doi:10.1093/nar/gkn275

WAR: Webserver for aligning structural RNAs

Elfar Torarinsson[1] and Stinus Lindgreen[2,*]

[1]Division of Genetics and Bioinformatics, IBHV, Faculty of Life Sciences, University of Copenhagen, Groennegaardsvej 3, DK-1870 Frederiksberg C and [2]Center for Bioinformatics, Department of Molecular Biology, University of Copenhagen, Ole Maaloes Vej 5, DK-2200 Copenhagen N, Denmark

Received January 14, 2008; Revised April 7, 2008; Accepted April 26, 2008

ABSTRACT

We present an easy-to-use webserver that makes it possible to simultaneously use a number of state of the art methods for performing multiple alignment and secondary structure prediction for noncoding RNA sequences. This makes it possible to use the programs without having to download the code and get the programs to run. The results of all the programs are presented on a webpage and can easily be downloaded for further analysis. Additional measures are calculated for each program to make it easier to judge the individual predictions, and a consensus prediction taking all the programs into account is also calculated. This website is free and open to all users and there is no login requirement. The webserver can be found at: http://genome.ku.dk/resources/war.

INTRODUCTION

Over the past few years, different studies have shown how noncoding RNAs (ncRNA) are involved in gene expression, cell specialization, multi-drug resistance, splicing etc. in all living cells (1,2). For instance, only a small part of mammalian genomes encodes protein coding genes, but experiments have shown that a large fraction of the genomes is transcribed (3). Thus, there is potential for a large number of ncRNA transcripts, and there is computational evidence for thousands of structured RNAs in several vertebrate genomes (4–6).

This has given rise to an increased interest in ncRNAs, and since the structure of these molecules is tightly linked to their function, structure prediction methods have received much attention. Many different methods have been developed and they vary greatly in their approaches to the problem. Previously, methods for folding a single sequence by predicting the minimum free energy conformation were pursued [mfold (7), RNAfold (8)]. Today, comparative methods are the norm where a multiple alignment of a set of related RNA sequences is a part of the approach. Either the alignment is predicted alongside the consensus structure [i.e. the Sankoff-approach (9)] or an alignment is part of the input to the structure prediction [e.g. RNAalifold (10)].

The benefit of using comparative methods is that more information is available than for single sequence approaches. Many programs have been published over the past few years, and it can be difficult for a user to determine which one to use, to judge the different predictions, and sometimes even to run the programs. Since the performance of the different programs depend on many factors such as sequence length and identity, a specific program will not always perform best. Using an ensemble of programs therefore makes it easier to get a good idea of the correct result.

Here, we present an easy way to run a selection of methods and get a combined view of the predictions. The user simply inputs the sequences to be analyzed, and a selection of programs is automatically run on the dataset. The predictions are analyzed in various ways to make the output more informative for the user. The results are presented on a webpage, where one can easily download the different predictions and compare the relative performance of the individual programs. We also present a combined consensus prediction based on the results.

MATERIALS AND METHODS

The webserver for aligning structural RNAs (WAR) performs multiple alignment and secondary structure prediction on a dataset using a number of programs. The input to the server is the RNA sequences to be analyzed in Fasta format. The sequences can either be uploaded as a file or copy–pasted to a field on the webpage. The methods chosen are (in alphabetical order):

CMfinder (11): an algorithm based on expectation maximization using covariance models. Searches for RNA motifs combined with structure prediction based on both folding energy and covariation.

[*]To whom correspondence should be addressed. Tel: +45 35321287; Fax: +45 3532 1281; Email: stinus@binf.ku.dk

The authors wish it to be known that, in their opinion, the first two authors should be regarded as joint First Authors

FoldalignM (12): based on the Sankoff approach, where the partition function (13) is used to calculate basepair probability matrices for each sequence. These matrices are then aligned using progressive alignment to produce a multiple alignment and predicted consensus structure. The approach is similar to PMcomp/PMmulti (14).

LaRA (15): a mathematical approach using Lagrangian transformed relaxation. The problem of optimizing alignment and structure is formulated as an integer programming problem, and a numerical optimization approach is used.

MASTR (16): a sampling approach using Markov chain Monte Carlo in a simulated annealing framework, where both structure and alignment is optimized by making small local changes. The score combines the log-likelihood of the alignment, a covariation term and the basepair probabilities.

RNAalifold (10) + ClustalW (17): ClustalW is one of the most widely used alignment programs. It performs progressive alignment using a simple guide tree. RNAalifold predicts the structure given in an alignment using both the free energy and a covariation measure to evaluate the basepairing regions.

RNAforester (18) + RNAcast (19): RNAforester performs multiple alignment based on an input set of sequences with secondary structures. The output is thus based on structural similarities. The input set is generated using RNAcast that predicts the common shape for all sequences and the energetically best structure for each sequence.

RNASampler (20): Possible stems are found for each sequence and the stems are then aligned by comparing all pairs of sequences. A conservation score considering both structure and sequence alignment measures the quality, and a structural alignment is built. Unpaired regions are aligned using ClustalW.

All the programs perform only global alignment, except for CMfinder which is capable of performing local multiple alignment. Optionally, one can use the webserver to perform local alignment. This is done by extracting the best scoring local motif predicted by CMfinder. Then all the selected programs in the webserver are run globally, as usual, on the local region selected by CMfinder.

All the programs are run using their default settings. If the user wants to try other parameter settings, we encourage the use of the webservers and/or source code of the individual programs. There are a few limitations on the use of WAR: the user must submit at least two sequences (note that CMfinder does not work with less than three sequences), no more than 50 sequences can be submitted in one job and a sequence can be no more than 250 nucleotides in length unless the local alignment box is checked; in that case, there is no length limit.

Postprocessing of the results

For each method, the result is presented and can be easily downloaded for further use. The multiple alignment is colored using the coloraln-script from the Vienna package (8) and is shown with a barplot visualizing the conservation for each column. The consensus structure is written on top of the alignment, and the predicted basepairs are color coded to highlight canonical basepairs (i.e. Watson–Crick interactions and GU-basepairs) and compensatory mutations.

The *consensus sequence* is shown along with the consensus structure as predicted by the program (both as dot-bracket and a Postscript-file). The consensus sequence is defined using the whole range of IUPAC ambiguity characters (21) and is similar to the *most informative sequence* (22).

To quantify the quality of the alignment, the *average pairwise identity* is calculated. When a reference alignment is not known, it is not a trivial task to measure the correctness of an alignment. This is one of the reasons why so many different alignment algorithms exist. The identity measure used in the WAR server is not as such a measure of correctness, but instead a measure of the quality of the alignment. For instance, the correct alignment of highly diverged sequences will by necessity have a lower overall identity than the correct alignment of closely related sequences. However, in the current setting, a number of different methods have all been used on the same dataset and one can get an idea of the quality by comparing the pairwise identity of the alignments. If one program obtains a comparatively low pairwise identity, the alignment is probably worse. The average pairwise identity is calculated by making all the pairwise comparisons between sequences in the multiple alignment and counting the number of aligned positions that are identical. The average fraction of identities is then reported for the whole alignment.

To estimate the thermodynamical stability of the predicted structure, the *average free energy* is used. The average free energy in itself cannot be used directly as a quality measure, but by comparing the average of the different predictions the relative performance of each can be assessed. For a given prediction, we map the consensus structure to each sequence in the alignment after removing the gaps. The sequence is then folded into this specified structure and the free energy calculated using RNAeval (8). The average free energy for the entire alignment given the predicted consensus structure is then reported.

To evaluate how well the sequence alignment supports the predicted structure, we calculate a *covariation score* for each basepair, given the alignment. The proposed structure will pair up columns in the alignment, and covariation measures the amount of evidence for a basepair. This is done by calculating how often a variation in one column leads to a variation in the other. We use the measure that proved to be best in a recent study comparing different covariation measures (23) and report the average covariation score for all basepairs. Note that this measure can be negative (due to a penalty term) and greater than 1.

The quality of the predicted structure is also measured as the *average basepair probability*. For each sequence, we calculate the basepair probability matrix using RNAfold (8,13). For a single sequence, we can then find the probability of each proposed basepair by simply looking in the matrix. For two pairing columns in the alignment, the probability of a basepair is then found as the average probability for that particular basepair in each sequence.

The average probability for the whole structure is then reported.

All these results are reported on a single webpage that makes it easy to compare the different methods for both similarities and differences. Alignments and structures are easily downloaded in different formats.

A consensus prediction is made based on all the programs in the following way: T-Coffee (24) is a program that performs multiple alignment using a library of all both local and global pairwise alignments in the set. The library is extended by realigning to a third sequence and weights are calculated based on consistencies within the library. The multiple alignment is then performed progressively based on these weights.

This method can also predict a consensus multiple alignment by building the library from a number of multiple alignments instead. A single consensus multiple alignment is constructed by giving the alignments from all the programs as input to T-Coffee. The consensus structure is found by taking each ungapped sequence and mapping the predicted structure from each program onto it. If a basepair is predicted by at least 50% of the programs, it becomes a part of the consensus structure for that sequence. This gives us a consensus structure based on all the programs for each sequence in the alignment. Each sequence is then aligned to the T-Coffee alignment, making it possible to compare the consensus structure for each sequence, and the basepairs that are present in at least 50% of the sequences become part of the consensus structure for the whole alignment.

WEBSERVER

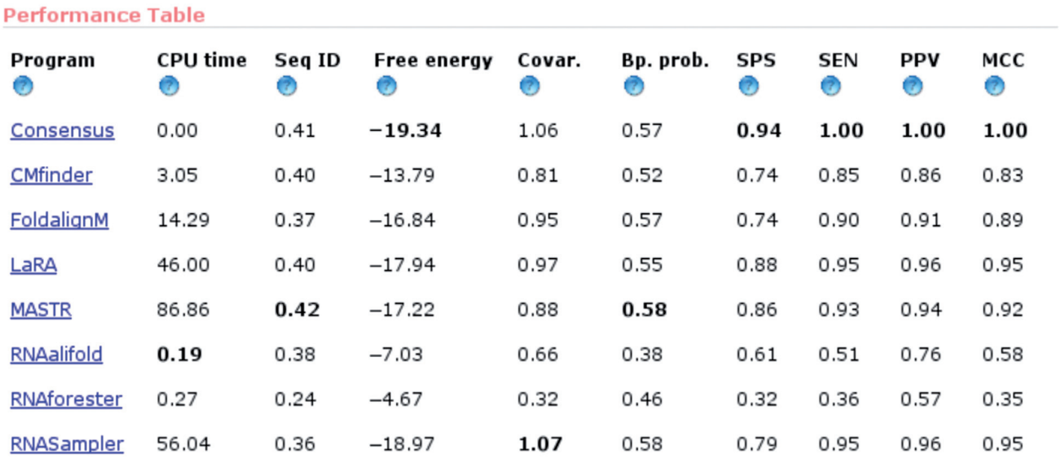
We illustrate the use of WAR in the following and especially show how the consensus prediction can be useful. For this purpose, we use a tRNA dataset consisting of 10 randomly chosen sequences, but the procedure is the same for any dataset.

If it is known, a reference alignment and structure can be uploaded along with the unaligned sequences. This makes it possible to compare the predictions to the correct answer and thus rank the methods. We use the following scores: the predicted alignment is compared to the reference alignment using the sum of pairs score (SPS), which is a sensitivity-like measure (25). It is based on the fraction of nucleotide pairs aligned in the prediction that are also present in the reference and yields a number between 0 and 1, where 1 is perfect prediction. The predicted structure is compared to the reference using Matthew's correlation coefficient (MCC), which shows the balance between sensitivity (SEN, i.e. the fraction of correct basepairs that are recovered by the method) and positive predictive value (PPV, i.e. the fraction of predicted basepairs that are also in the reference). MCC lies between −1 and 1, where 1 is perfect prediction.

Using WAR is simple: on the input form, you have to specify the input sequences either in the box or as a file. You also have to input a valid email address to receive notification when the job is complete (note that the email is used only for this notification). There are also some optional settings: you can choose to perform local alignment, you can specify a reference alignment file if available, you can name the submission and you can choose only to run a selection of the programs.

When the programs have finished, a table is shown that summarizes the predictions (Figure 1). If a reference alignment was given, the four rightmost columns summarize the performance (SPS, SEN, PPV and MCC). Otherwise, only the first six columns are shown (program, CPU time, average sequence identity, average free energy, covariation and average basepair probability). These measures are calculated as described earlier.

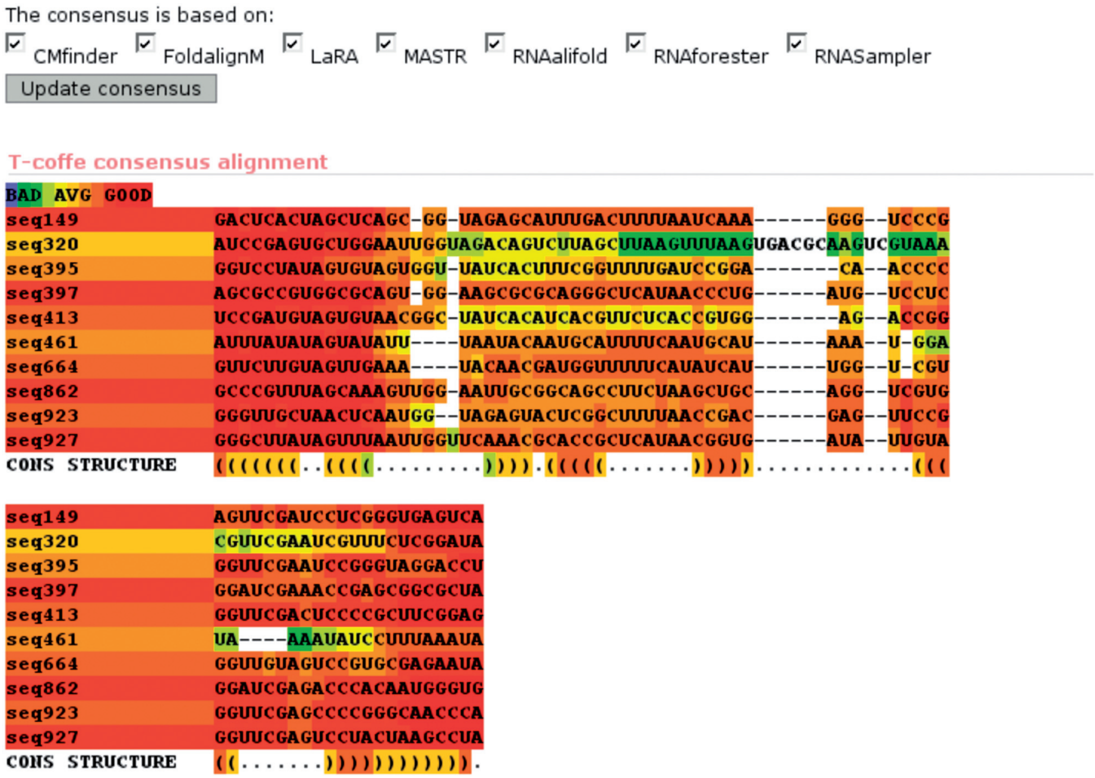
In the current example, LaRA and RNASampler gave the most correct structure predictions with $MCC = 0.95$. This corresponds with the two programs having the two lowest average free energies (−17.94 and −18.75, respectively) and the two highest average covariations (0.97 and 1.06, respectively). The average basepair probabilities are also high in both cases (0.55 and 0.58, respectively). The most correct alignment was also predicted by LaRA with $SPS = 0.88$. The measure of pairwise identity in the prediction is in this case 0.40, which is only the second largest. Notice, however, that the program with the highest pairwise identity also shows a very good SPS of 0.86.



Performance Table

Program	CPU time	Seq ID	Free energy	Covar.	Bp. prob.	SPS	SEN	PPV	MCC
Consensus	0.00	0.41	**−19.34**	1.06	0.57	**0.94**	**1.00**	**1.00**	**1.00**
CMfinder	3.05	0.40	−13.79	0.81	0.52	0.74	0.85	0.86	0.83
FoldalignM	14.29	0.37	−16.84	0.95	0.57	0.74	0.90	0.91	0.89
LaRA	46.00	0.40	−17.94	0.97	0.55	0.88	0.95	0.96	0.95
MASTR	86.86	**0.42**	−17.22	0.88	**0.58**	0.86	0.93	0.94	0.92
RNAalifold	**0.19**	0.38	−7.03	0.66	0.38	0.61	0.51	0.76	0.58
RNAforester	0.27	0.24	−4.67	0.32	0.46	0.32	0.36	0.57	0.35
RNASampler	56.04	0.36	−18.97	**1.07**	0.58	0.79	0.95	0.96	0.95

Figure 1. The table showing the results from the different programs.



The consensus is based on:

☑ CMfinder ☑ FoldalignM ☑ LaRA ☑ MASTR ☑ RNAalifold ☑ RNAforester ☑ RNASampler

T-coffe consensus alignment

```
seq149          GACUCACUAGCUCAGC-GG-UAGAGCAUUUGACUUUUAAUCAAA------GGG--UCCCG
seq320          AUCCGAGUGCUGGAAUUGGUAGACAGUCUUAGCUUAAGUUUAAGUGACGCAAGUCGUAAA
seq395          GGUCCUAUAGUGUAGUGGU-UAUCACUUUCGGUUUUGAUCCGGA-------CA--ACCCC
seq397          AGCGCCGUGGCGCAGU-GG-AAGCGCGCAGGGCUCAUAACCCUG------AUG--UCCUC
seq413          UCCGAUGUAGUGUAACGGC-UAUCACAUCACGUUCUCACCGUGG-------AG--ACCGG
seq461          AUUUAUAUAGUAUAUU----UAAUACAAUGCAUUUUCAAUGCAU------AAA--U-GGA
seq664          GUUCUUGUAGUUGAAA----UACAACGAUGGUUUUUCAUAUCAU------UGG--U-CGU
seq862          GCCCGUUUAGCAAAGUUGG-AAUUGCGGCAGCCUUCUAAGCUGC------AGG--UCGUG
seq923          GGGUUGCUAACUCAAUGG--UAGAGUACUCGGCUUUUAACCGAC------GAG--UUCCG
seq927          GGGCUUAUAGUUUAAUUGGUUCAAACGCACCGCUCAUAACGGUG------AUA--UUGUA
CONS STRUCTURE  (((((((..((((.........)))).(((((.......))))).............(((

seq149          AGUUCGAUCCUCGGGUGAGUCA
seq320          CGUUCGAAUCGUUUCUCGGAUA
seq395          GGUUCGAAUCCGGGUAGGACCU
seq397          GGAUCGAAACCGAGCGGCGCUA
seq413          GGUUCGACUCCCCGCUUCGGAG
seq461          UA----AAAUAUCCUUUAAAUA
seq664          GGUUGUAGUCCGUGCGAGAAUA
seq862          GGAUCGAGACCCACAAUGGGUG
seq923          GGUUCGAGCCCCGGGCAACCCA
seq927          GGUUCGAGUCCUACUAAGCCUA
CONS STRUCTURE  ((.......)))))))))))).
```

Figure 2. The top of the consensus result webpage showing the alignment as a heatmap.

When looking at the consensus prediction, it is evident that combining the structure predictions from all seven programs yields an improvement in both SEN, PPV and MCC to 1.0, which is better than the previous best prediction with $MCC = 0.95$. Also, the alignment is improved to $SPS = 0.94$, which is better than the single best prediction with $SPS = 0.88$. Looking at the other measures, the average free energy (−19.34) and covariation (1.06) are also best for the consensus, with the pairwise sequence identity (0.41) and basepair probability (0.57) being second best. In this case, using the consensus prediction based on the seven programs is clearly an improvement to using any single program.

Of course, that is not always the case, especially when only a few of the programs make reasonable predictions. Therefore, it is possible to update the consensus (see below) using only selected programs and not, as per default, all of them. For example, if some of the programs predict a very unstable consensus structure with a high free energy, it might be a good idea to update the consensus by removing those programs that perform poorly.

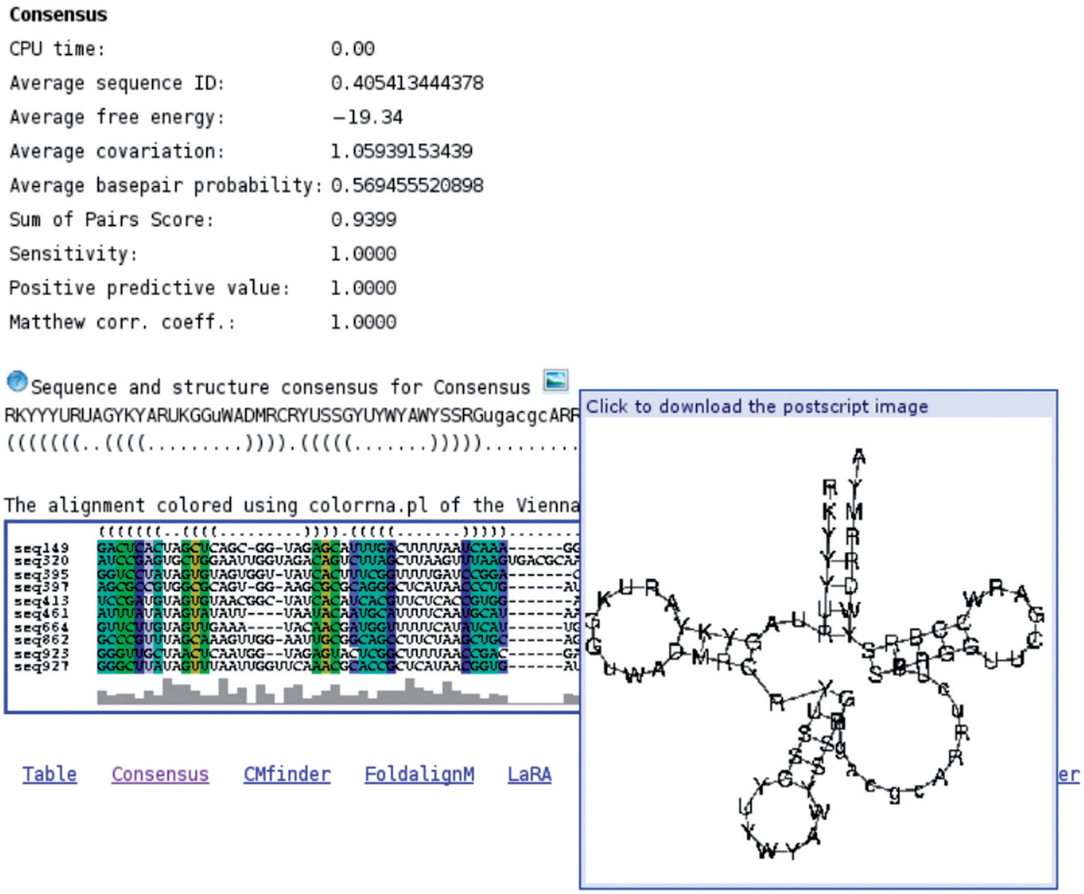
Clicking on any of the links in the first column shows a detailed description of the prediction from the chosen program. These pages are similar for all programs, only the consensus link differs. On the consensus page, the alignment is shown as a heat map at the top illustrating the confidence of the different parts of the alignment (from blue being low to red being high, see Figure 2). If necessary, one can choose which program to include in the consensus and update the alignment and structure. If some of the predictions are very different from the rest, the consensus might be improved by excluding these from the calculation.

Further down the page is additional information, which is calculated for each program (Figure 3). The measures such as average pairwise identity and—if a reference was uploaded—MCC etc. are shown. Furthermore, the predicted consensus structure is shown with the calculated consensus sequence. By pointing with the mouse at the small image-icon, the structure is shown in a small pop-up window and by clicking it is downloaded as a Postscript-file. Below the structure a color-coded image of the alignment is shown, which can be downloaded as a gif-file or a Postscript-file. The raw output data from the program can be downloaded in various formats (Fasta, Clustal, Stockholm and col) in the top right corner of the page.

The consensus prediction is particularly useful when studying unknown structural RNAs. With no prior knowledge, it is hard to judge if getting a good prediction from a single program is reliable. But if all or several of the programs agree, and the heatmap shows a reliable consensus alignment and structure, you can have higher confidence in the results.

CONCLUSION

The WAR webserver makes it easy to use a number of methods for aligning and predicting the secondary structure for a set of structural RNAs. With all the focus on ncRNAs, this is a very useful tool for any researcher who wants to analyze sequences, but does not



```
Consensus
CPU time:                      0.00
Average sequence ID:           0.405413444378
Average free energy:           −19.34
Average covariation:           1.05939153439
Average basepair probability:  0.569455520898
Sum of Pairs Score:            0.9399
Sensitivity:                   1.0000
Positive predictive value:     1.0000
Matthew corr. coeff.:          1.0000

Sequence and structure consensus for Consensus
RKYYYURUAGYKYARUKGGuWADMRCRYUSSGYUYWYAWYSSRGugacgcARR
(((((((..((((.........)))).(((((.......))))).........

The alignment colored using colorrna.pl of the Vienna
          (((((((..((((.........)))).(((((.......))))).......
seq149    GACUCACUAGCUCAGC-GG-UAGAGCAUUUGACUUUUAAUCAAA------GG
seq320    AUCCGAGUGCUGGAAUUGGUAGACAGUCUUAGCUUAAGUUUAAGUGACGCAA
seq395    GGUCCUAUAGUGUAGUGGU-UAUCACUUUCGGUUUUGAUCCGGA-------C
seq397    AGCGCCGUGGCGCAGU-GG-AAGCGCGCAGGGCUCAUAACCCUG------AU
seq413    UCCGAUGUAGUGUAACGGC-UAUCACAUCACGUUCUCACCGUGG-------A
seq461    AUUUAUAUAGUAUAUU----UAAUACAAUGCAUUUUCAAUGCAU------AA
seq664    GUUCUUGUAGUUGAAA----UACAACGAUGGUUUUCAUAUCAU------UG
seq862    GCCCGUUUAGCAAAGUUGG-AAUUGCGGCAGCCUUCUAAGCUGC------AG
seq923    GGGUUGCUAACUCAAUGG--UAGAGUACUCGGCUUUUAACCGAC------GA
seq927    GGGCUUAUAGUUUAAUUGGUUCAAACGCACCGCUCAUAACGGUG------AU
```

Figure 3. The part of the consensus result page showing the predicted structure (with the pop-up window visible), the consensus sequence and the color-coded alignment. The consensus sequence is calculated from the actual nucleotide frequencies in each alignment column compared to the background frequencies to find overrepresented nucleotides. For instance, if only nucleotide 'A' is overrepresented, an 'A' will be added to the consensus. If, on the other hand, both 'A' and 'C' are overrepresented, an 'M' is used in the consensus. If more than half the sequences contain a gap in a column, a lower case letter is used. Otherwise, we use upper case. This part of the page is similar to the result pages for the other programs.

want to get the different programs to run him/herself. In time, WAR will be extended with new methods that perform well, and other combinations of alignment tools and structure prediction tools might be pursued.

What makes WAR especially useful is the postprocessing. The different measures calculated as part of the pipeline make it easier for the user to judge the quality of both the alignment and predicted structure, as well as compare the different methods. The calculated consensus alignment and structure is a valuable indicator of the quality of the predictions, especially for new, unknown sequences, where the user can get a good idea of how trustworthy the predictions are. Finally, WAR makes it easy to download the alignments and structures, both for the individual programs and the consensus, for further analysis.

It should be stressed again that the performance of the individual methods depend strongly on the actual dataset: the RNA family, the sequence length, the overall identity—all of this will affect the performance. The strength of the WAR server is the ease with which the different methods can be compared on different RNA datasets. In the example above, the tRNA dataset is fairly divergent and will thus be hard for methods that rely on a good, purely sequence-based alignment. A set of more closely related sequences will on the other hand be a challenge to programs that need a strong evolutionary signal from covariation. The goal of the above example is not to compare the individual methods but to show how the server works. On the website, the result of a different, less divergent dataset is available as an example.

ACKNOWLEDGEMENTS

We thank Jan Gorodkin, Stefan Seemann and Jakob S. Pedersen for their comments, and Paul Gardner for some rewarding conversations. Funding to pay the Open Access publication charges for this article was provided by a faculty PhD stipend.

Conflict of interest statement. None declared.

REFERENCES

1. Bompfünewerer,A.F., Backofen,R., Bernhart,S.H., Hertel,J., Hofacker,I.L., Stadler,P.F. and Will,S. (2007) Variations on RNA folding and alignment: lessons from Benasque. *J. Math. Biol.*, **56**, 129–144.

2. Bompfünewerer,A.F., Flamm,C., Fried,C., Fritzsch,G., Hofacker,I.L., Lehmann,J., Missal,K., Mosig,A., Müller,B., Prohaska,S.J. *et al.* (2005) Evolutionary patterns of non-coding RNAs. *Theor. Biosci.*, **123**, 301–369.
3. Cheng,J., Kapranov,P., Drenkow,J., Dike,S., Brubaker,S., Patel,S., Long,J., Stern,D., Tammana,H., Helt,G. *et al.* (2005) Transcriptional maps of 10 human chromosomes at 5-nucleotide resolution. *Science*, **308**, 1149–1154.
4. Washietl,S., Hofacker,I.L., Lükasser,M., Huttenhofer,A. and Stadler,P.F. (2005) Mapping of conserved RNA secondary structures predicts thousands of functional noncoding RNAs in the human genome. *Nat. Biotechnol.*, **23**, 1383–1390.
5. Pedersen,J.S., Bejerano,G., Siepel,A., Rosenbloom,K., Lindblad-Toh,K., Lander,E.S., Kent,J., Miller,W. and Haussler,D. (2006) Identification and classification of conserved RNA secondary structures in the human genome. *PLoS Comput. Biol.*, **2**, e33.
6. Torarinsson,E., Sawera,M., Havgaard,J.H., Fredholm,M. and Gorodkin,J. (2006) Thousands of corresponding human and mouse genomic regions unalignable in primary sequence contain common RNA structure. *Genome Res.*, **16**, 885–889.
7. Zuker,M. and Stiegler,P. (1981) Optimal computer folding of large RNA sequences using thermodynamics and auxiliary information. *Nucleic Acids Res.*, **9**, 133–148.
8. Hofacker,I.L. (2003) Vienna RNA secondary structure server. *Nucleic Acids Res.*, **31**, 3429–3431.
9. Sankoff,D. (1985) Simultaneous solution of the RNA folding, alignment and protosequence problems. *SIAM J. Appl. Math.*, **45**, 810–825.
10. Hofacker,I.L., Fekete,M. and Stadler,P.F. (2002) Secondary structure prediction for aligned RNA sequences. *J. Mol. Biol.*, **319**, 1059–1066.
11. Yao,Z., Weinberg,Z. and Ruzzo,W.L. (2006) CMfinder – a covariance model based RNA motif finding algorithm. *Bioinformatics*, **22**, 445–452.
12. Torarinsson,E., Havgaard,J.H. and Gorodkin,J. (2007) Multiple structural alignment and clustering of RNA sequences. *Bioinformatics*, **23**, 926–932.
13. McCaskill,J.S. (1990) The equilibrium partition function and base pair binding probabilities for RNA secondary structure. *Biopolymers*, **29**, 1105–1119.
14. Hofacker,I.L., Bernhart,S.H.F. and Stadler,P.F. (2004) Alignment of RNA base pairing probability matrices. *Bioinformatics*, **20**, 2222–2227.
15. Bauer,M., Klau,G.W. and Reinert,K. (2007) Accurate multiple sequence-structure alignment of RNA sequences using combinatorial optimization. *BMC Bioinform.*, **8**.
16. Lindgreen,S., Gardner,P.P. and Krogh,A. (2007) MASTR: multiple alignment and structure prediction of non-coding RNAs using simulated annealing. *Bioinformatics*, **23**, 3304–3311.
17. Thompson,J.D., Higgins,D.G. and Gibson,T.J. (1994) CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Res.*, **22**, 4673–4680.
18. Höchsmann,M., Töller,T., Giegerich,R. and Kurtz,S. (2003) Local similarity of RNA secondary structures. In *Proceedings of the IEEE Bioinformatics Conference*. Stanford University, Stanford, CA, pp. 159–168.
19. Reeder,J. and Giegerich,R. (2005) Consensus shapes: an alternative to the Sankoff algorithm for RNA consensus structure prediction. *Bioinformatics*, **21**, 3516–3523.
20. Xu,X., Ji,Y. and Stormo,G.D. (2007) RNA Sampler: a new sampling based algorithm for common RNA secondary structure prediction and structural alignment. *Bioinformatics*, **23**, 1883–1891.
21. Cornish-Bowden,A. (1985) Nomenclature for incompletely specified bases in nucleic acid sequences: recommendations 1984. *Nucleic Acids Res.*, **13**, 3021–3030.
22. Freyhult,E., Moulton,V. and Gardner,P.P. (2005) Predicting RNA structure using mutual information. *Appl. Bioinform.*, **4**, 53–59.
23. Lindgreen,S., Gardner,P.P. and Krogh,A. (2006) Measuring covariation in RNA alignments: physical realism improves information measures. *Bioinformatics*, **22**, 2988–2995.
24. Notredame,C., Higgins,D. and Heringa,J. (2000) T-Coffee: a novel method for multiple sequence alignments. *J. Mol. Biol.*, **302**, 205–217.
25. Thompson,J.D., Plewniak,F. and Poch,O. (1999) A comprehensive comparison of multiple sequence alignment programs. *Nucleic Acids Res.*, **27**, 2682–2690.