IPSJ Transactions on Bioinformatics Vol. 4 2–8 (Jan. 2011)

Database/Software Paper

A Web Server for Multi-objective Pairwise RNA Sequence Alignment with an Index for Selecting Accurate Alignments

Akito Taneda[†1]

The importance of non-coding RNAs and their informatics tools has grown for a decade due to a drastic increase of known non-coding RNAs. RNA sequence alignment is one of the most important technologies in such informatics tools. Recently, we have proposed a multi-objective genetic algorithm, Cofolga2mo, for obtaining an approximate set of weak Pareto optimal solutions for global pairwise RNA sequence alignment, where a sequence similarity and a secondary structure contribution are taken into account as objective functions. In the present study, we have developed a web server for obtaining RNA sequence alignments by Cofolga2mo and for assisting the decision making from the alignments. Furthermore, we introduced an index for reducing the number of alignments output by Cofolga2mo. As a result, we successfully reduced the maximum number of alignments for an input RNA sequence pair from fifty to ten without a significant loss of accurate alignments. By using the BRAliBase 2.1 benchmark dataset, we show that a set of alignments output by Cofolga2mo for an input RNA sequence pair, which has at most ten alignments, includes an accurate alignment compared to those of the previous mono-objective RNA sequence alignment programs.

1. Introduction

Non-coding RNAs (ncRNAs) play various important roles in cells such as regulation [1], splicing [2] and maturation of other functional biomolecules [3] in addition to the well-known functions in translation (e.g., transfer RNAs and ribosomal RNAs). Since ncRNAs usually have characteristic secondary structures in accordance with their cellular functions, the secondary structures of ncRNAs which have a same function can be conserved even when the nucleotide sequences do not conserved. For this reason, inclusion of a structural information into an alignment algorithm is essential for RNA sequence alignment and recent RNA sequence alignment methods have utilized secondary structures to obtain an accurate alignment.

Fusion of sequence alignment and RNA secondary structure prediction has first been realized in Sankoff's algorithm [4]. Since Sankoff's algorithm is an attractive idea and the computational complexity of its naive implementation is very high, many variations of Sankoff's algorithm have been proposed and used to align structure-unknown RNA sequences; the variations range from dynamic programming to stochastic algorithms [5)–12]. In addition to these softwares distributed as a downloadable file at each website, several web servers have also been developed and users can freely perform an RNA sequence alignment at the web servers through the Internet: e.g., SCARNA and Murlet servers at ncRNA.org [13], LocARNA server at Vienna RNA web servers [14], MASTR server [11], R-Coffee server [15] and the FOLDALIGN web server [16]. The number of the websites which provide a web service of RNA sequence alignment is scarce compared to the number of RNA sequence alignment programs distributed as a downloadable file. One possible reason for this is the relatively high computational complexities of the RNA sequence alignment algorithms. Since installation of a downloaded software into each PC is tedious and web services can give an install-less richer usability including graphical interfaces, increasing the number of the websites where users can freely perform RNA sequence alignment is important for the researchers who have an interest in analyzing RNA sequences.

Recently, we have developed an efficient RNA sequence alignment program called Cofolga2mo [17], which is based on multi-objective genetic algorithm (MOGA) [18]. Cofolga2mo explores weak Pareto optimal solutions in an objective function space composed of a sequence similarity and a consensus structure score, and outputs an approximate set of weak Pareto optimal alignments. In RNA sequence alignment, sequence similarity and consensus structure score are usually conflicted, i.e., there is a tradeoff between these two objective functions. Since a MOGA outputs not a single but multiple optimal solutions when a tradeoff exists between objective functions, Cofolga2mo also usually outputs multiple solutions (the approximate set). In our previous paper [17], we showed that the approximate set (composed of at most fifty alignments) includes an accurate alignment

†1 Graduate School of Science and Technology, Hirosaki University

© 2011 Information Processing Society of Japan

compared to the alignments obtained by previous state-of-the-art mono-objective RNA sequence alignment programs. In the present paper, we propose a simple index for reducing the number of the solutions output from Cofolga2mo; by sorting the fifty alignments output from Cofolga2mo with the index, we can reduce the maximum number of alignments output for an input RNA sequence pair to ten without a significant loss of an accurate alignment included in the original fifty alignments.

When we use a MOGA to solve a combinatorial optimization problem, a user (decision maker) has to inspect a set of solutions produced by the MOGA in order to find a result suitable for his/her purpose since MOGA usually outputs multiple solutions. To inspect the results, not a character-based user interface (CUI) but a graphical user interface (GUI) is useful for MOGA. In this paper, we present a web server for executing Cofolga2mo and browsing the approximate set of weak Pareto optimal alignments with a GUI. In addition to the output alignments, the consensus secondary structure predicted for each output alignment can also be browsed at the web server in an interactive way. By inspecting the consensus secondary structure predicted for each output alignment, users can select an RNA sequence alignment more intuitively than when using the CUI version of Cofolga2mo.

2. Methods

2.1 Cofolga2mo Algorithm

Since the detail of Cofolga2mo algorithm is published in the previous paper [17], we briefly describe the algorithm here. The aim of Cofolga2mo is to obtain an approximate set of weak Pareto optimal solutions for RNA sequence alignment problem, where a solution is represented by a global pairwise alignment of a user-given RNA sequence pair. Weak Pareto optimal solutions are defined as a set of solutions which are not *strongly dominated* by any other solution, where Solution A is said to strongly dominate Solution B if the values of all objective functions of Solution A are strictly better than those of Solution B. In Cofolga2mo, a sequence similarity score s and a consensus structure score P are used as objective functions. A higher value is better for these two objective functions. The s is calculated with RIBOSUM85-60 [19] and affine gap penalties (an opening gap penalty = 30 and an elongation gap penalty = 4; terminal gap penalties are set to zero). These values of gap penalties are taken from our previous paper [17]. The P is calculated by the following formula:

$$P = \sum_{i<j} b_{ij}, \tag{1}$$

where b_{ij} is an arithmetic mean of the base paring probabilities of two RNA sequences, where i and j indicate alignment column positions and the base paring probabilities are computed by RNAfold [20]; b_{ij} is set to zero if the base pairing probability of one of the RNAs is zero. Exploration of weak Pareto optimal solutions in Cofolga2mo proceeds on the basis of a standard genetic algorithm (GA) [21], where a population of solutions (pairwise alignments) is randomly generated at initialization step and then the population is improved by iteratively applying evaluation and reproduction procedures; a GA population size of fifty is used in the present web server. In evaluation step, the s and P of each alignment are calculated and a 'dominance rank' is assigned to each alignment. A lower value of the dominance rank indicates the solution is closer to the weak Pareto optimal solutions, and a dominance rank of one is assigned to the best solutions in each population. The dominance rank is computed by using non-dominated sorting [18]. In reproduction step, solutions with a dominance rank of one in the current population are copied to the population of the next GA iteration (elite-preserving strategy). Thereafter child solutions are generated by applying GA operators to parent solutions until the population for the next GA iteration is fulfilled by the elite and newly generated child solutions, where the parents are randomly selected with a dominance-rank-based roulette wheel selection. The iteration between evaluation and reproduction steps is stopped when a maximum iteration number of 200 is reached or the number of solutions with a dominance rank of one has not changed for a continuous thirty GA iterations.

2.2 A Linear Weight Index for Selecting Accurate Alignments

A set of alignments output by Cofolga2mo can include various alignments, which range from that with a low s and a high P to that of a high s and a low P. When a user aligns an RNA sequence pair having a low sequence identity, we can expect that good alignments exist in the alignments with a high P and a low s, and *vice versa*. By utilizing this idea, we can reduce the number of alignments

© 2011 Information Processing Society of Japan

output by Cofolga2mo before manually inspecting the results. In the present study, we propose a linear weight index I for reducing the number of alignments output by Cofolga2mo. The I is calculated by the following linear combination of s and P:

$$I = s + wP, \qquad w = \alpha\sigma + \beta, \tag{2}$$

where w is a weight factor and σ is the percent sequence identity. The σ is calculated based on the alignment with the highest s in the approximate set of weak Pareto optimal alignments obtained by Cofolga2mo; α and β are parameters to be determined based on a training data. After obtaining an approximate set of weak Pareto optimal solutions, we can assign I to each alignment by using their s and P. Then we sort all alignments contained in the approximate set in descending order of I and obtain a sorted alignment list. By picking up the top alignments in the sorted alignment list, we can select a subset of the approximate set in accordance with the percent sequence identity σ of the input RNA sequence pair.

In the present study, the values of α and β for the I were determined by using 5,010 reference pairwise alignments (the k2-dataset used in the previous paper [17]) taken from the BRAliBase 2.1 [22], which are composed of alignments with a wide range of sequence identities. The best set of α and β was determined by finding the set of α and β which gives the highest mean Mathews correlation coefficient (CC). The CC is defined as follows:

$$\mathrm{CC} = \frac{TP \cdot TN - FP \cdot FN}{\sqrt{(TP+FP)(TP+FN)(TN+FP)(TN+FN)}}, \tag{3}$$

where TP, TN, FP and FN indicate the number of correctly predicted base pairs, the number of negative pairs predicted as negative, the number of negative pairs incorrectly predicted as positive and the number of positive pairs incorrectly predicted as negative, respectively. The reference base pairs were obtained by mapping the secondary structures of Rfam 7.0 [23] to the reference alignments. The predicted base pairs used for the alignment performance evaluation were generated by applying RNAalifold [24] to a predicted alignment. We evaluated the alignment performance of Cofolga2mo by a mean CC obtained for the BRAliBase 2.1 k2-dataset. The mean CC of Cofolga2mo was calculated by using the best CC in the top alignments in each alignment list sorted by I, where the alignment list is output by Cofolga2mo for each input RNA sequence pair.

2.3 Web Server

Cofolga2mo web server is available at our website [⋆1]. In Cofolga2mo web server, two RNA sequences whose lengths are ≤ 200 nucleotides can be aligned by pasting the two RNA sequences (in fasta format) into the input form of the website. In the submission page of the website, a user can select the number, n_{out}, of output alignments and the measure for sorting the alignments before submitting a job. As a measure for the sorting, I, s and P are available (a default setting is I). An initial random number can also be selected at the submission page.

The results of Cofolga2mo are browsed through two types of web pages. The first one (sorted alignment list) contains a table of weak Pareto optimal alignments computed by Cofolga2mo. In this page, top n_{out} alignments are tabulated in descending order of the measure selected in the submission page. From the first web page, we can move to the second web page (predicted consensus structure), where a user can interactively browse the consensus secondary structure predicted for each alignment. Since Cofolga2mo does not have an ability for outputting a consensus structure, the consensus structure prediction is performed with RNAalifold [24]. In the web page for browsing the consensus structure, the predicted secondary structure of each of two input RNA sequences is visualized by using a Java API of VARNA 3.7 [25]. In addition, at the bottom of the web page for browsing the structure, the predicted consensus secondary structure and an alignment for which the structure was predicted are shown.

In both the first and second web pages, a plot in a s-P plane drawn with gnuplot is revealed at the top of each page. By browsing the plots, user can know the distribution of the obtained solutions in the objective function space. Cofolga2mo web server is implemented in perl.

3. Results

3.1 Parameter Determination

As mentioned above, the values of parameters α and β have to be determined in order to reduce the number of output alignments by using a linear weight

⋆1 http://rna.eit.hirosaki-u.ac.jp/cofolga2mo/srv/

© 2011 Information Processing Society of Japan

index I. In the present study, the parameter space to be explored is represented by a coarse grid $(\alpha,\ \beta) \in A \times B$, where $A = \{-0.1, -0.2, \cdots, -1.9, -2.0\}$ and $B = \{10, 20, \cdots, 90, 100\}$. The set of α and β at the grid point which gives the highest mean CC was adopted as the best set of α and β. The mean CC was obtained by using the best CC in the top ten alignments in each alignment list sorted by I. As a result of the optimization using 5,010 RNA sequence pairs taken from the BRAliBase 2.1 k2-dataset, we obtained $\alpha = -0.4$ and $\beta = 40$, which gives a mean CC = 0.782. In addition to the parameter determination based on the whole dataset, we performed 5-fold cross validation for the same dataset. The BRAliBase 2.1 k2-dataset was divided into five sub-datasets, and five training sub-datasets were constructed by subtracting one of the five sub-datasets from the whole dataset. The whole dataset was divided in such a way that the sequence identity and RNA type distributions of the sub-datasets are not biased. The subtracted sub-dataset was used as a test dataset for each training sub-dataset. After that, the parameter determination procedure same with that mentioned above was applied to each of five training sub-datasets and the obtained parameters were tested with a corresponding test sub-dataset. As a result, we obtained mean CCs of 0.776, 0.783, 0.777, 0.787 and 0.785 for the five test sub-datasets.

Figure 1 shows the n_{out} dependence of the mean CC for Cofolga2mo, where n_{out} is the number of output alignments and each mean CC was calculated by using the best CC in the top n_{out} alignments. In this figure, mean CCs for Foldalign 2.1.0 [8] and Dynalign 4.5 [9], which gave the highest mean CCs in our previous benchmark with the same dataset [17], are also plotted. Although the mean CC (0.782) obtained for the top ten alignments ($n_{\mathrm{out}} = 10$) is lower than that (0.801) for all alignments ($n_{\mathrm{out}} = 50$), still it is better than those of the previous mono-objective RNA sequence alignment programs (unfortunately, when we used the alignments with the best I [$n_{\mathrm{out}} = 1$], we obtained the mean CC worse than those of the mono-objective methods). Thus, by inspecting the ten alignments output by the present web server, users can obtain accurate alignments compared to those of previous mono-objective methods. It is noted that, however, inaccurate alignments can also be included in the top ten alignments of the alignment list sorted by I (i.e., there is no guarantee that all alignments in the top ten are accurate).

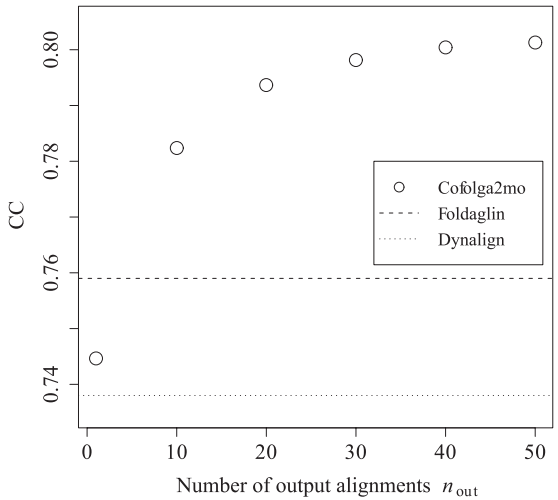

Fig. 1 The n_{out} dependence of the mean CC for Cofolga2mo, where n_{out} is the number of output alignments. Mean CCs for Foldalign 2.1.0 and Dynalign 4.5 are also plotted by dashed and dotted lines, respectively.

3.2 Sorted Alignment List

By pasting two RNA sequences into a form and clicking the submit button at the submission page of Cofolga2mo web server, a user can browse the alignment results at the 'web page for the alignment list'. **Figure 2** shows a screen shot of the web page for the alignment list output by the server. The top of the web page, a plot in s-P plane is shown. In this plot, open triangles indicate n_{out} alignments output by Cofolga2mo, where n_{out} is the number of output alignments; n_{out} can be specified by the user at the submission page. Just below the s-P plot, the sorted alignment list is tabulated. In this list, n_{out} alignments obtained by Cofolga2mo are contained in each row with its I, P and s. The alignments are sorted in descending order of a measure which was selected by the user at the submission page. At the left most column (a 'Fold' column) of the table, there is a button for browsing a consensus secondary structure; by clicking the button, the second web page is opened and the user can browse the consensus secondary structure predicted for each alignment.

© 2011 Information Processing Society of Japan

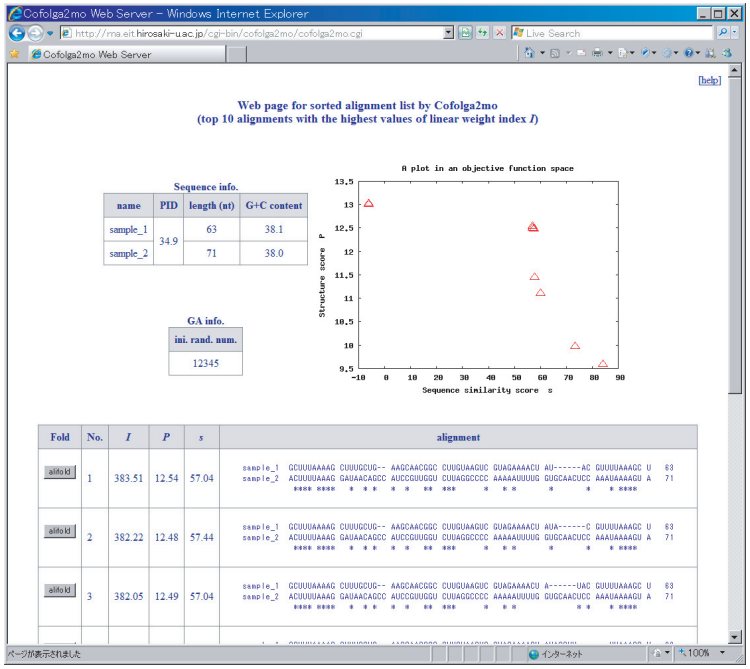

Fig. 2 A screenshot of the alignment list output by Cofolga2mo web server.

3.3 Predicted Consensus Secondary Structure

In the second web page for browsing the predicted consensus secondary structure (**Fig. 3**), the user can see a s-P plot, predicted secondary structures for two RNA sequences, and a consensus secondary structure predicted for the selected alignment. In the s-P plot shown at the top of the web page, a point corresponding to the selected alignment is plotted by a blue solid triangle in addition to the points for the other alignments denoted by open triangles. Below the s-P plot, predicted secondary structures for two RNA sequences are visualized by using VARNA. These structures are derived from the consensus secondary structure predicted for the alignment. By virtue of the various functions implemented in VARNA, the user can not only interactively examine the structures (e.g., rotation, zoom and shift of a structure), but also download the secondary structures

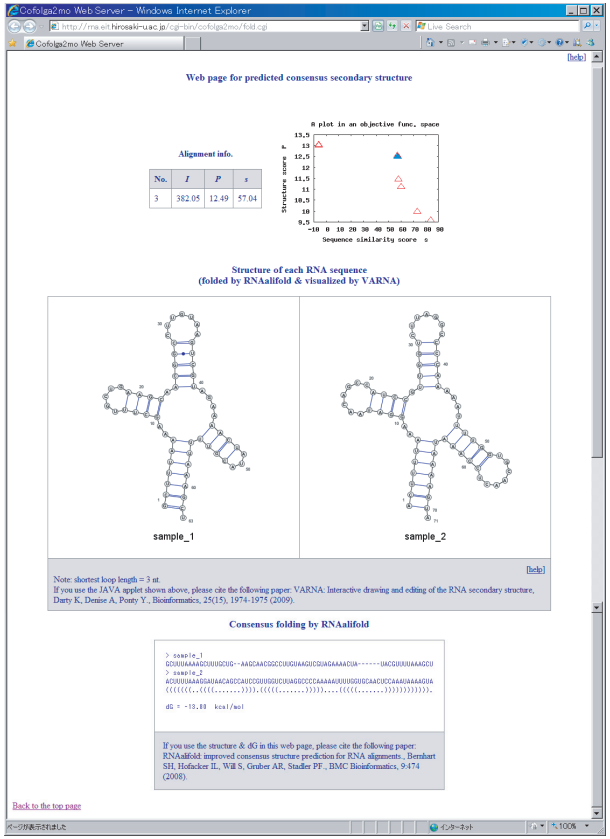

Fig. 3 A screenshot of the predicted consensus secondary structure. A user can browse the structure for the alignment selected at the alignment list. Two screenshots were concatenated for this figure.

in various formats including SVG, EPS and PNG. A base pair between the nucleotide positions i and j is eliminated from the predicted secondary structures visualized by VARNA if $|i - j| \leq 3$. At the bottom of this web page, a consensus secondary structure predicted for the alignment is shown. Accompanying with the alignment and the consensus secondary structure in bracket notation, a free energy output by RNAalifold is also displayed to assess the stability of the

© 2011 Information Processing Society of Japan

predicted consensus secondary structure.

4. Discussion

Let us consider a mono-objective RNA sequence alignment method, where we maximize an objective function $f(s, P)$ which monotonically increases with both s and P (i.e., $f(s_a, P) < f(s_b, P)$ if $s_a < s_b$ and $f(s, P_a) < f(s, P_b)$ if $P_a < P_b$). In such an optimization, the optimal solution for any $f(s, P)$ becomes one of Pareto optimal solutions. A Pareto optimal solution is defined as a solution which is not *dominated* by any other solution, where Solution A is said to *dominate* Solution B if '$s_A \geq s_B$ and $P_A \geq P_B$' and '$s_A \neq s_B$ or $P_A \neq P_B$' in the present study. According to this definition, it is guaranteed that, for any non-Pareto optimal solution, there exists a Pareto optimal solution whose value of the $f(s, P)$ is higher than that of the non-Pareto optimal solution, since the s or P of any non-Pareto optimal solution is strictly smaller than the corresponding value of a Pareto optimal solution. Thus, obtaining Pareto optimal solutions (which are included in weak Pareto optimal solutions) is equivalent to obtaining the optimal solutions for various $f(s, P)$ at once. This is the reason why accurate alignments can be included in Pareto optimal solutions. The reason for using the notion of not Pareto but weak Pareto optimal solutions in Cofolga2mo is mentioned in our previous paper [17].

In the present paper, we propose a web server for performing a global pairwise RNA sequence alignment on the basis of multi-objective genetic algorithm. Since our server can quickly give an accurate alignment even for a sequence pair with a low sequence identity [17], the present web server will be useful for users who are interested in RNA sequence alignment. In addition, we proposed a linear weight index in the present paper, by which we can reduce the maximum number of alignments output by Cofolga2mo with a small loss of accurate alignments. Cofolga2mo algorithm can be extended to multiple alignment if we use genetic operators for multiple alignment. We are now developing such a multiple alignment version of Cofolga2mo.

One known problem of the present server is a redundancy of predicted consensus structures. E.g., when we generate ten RNA sequence alignments, we can predict ten consensus structures based on the ten alignments. However, it is also possible that an exactly same structure is predicted for several different alignments. Since we cannot use the consensus structure to compare the alignments in such a situation, it is desirable that similar predicted consensus structures are detected automatically and indicated in the alignment list. Development of such a version of the web server is currently in progress. Another problem is that of common to all stochastic algorithms, i.e., Cofolga2mo can output different solutions in accordance with the value of initial random number. Therefore, we recommend users to try multiple runs with different initial random numbers if the user wants to examine the random number dependence of the results of Cofolga2mo. This function is implemented in the submission page of Cofolga2mo web server.

Acknowledgments We would like to thank the anonymous reviewers for their kind comments. This work was supported by KAKENHI (22700304).

References

1) Griffiths-Jones, S., Saini, H.K., van Dongen, S. and Enright, A.J.: miRBase: Tools for microRNA genomics, *Nucleic Acids Res.*, Vol.36, pp.D154–158 (2008).
2) Rogers, J. and Wall, R.: A mechanism for RNA splicing, *Proc. Natl. Acad. Sci. U.S.A.*, Vol.77, pp.1877–1879 (1980).
3) Brown, J.W.: The Ribonuclease P Database, *Nucleic Acids Res.*, Vol.27, p.314 (1999).
4) Sankoff, D.: Simultaneous solution of the RNA folding, alignment and protosequence problems, *SIAM J. Appl. Math.*, Vol.45, pp.810–825 (1985).
5) Tabei, Y., Tsuda, K., Kin, T. and Asai, K.: SCARNA: Fast and accurate structural alignment of RNA sequences by matching fixed-length stem fragments, *Bioinformatics*, Vol.22, pp.1723–1729 (2006).
6) Hamada, M., Sato, K., Kiryu, H., Mituyama, T. and Asai, K.: CentroidAlign: Fast and accurate aligner for structured RNAs by maximizing expected sum-of-pairs score, *Bioinformatics*, Vol.25, pp.3236–3243 (2009).
7) Will, S., Reiche, K., Hofacker, I.L., Stadler, P.F. and Backofen, R.: Inferring Noncoding RNA Families and Classes by Means of Genome-Scale Structure-Based Clustering, *PLoS Comp. Biol.*, Vol.3, p.e65 (2007).
8) Havgaard, J., Torarinsson, E. and Gorodkin, J.: Fast pairwise structural RNA alignments by pruning of the dynamical programming matrix, *PLoS Comput. Biol.*, Vol.3, pp.1896–1908 (2007).
9) Harmanci, A., Sharma, G. and Mathews, D.: Efficient pairwise RNA structure prediction using probabilistic alignment constraints in Dynalign, *BMC Bioinformatics*,

© 2011 Information Processing Society of Japan

Vol.8, p.130 (2007).
10) Kiryu, H., Tabei, Y., Kin, T. and Asai, K.: Murlet: A practical multiple alignment tool for structural RNA sequences, *Bioinformatics*, Vol.23, pp.1588–1598 (2007).
11) Lindgreen, S., Gardner, P. and Krogh, A.: MASTR: Multiple alignment and structure prediction of non-coding RNAs using simulated annealing, *Bioinformatics*, Vol.23, pp.3304–3311 (2007).
12) Taneda, A.: An efficient genetic algorithm for structural RNA pairwise alignment and its application to non-coding RNA discovery in yeast, *BMC Bioinformatics*, Vol.9, p.521 (2008).
13) Asai, K., Kiryu, H., Hamada, M., Tabei, Y., Sato, K., Matsui, H., Sakakibara, Y., Terai, G. and Mituyama, T.: Software.ncrna.org: Web servers for analyses of RNA sequences, *Nucleic Acids Res.*, Vol.36, pp.W75–78 (2008).
14) Hofacker, I.: Vienna RNA secondary structure server, *Nucleic Acids Res.*, Vol.31, pp.3429–3431 (2003).
15) Moretti, S., Wilm, A., Higgins, D.G., Xenarios, I. and Notredame, C.: R-Coffee: A web server for accurately aligning noncoding RNA sequences, *Nucleic Acids Res.*, Vol.36, pp.W10–13 (2008).
16) Havgaard, J., Lyngso, R. and Gorodkin, J.: The FOLDALIGN web server for pairwise structural RNA alignment and mutual motif search, *Nucleic Acids Res.*, Vol.33, pp.W650–653 (2005).
17) Taneda, A.: Multi-objective pairwise RNA sequence alignment, *Bioinformatics*, Vol.26, pp.2383–2390 (2010).
18) Deb, K.: *Multi-Objective Optimization using Evolutionary Algorithms*, Wiley-Interscience Series in Systems and Optimization, John Wiley & Sons, Chichester (2001).
19) Klein, R. and Eddy, S.: RSEARCH: Finding homologs of single structured RNA sequences, *BMC Bioinformatics*, Vol.4, p.44 (2003).
20) Hofacker, I., Fontana, W., Stadler, P., Bonhoeffer, L., Tacker, M. and Schuster, P.: Fast Folding and Comparison of RNA Secondary Structures, *Monatsh. Chem.*, Vol.125, pp.167–188 (1994).
21) Goldberg, D.E.: *Genetic Algorithms in Search, Optimization and Machine learning*, Addison-Wesley, New York (1987).
22) Wilm, A., Mainz, I. and Steger, G.: An enhanced RNA alignment benchmark for sequence alignment programs, *Algorithms Mol. Biol.*, Vol.1, p.19 (2006).
23) Gardner, P.P., Daub, J., Tate, J.G., Nawrocki, E.P., Kolbe, D.L., Lindgreen, S., Wilkinson, A.C., Finn, R.D., Griffiths-Jones, S., Eddy, S.R. and Bateman, A.: Rfam: Updates to the RNA families database, *Nucleic Acids Res.*, Vol.37, pp.D136–140 (2009).
24) Bernhart, S., Hofacker, I., Will, S., Gruber, A. and Stadler, P.: RNAalifold: Improved consensus structure prediction for RNA alignments, *BMC Bioinformatics*, Vol.9, p.474 (2008).
25) Darty, K., Denise, A. and Ponty, Y.: VARNA: Interactive drawing and editing of the RNA secondary structure, *Bioinformatics*, Vol.25, pp.1974–1975 (2009).

(Received September 26, 2010)
(Accepted October 18, 2010)
(Released January 25, 2011)

(Communicated by *Kengo Sato*)

Akito Taneda is an assistant professor of Graduate School of Science and Technology, Hirosaki University. He received his B.E, M.E. and Ph.D. degrees from Tohoku University in 1996, 1997 and 2000, respectively. He worked as a research associate at Hirosaki University from 2000 to 2007. His current research interests are bioinformatics, soft computing, and computer simulation of nanoscale materials. He is a member of IPSJ and JSBi.

© 2011 Information Processing Society of Japan