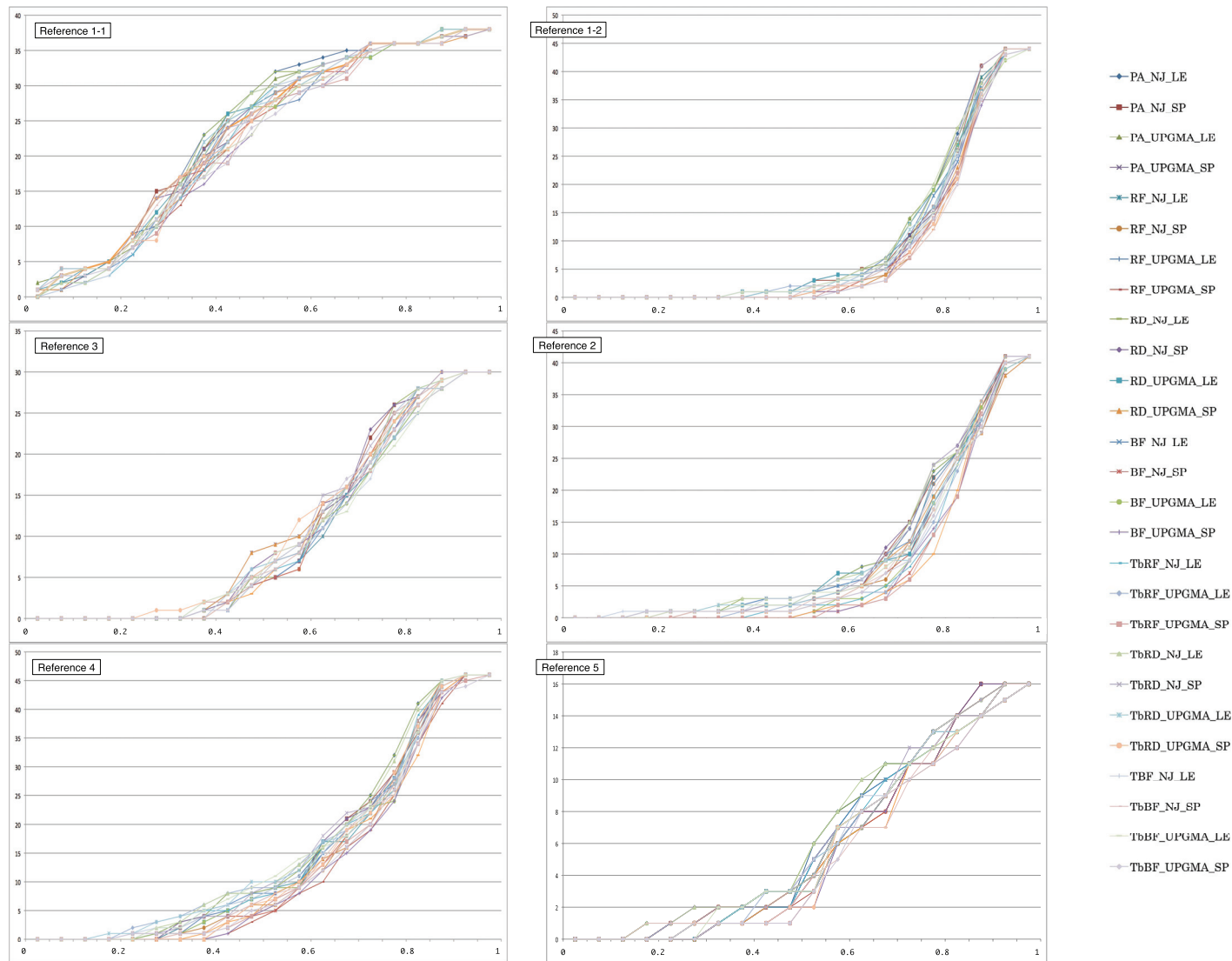


Fig. 2 CF values of BSP scores.

**Table 3** Sum of CF values. The highest value for each reference is highlighted in bold with underline, the lowest value is highlighted in italic with underline.

Reference	PA				RF			
	NJ		UPGMA		NJ		UPGMA	
	LE	SP	LE	SP	LE	SP	LE	SP
1-1	<u>477</u>	469	472	452	467	451	456	445
1-2	<u>210</u>	201	207	183	193	176	200	175
2	216	207	199	185	205	195	197	182
3	206	211	205	218	<b>202</b>	207	<b>202</b>	209
4	315	293	293	275	297	276	289	<b>254</b>
5	129	124	119	115	121	120	118	111

Reference	RD				BF			
	NJ		UPGMA		NJ		UPGMA	
	LE	SP	LE	SP	LE	SP	LE	SP
1-1	472	467	468	452	460	453	453	441
1-2	207	188	203	179	192	178	191	175
2	220	212	203	188	207	162	177	161
3	209	215	207	<u>219</u>	<b>202</b>	207	<b>202</b>	209
4	316	299	293	277	300	274	289	257
5	131	127	118	116	121	120	117	110

Reference	TbRF				TbRD			
	NJ		UPGMA		NJ		UPGMA	
	LE	SP	LE	SP	LE	SP	LE	SP
1-1	451	466	453	445	466	466	463	452
1-2	191	177	191	172	209	192	191	173
2	174	161	177	<b>158</b>	<u>221</u>	213	196	188
3	206	209	205	209	210	215	213	220
4	298	272	312	271	<u>318</u>	306	309	282
5	122	117	115	111	<u>134</u>	124	120	111

Reference	TbBF			
	NJ		UPGMA	
	LE	SP	LE	SP
1-1	451	460	446	<b>440</b>
1-2	189	176	190	<b>171</b>
2	200	194	192	185
3	205	208	204	208
4	298	281	306	267
5	122	113	118	<b>107</b>

**Table 2** shows the mean values of the BSP scores. One of the scoring functions (LE or SP) and one of the sequence clustering methods (NJ or UPGMA) were paired for each strategy. The best and worst scores for each reference have been underlined in the table. **Figure 2** shows the CFs for the BSP scores. It is preferable that the frequency is low for low score and increases rapidly for

**Table 4** Results of Friedman ANOVA test of BSP scores.

References	P value
Reference 1-1	5.41.E-05
Reference 1-2	6.23.E-30
Reference 2	9.50.E-45
Reference 3	1.37.E-03
Reference 4	3.42.E-35
Reference 5	1.37.E-12

the high scores. **Table 3** shows the sums of the CF values used to assess the performance of the CF. It is preferable that the sums of the CF values are low. The best and worst scores for each reference have been underlined in the table.

The statistical significance of the BSP scores was determined using rank statistics. **Table 4** shows the results of the Friedman ANOVA test on the BSP scores. It was clear that there were significant differences among the alignment strategies for each reference.

In the next step, the typical differences among the alignment strategies for each reference have to be determined. **Table 5** shows the rank sum value of each alignment strategy. It is preferable that the rank sum value is high. The best and worst scores for each reference have been underlined in the table. We performed a post-hoc multiple comparison test, that was, the Scheffé multiple comparison test for all-pair comparisons in order to determine the statistical significance of the alignment strategies. Our results showed some significant differences in Reference 1-2, 2, and 4. On the basis of rank sum order and the significant difference, we identified efficient alignment strategies in **Table 6**. We also identified some inefficient strategies in **Table 7**. The multiple comparison test did not reveal any significant difference in Reference 1-1, 3, and 5.

The above-mentioned results indicate that the UPGMA clustering method provides better performance than NJ clustering method. NJ is known as the most reliable method in predicting the correct phylogenetic tree, because the branch lengths of trees are allowed to vary in a manner that simulates varying levels of evolutionary change. UPGMA assumes the same evolutionary speed on all lineages. It is generally not considered a suitable method for construction of phylogenetic trees as it relies on the rates of evolution among different lineages to be approximately equal. However, Hirosawa, et al. and Wallace, et al. used

**Table 5** Rank sum values of BSP scores. The highest value for each reference is highlighted in bold with underline, the lowest value is highlighted in italic with underline.

Reference	PA				RF			
	NJ		UPGMA		NJ		UPGMA	
	LE	SP	LE	SP	LE	SP	LE	SP
1-1	462.0	<u>402.0</u>	487.0	582.0	542.0	542.5	617.0	593.5
1-2	<u>326.5</u>	388.5	500.0	629.0	603.0	727.5	669.5	765.0
2	335.0	395.5	693.5	758.0	529.5	644.5	771.0	<b>821.0</b>
3	444.5	401.0	476.0	373.0	527.5	464.0	<b>531.0</b>	453.0
4	<u>422.5</u>	558.5	679.5	828.5	664.0	811.0	876.5	<b>1030.0</b>
5	133.5	160.5	238.5	297.5	198.0	200.0	250.5	323.0

Reference	RD				BF			
	NJ		UPGMA		NJ		UPGMA	
	LE	SP	LE	SP	LE	SP	LE	SP
1-1	468.5	440.0	498.0	597.0	574.5	535.0	632.5	638.5
1-2	364.0	454.0	564.5	684.0	645.5	736.5	704.0	801.5
2	<u>256.0</u>	331.0	577.0	634.0	537.5	645.0	754.0	788.5
3	388.0	333.5	421.5	<u>312.0</u>	525.5	469.0	<b>531.0</b>	445.0
4	380.0	528.5	635.0	<u>766.5</u>	658.5	822.5	863.5	961.0
5	<u>110.5</u>	135.0	233.0	275.0	211.0	226.5	263.0	326.0

Reference	TbRF				TbRD			
	NJ		UPGMA		NJ		UPGMA	
	LE	SP	LE	SP	LE	SP	LE	SP
1-1	586.0	449.0	614.5	660.5	516.5	409.5	539.0	592.0
1-2	650.0	700.5	<b>847.0</b>	842.5	389.5	468.0	669.0	715.5
2	626.5	549.0	714.5	760.5	263.0	279.5	614.5	627.5
3	435.5	402.5	503.0	438.0	389.0	326.5	374.5	345.0
4	640.0	844.5	718.0	831.5	352.5	514.0	673.5	701.5
5	200.0	212.0	308.5	303.0	112.0	142.5	241.5	289.0

Reference	TbBF			
	NJ		UPGMA	
	LE	SP	LE	SP
1-1	628.0	489.5	<b>671.0</b>	660.5
1-2	666.0	722.5	800.5	830.0
2	632.0	631.0	766.0	711.0
3	436.0	457.0	506.0	471.5
4	645.5	844.5	739.5	903.0
5	209.0	266.0	276.5	<b>354.5</b>

**Table 6** Efficient strategies (high score order).

Reference 1-2	Reference 2	Reference 4
TbRF using UPGMA and LE	RF using UPGMA and SP	RF using UPGMA and SP
TbRF using UPGMA and SP	BF using UPGMA and SP	BF using UPGMA and SP
TbBF using UPGMA and SP	RF using UPGMA and LE	TbBF using UPGMA and SP
	TbBF using UPGMA and LE	RF using UPGMA and LE
	TbRF using UPGMA and SP	
	PA using UPGMA and SP	
	BF using UPGMA and LE	

**Table 7** Inefficient strategies.

Reference 1-2	Reference 2	Reference 4
PA using NJ and LE	RD using NJ and LE	TbRD using NJ and LE
	TbRD using NJ and LE	RD using NJ and LE
	TbRD using NJ and SP	PA using NJ and LE
	RD using NJ and SP	TbRD using NJ and SP
	PA using NJ and LE	

give good performance.

In previous studies, Hirosawa, et al. showed that PA algorithm gave the worst mean scores. Hirosawa, et al. used 30 protein kinase data sets that were the same characteristics test sequence sets. Reference 1-1 and 1-2 have the similar characteristics to these. Table 2 shows PA algorithm gave the worst mean scores in Reference 1-1 and 1-2, those were the same as Hirosawa, et al.'s results. And Hirosawa, et al. showed that TbRF, TbRD, and TbBF algorithms had the same performance. However in this study, TbBF showed good mean scores in Reference 1-1 and 1-2. The supposable reason might be the restrictive data sets. Hirosawa, et al. used 30 data sets and the sequence length was limited to 80, although we covered the sequences of various length. Wallace, et al. showed RD and TbRD gave bad mean scores and TbBF gave the best mean scores, those were the same as our results. On the other hand, Wallace, et al. showed the LE scoring function provides better performance. However, the weighted SP scoring function provided better performance than LE in this study. The possible reason might be the difference of score in the iteration judgement step. We used LE scoring function in the iteration judgement step, when LE were used as alignment process, ditto with weighted SP scoring function. Generally the scoring function used in the alignment process should be used in the iteration judgment step.

not NJ but UPGMA without reason. Katoh, et al.<sup>22)</sup> indicated that in the case of construction of guide tree for the progressive alignment, UPGMA was more efficient method than NJ. Our results matched to these indications.

Table 7 shows RD and TbRD algorithms gave inefficient performance in Reference 2 and 4. Also in the other references, RD and TbRD algorithms did not

However, Wallace, et al. used the SP scoring function by multiple sequence alignment package SAGA in all iteration judgement steps. This difference of scoring function for the iteration judgement step might have influenced the results.

## 7. Conclusion

We evaluated different iterative refinement algorithms statistically. In this study, we performed a comprehensive analyses of alignment strategies computing seven alignment algorithms, two scoring functions for the iteration judgement step, and two sequence clustering methods. We considered the characteristics of different types of data sets and evaluated the performance of all strategies on the basis of sequence types on the BALiBASE benchmark database.

From the results of nonparametric statistical tests, we found that there were significant statistical differences among the alignment strategies for all BALiBASE references and identified efficient strategies for each reference.

## References

- 1) Do, C.B., Mahabhashyam, M.S., Brudno, M. and Batzoglou, S.: ProbCons: Probabilistic consistency-based multiple sequence alignment, *Genome Res.*, Vol.15, pp.330–340 (2005).
- 2) Katoh, K., Kuma, K., Toh, H. and Miyata, T.: MAFFT version 5: Improvement in accuracy of multiple sequence alignment, *Nucleic Acids Res.*, Vol.33 pp.511–518 (2005).
- 3) Notredame, C. and Higgins, D.G.: SAGA: sequence alignment by genetic algorithm, *Nucleic Acids Research*, Vol.24, pp.1515–1524 (1996).
- 4) Taylor, W.R.: Multiple sequence alignment by a pairwise algorithm, *Comput. Appl. Biosci.*, Vol.3, pp.81–87 (1987).
- 5) Thompson, J.D., Higgins, D.G. and Gibson, T.J.: ClustalW: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice, *Nucleic Acids Res.*, Vol.22, pp.4673–4680 (1994).
- 6) Notredame, C., Higgins, D.G. and Heringa, J.: T-Coffee: A novel method for fast and accurate multiple sequence alignment, *Journal of Molecular Biology*, Vol.302, pp.205–217 (2000).
- 7) Edgar, R.C.: MUSCLE: Multiple sequence alignment with high accuracy and high throughput, *Nucleic Acids Res.*, Vol.32, pp.1792–1797 (2004).
- 8) Gotoh, O.: Significant improvement in accuracy of multiple protein sequence alignments by iterative refinement as assessed by reference to structural alignments, *J. Mol. Biol.*, Vol.264, pp.823–838 (1996).
- 9) Thompson, J.D., Plewniak, F. and Poch, O.: A comprehensive comparison of multiple sequence alignment programs, *Nucleic Acids Res.*, Vol.27, No.13, pp.2682–2690 (July 1999).
- 10) Notredame, C.: Recent progresses in multiple sequence alignment: A survey, *Pharmacogenomics*, Vol.31, No.1, pp.131–144 (2002).
- 11) Notredame, C.: Recent evolutions of multiple sequence alignment algorithms, *PLoS. Computational Biology*, Vol.3, No.8, pp.e123 (2007).
- 12) Wallace, I.M., Blackshields, G. and Higgins, D.G.: Multiple sequence alignments, *Curr. Opin. Struct. Biol.*, Vol.15, No.3, pp.261–266 (2005).
- 13) Pirovano, W. and Heringa, J.: Multiple sequence alignment, *Methods Mol. Bio.*, Vol.452, pp.143–161 (2008).
- 14) Hirosawa, M., Totoki, Y., Hoshida, M. and Ishikawa, M.: Comprehensive study on iterative algorithms of multiple sequence alignment, *Comput. Appl. Biosci.*, Vol.11, No.1, pp.13–18 (1995).
- 15) Wallace, I.M., O'Sullivan, O. and Higgins, D.G.: Evaluation of iterative alignment algorithms for multiple alignment, *Bioinformatics*, Vol.21, No.8, pp.1408–1414 (2005).
- 16) Mizuguchi, K.: HOMSTRAD: A database of protein structure alignments for homologous families., *Protein Sci.*, Vol.7, pp.2469–2471 (1998).
- 17) Friedman, M.: A comparison of alternative tests of significance for the problem of m rankings, *The Annals of Mathematical Statistics*, Vol.11, No.1, pp.86–92 (1940).
- 18) Scheffé, H.: A method for judging all contrasts in the analysis of variance, *Biometrika*, Vol.40, pp.87–104 (1953).
- 19) Thompson, J.D., Koehl, P., Ripp, R. and Poch, O.: BALiBASE 3.0: latest developments of the multiple sequence alignment benchmark, *Proteins: Structure, Function, and Bioinformatics*, Vol.61, pp.127–136 (2005).
- 20) Needleman, S.B. and Wunsch, C.D.: A general method applicable to the search for similarities in the amino acid sequence of two proteins, *J. Mol. Biol.*, Vol.48, pp.443–453 (1970).
- 21) Muller, T., Spang, R. and Vingron, M.: Estimating amino acid substitution models: A comparison of Dayhoff's estimator, the resolvent approach and a maximum likelihood method, *Mol. Biol.*, Vol.19, No.1, pp.8–13 (2002).
- 22) Katoh, K. and Misawa, K.: Multiple Sequence Alignment — overview of recent softwares, *Biochem. J.*, Vol.46, No.6, pp.312–317 (2006).

(Received January 7, 2009)

(Accepted March 14, 2009)

(Released June 22, 2009)

(Communicated by Susumu Goto)





**Daigo Wakatsu** was born in 1984. He received his M.E. degree from University of the Ryukyus in 2009. He has been working in NEC Soft Okinawa, Ltd. since 2009.



**Takeo Okazaki** was born in 1965. He received his M.Sc. degree from Kyushu University in 1989. He was a research assistant at Kyushu University from 1989, and has been a lecturer at University of the Ryukyus from 1995. His current research interests are the genetic causal network estimation, the normalization of gene expression data and the sequence assembling.

---