


SOFTWARE

Open Access



RNA-TVcurve: a Web server for RNA secondary structure comparison based on a multi-scale similarity of its triple vector curve representation

Ying Li^{1,2} , Xiaohu Shi^{1,2}, Yanchun Liang^{1,2,3}, Juan Xie^{4,5,6}, Yu Zhang^{1,2*} and Qin Ma^{4,5,6*}

Abstract

Background: RNAs have been found to carry diverse functionalities in nature. Inferring the similarity between two given RNAs is a fundamental step to understand and interpret their functional relationship. The majority of functional RNAs show conserved secondary structures, rather than sequence conservation. Those algorithms relying on sequence-based features usually have limitations in their prediction performance. Hence, integrating RNA structure features is very critical for RNA analysis. Existing algorithms mainly fall into two categories: alignment-based and alignment-free. The alignment-free algorithms of RNA comparison usually have lower time complexity than alignment-based algorithms.

Results: An alignment-free RNA comparison algorithm was proposed, in which novel numerical representations RNA-TVcurve (triple vector curve representation) of RNA sequence and corresponding secondary structure features are provided. Then a multi-scale similarity score of two given RNAs was designed based on wavelet decomposition of their numerical representation. In support of RNA mutation and phylogenetic analysis, a web server (RNA-TVcurve) was designed based on this alignment-free RNA comparison algorithm. It provides three functional modules: 1) visualization of numerical representation of RNA secondary structure; 2) detection of single-point mutation based on secondary structure; and 3) comparison of pairwise and multiple RNA secondary structures. The inputs of the web server require RNA primary sequences, while corresponding secondary structures are optional. For the primary sequences alone, the web server can compute the secondary structures using free energy minimization algorithm in terms of RNAfold tool from Vienna RNA package.

Conclusion: RNA-TVcurve is the first integrated web server, based on an alignment-free method, to deliver a suite of RNA analysis functions, including visualization, mutation analysis and multiple RNAs structure comparison. The comparison results with two popular RNA comparison tools, RNAdpdist and RNAdistance, showcased that RNA-TVcurve can efficiently capture subtle relationships among RNAs for mutation detection and non-coding RNA classification. All the relevant results were shown in an intuitive graphical manner, and can be freely downloaded from this server. RNA-TVcurve, along with test examples and detailed documents, are available at: <http://ml.jlu.edu.cn/tvcurve/>.

Keywords: RNA-TVcurve, RNA structure comparison, Multi-scale similarity, Phylogenetic tree, Numerical representations

* Correspondence: zy26@jlu.edu.cn; qin.ma@sdstate.edu

¹College of Computer Science and Technology, Jilin University, Changchun 130012, China

⁴Department of Mathematics and Statistics, South Dakota State University, Brookings, SD 57007, USA

Full list of author information is available at the end of the article

Background

RNAs are known to carry important functionalities among diverse species, while those with similar functions are usually less conserved at sequence level in structured regions. Hence, it is critical to consider the structural information in RNA comparison, aiming to elucidate their functional and evolutionary relationship. RNA three dimensional structure mainly determine the Current methods for RNA secondary structure comparison can be generally classified into two categories, i.e., alignment-based and alignment-free.

The alignment-based methods are mostly conducted in a dynamic programming framework, relying on a string or tree representation of RNA secondary structure [1–6]. The RNA secondary structure alignments generally fall into two broad categories: Sankoff-based and Non-Sankoff RNA alignment. The Sankoff algorithm simultaneously folds and aligns two or more RNA sequences using free energy minimization [6]. Its time complexity and space complexity are $O(n^{3M})$ and $O(n^{2M})$, respectively, given M RNA sequences with lengths of n . Although this algorithm has a very good performance in prediction, its heavy computational complexity greatly limits its application.

Therefore, in order to be more practical, there are many simplifications of Sankoff algorithm [7] including but not limited to Consan [2], Dynalign [8, 9], PMcomp [10], Stemloc [11], Foldalign [12, 13], locARNA [14], SPARSE [15], MARNA [16], FoldAlignM [17], Murlet [18], CARNA [19] and RAF [20]. PMcomp performs pairwise and progressive multiple alignments based on base pairing probability matrices computed from McCaskill's algorithm [21], which simplifies Sankoff's model by predicting only a single consensus structure. LocARNA improves PMcomp using local alignment. CARNA is an improved algorithm to consider RNA pseudoknot structures based on the PMcomp model. SPARSE specifies prediction and alignment of RNAs based on their structure ensembles in quadratic time. Dynalign calculates a common structure by combining free energy minimization and comparative sequence analysis to find a low free energy structure common to two sequences without any sequence identity. Non-Sankoff algorithms separate these two processes: folding and alignments. The main methods include CMfinder [22], LARA [23], RNAdistance [4], RNAStrAt [24], RNAforester [25], SCARNA [26], gardenia [27], ERA [28] Web-Beagle [29] and RNAdist [30]. RNAdistance, RNAforester, RNAdist and RNAStrAt are all tree-based approaches. RNAdistance compares RNA secondary structures based on tree edit distance, and RNAforester calculates the pairwise- or multiple-alignment of structures based on tree alignment. RNAdist is another program based on the base pairing probabilities, which has

been included in the Vienna RNA package [31]. RNAStrAt uses a conservative edit distance and mapping between two RNA stem-loops based on a tree representation. LARA is a graph-based representation for structural alignments using an integer linear program. CMfinder is based on a covariance model (CM). Gardenia is based on the definition of the common arc-annotated supersequence. ERA is an efficient and accurate RNA secondary structure alignment tool using the sparse dynamic programming technique. In [32], a new context-aware encoding representation for RNA secondary structure named as BEAR is designed and an RNA structural alignment based on BEAR encoding is proposed by integrating a constructed substitution matrix of RNA structural elements with Needleman-Wunsch algorithm. The web server Web-Beagle [29] of RNA structural alignment based on BEAR encoding is developed. Among the above alignment-based tools, RNAdistance and RNAdist are the two most widely-used tools, which have been included in the Vienna RNA package [31]. It is noteworthy that the above alignment-based methods have an intensive requirement in time complexity.

An alignment-free algorithm is usually based on a numerical representation of sequences, which requires a novel idea and alternative way to visualize, analyze and compare DNAs, RNAs and proteins. Hence, compared to the alignment-based methods for RNA structure comparison, the studies on alignment-free algorithms are limited. In [33], various graphical representations of protein, DNA and the secondary structure of RNA were reviewed. A three-dimensional curve that constitutes a unique representation of DNA sequence was proposed by Chun-ting Zhang [34], so-called Z-curve. It is one of the representative studies in this field and has been successfully applied to many different bioinformatics areas, including identification of replication origin, proteins-coding genes, genomic island, GC content variations, and phylogenetic analysis [35–37]. As far as we known, noncoding RNAs are more conserved at structure levels [38]. It is very important to consider the structure features for alignment-free RNA comparison algorithm. The sequence and structure features of RNAs have been changed into different representations, including 2 dimensions, 3 dimensions, 4 dimensions and image [39–46]. For instance, the features of the sequence and base pairing were transformed into a linear sequence, then the Lempel-Ziv algorithm was applied to compute the similarity [45]. GraphClust is another alignment-free algorithm based on a fast graph kernel technique for clustering of local RNA secondary structures [46]. Recently, signal and image processing methods have been found useful in elucidating the similarity, and more details could be found in [47]. Alignment-free methods for RNA comparison usually have lower time complexity and less accuracy compared with alignment-based methods.

In addition, with steadily increasing numbers of known 3D RNA structures, some 3D structure alignment algorithms and their web servers including SETTER [48, 49], MultiSETTER [50, 51], R3Dalign [52, 53], Rclick [54, 55], R3D-BLAST [56] and R3D-2-MSA [57] have been developed.

We developed a web server, RNA-TVcurve, for inferring RNAs' relationship based on comparing their secondary-structures. The underlying alignment-free algorithm is summarized in the next section and more details can be found in [58]. This method has two unique features: 1) a novel numerical representation of RNA sequence and its secondary structure; and 2) a novel multi-scale similarity metric of above numerical representation based on wavelet decomposition. RNA-TVcurve has three functions: 1) visualization of RNA secondary structure; 2) detection of single-point mutation based on RNA secondary structure; and 3) construction of phylogenetic trees for multiple RNAs. Evaluations of our in-house algorithm and web server has been carried out on single point mutation identification for RNA virus and phylogenetic trees construction for non-coding RNA families. These performance comparison analyses were only conducted between RNA-TVcurve and two alignment-based programs, RNAdistance and RNAdpdist, as no other web servers based on alignment-free methods for RNA structure comparison are available in the public domain. The results showcased that RNA-TVcurve is more reliable than RNAdistance and RNAdpdist. The main results are shown using intuitive figures. All the related results and execution files can be freely downloaded at <http://ml.jlu.edu.cn/tvcurve/>.

Implementation

Methodology for pairwise RNA secondary structure comparison based on RNA-TVcurve

The details of the underlying algorithm of RNA-TVcurve can be found in [58]. Here we firstly summarize the core methodology for pairwise RNA secondary structure comparison in the following four steps (Fig. 1).

Step 1: The minimum free energy method in terms of Vienna's RNAfold [59] is used to predict the secondary structure for any given RNA sequence if users do not provide secondary structure.

Step 2: The characteristic representation of RNA secondary structure is designed based on eight symbols, in which A, C, G and U were used for the unpaired nucleotide bases, and A', C', G', and U' were used for the same bases if paired by hydrogen bonds.

Step 3: A numerical representation of RNA is constructed, named as Triple vector curves (TV-curve), by integrating RNA sequence and secondary structure features. The triple vectors are designed to represent the

eight A, T, C, G, A', U', G' and C' symbols respectively as follows:

$$\begin{aligned} (1, 1), (1, 1), (1, 1) &\Rightarrow A, (1, -1), (1, -1), (1, 1) \Rightarrow A' \\ (1, 1), (1, -1), (1, 1) &\Rightarrow U, (1, -1), (1, 1), (1, 1) \Rightarrow U' \\ (1, 1), (1, -1), (1, -1) &\Rightarrow G, (1, 1), (1, 1), (1, -1) \Rightarrow G' \\ (1, -1), (1, 1), (1, -1) &\Rightarrow C, (1, 1), (1, -1), (1, -1) \Rightarrow C' \end{aligned}$$

The schematic diagram of triple vectors representing eight letters is shown in subfigure of Step 3 in Fig. 1. There is a one-to-one mapping between the TV-curves and the RNA information of sequences and secondary structures. The TV-curve is a visualization method to represent the information of the primary and secondary structure of an RNA molecular, which can provide users another angle to display and infer the difference between RNAs, especially for the long RNAs.

Step 4: Firstly, the wavelet decomposition [60] with L -level of RNA TV-Curves are computed. After the L -level transformation, the detailed coefficients and approximation coefficients at the different resolution are obtained. L is usually taken as four by default. Next, a novel similarity metric, named multi-scale similarity, was designed to capture both the global and local properties of the constructed TV-curve based on wavelet decomposition. The Pearson correlation coefficients at different decomposition levels are calculated and the weighted sum is taken in the multi-scale similarity calculation.

Detection of maximal structure difference among all RNA single point mutants

For an RNA sequence of length N , there are three possible types of mutations for each position. For the original RNA sequence, if the original structure is not provided, the secondary structure of the original RNA sequence will be predicted with the minimum free energy algorithm. All the 3^*N single point mutated RNA sequences are constructed. For each single point mutated RNA sequence, the minimum free energy algorithm is used to predict the secondary structure. In order to identify the maximal structural mutation, the similarity between each mutated RNA sequence and original RNA sequence based RNA-TVcurve is computed. The similarity score is changed into a distance score by one minus it. The mutation with the maximal structural distance is identified. Furthermore, the histogram of the structure distance scores of all the single point mutation is provided to help users to comprehensively understand the landscape of all the single point mutation. In addition, to describe the structural stability of each position, the deleterious profile is defined as the maximum distance in structures between the wild-type and the possible single point mutants at each position. The structurally important sites can be easily identified by

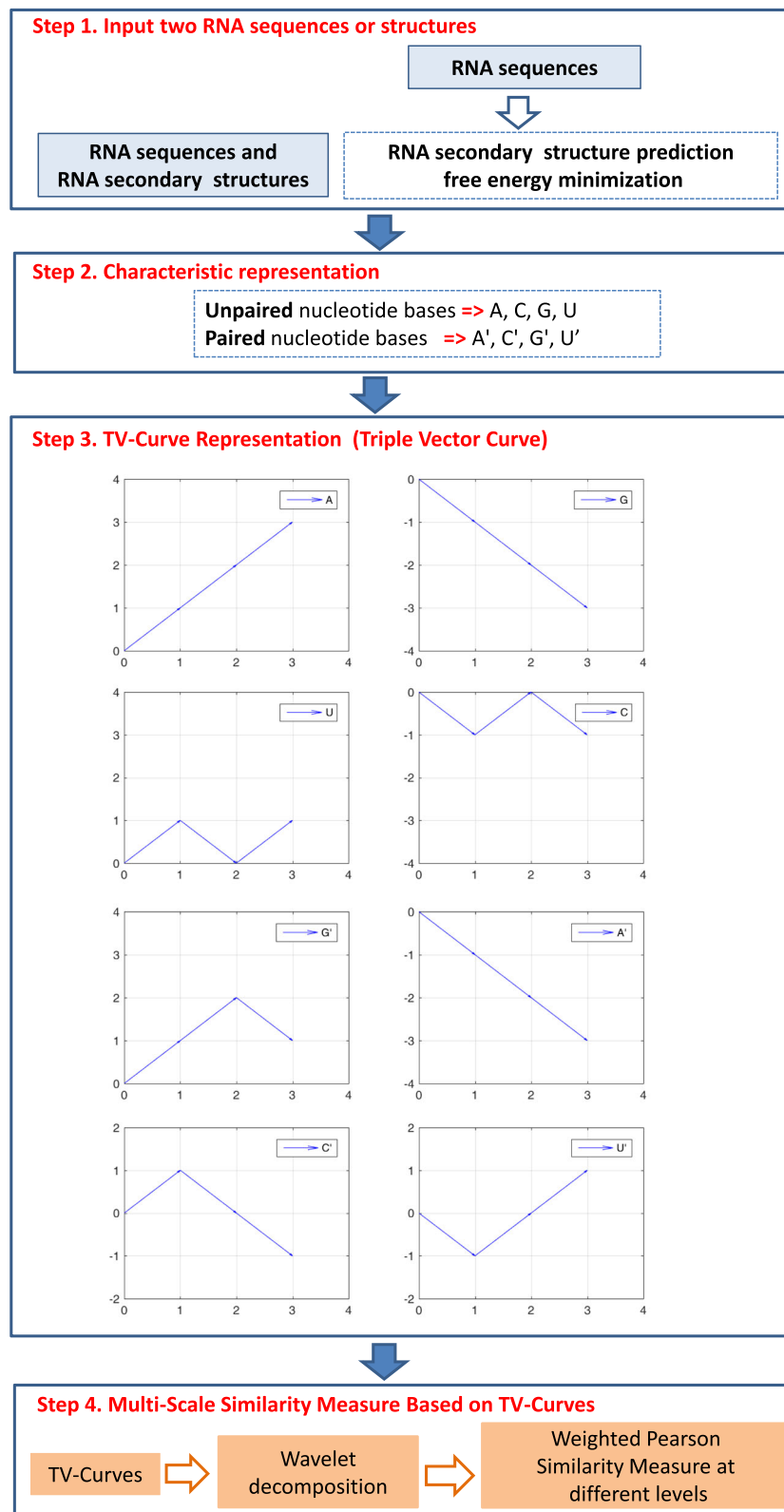


Fig. 1 Core methodology of RNA-TVcurve including four sub-block diagram: (1) RNA secondary structure prediction, (2) Characteristic representation for combining RNA sequence and secondary structure, (3) Construction of TV-Curve (Triple Vector Representation), (4) Wavelet decomposition of TV-Curves and multi-scale similarity measure based on TV-Curves

the peaks with a larger structural distance on the profile. The distributions of the distances between the wild-type and all the possible single mutants are shown as histograms. The histograms of the distances between the wild-type and all the possible single mutants provide the visual understanding on RNA mutational robustness.

Construction of RNA phylogenetic trees

The main steps of construction the RNA phylogenetic tree are listed as follows:

Step 1: Compute TV-curve for each sequence in a given RNA set.

Step 2: Calculate similarity matrix of $N(N-1)/2$ similarity between all pairs of N sequences by pairwise RNA secondary structure comparison based on RNA-TVcurve as mentioned above.

Step 3: Converting raw similarity scores to distances scores by 1-minus-similarity.

Step 4: Construct a phylogenetic tree from the distance matrix by UPGMA (Unweighted Pair Group Method with Arithmetic Mean). Firstly, cluster a pair of RNAs (taxa) with the smallest distance. Then recalculate an average distance between the new cluster and other taxa, obtaining a new distance matrix. Finally, repeat previous step until clusters converge.

Usage of Web server

RNA-TVcurve offers four different functional modules (Additional file 1: Figure S1): (1) **TVCurve**: visualization and generation of TV-curve for a provided RNA sequence; (2) **RNA Mutation**: RNA single point mutation detection for a given RNA sequence; (3) **RNA Multiple**: construction of RNA phylogenetic tree for multiple RNAs. (4) **RNA Pairwise**: perform pairwise comparison based on RNA-TVcurve. All these usages are freely available and no login information is required.

The server is accessible at: <http://ml.jlu.edu.cn/tvcurve/> with standard web browsers (Mozilla Firefox, Google Chrome, Internet Explorer, Safari). Each functional module of the web serve of RNA-TVcurve has been tested through four examples. The test samples for each module and the examples shown in this paper are listed in the server and also included in Additional file 2, aiming to help users to understand and use each of the functional modules. The input of the web server can be either entered in a text box or uploaded as an individual file from a local computer. Users can either directly run example datasets on the web server or upload their files to test different functional modules. The minimal length of the input RNA sequence for RNA-TVcurve is 10.

All functional modules require the input in FASTA format or a modified FASTA including the secondary structure in dot-bracket. Upon a job is submitted to the server, a job ID will be generated. A user can just save the job ID

and use it to extract the results. For a submitted job, its results are saved up to 1 month so that the user can query and retrieve. The probable long-running jobs include the RNA structural mutation detection for a longer RNA sequence with a length larger than 1000 nt, and the RNA phylogenetic tree construction for more than 300 RNA sequences. Some important results including the TV-curve, the mutated structure with the maximal structural difference and the phylogenetic tree are shown in a more intuitive graphical manner. The output page also offers the download of the results as a single packed file in “.zip” format for offline analysis. The README file in the packed file provides information on of the downloaded results. The detailed steps and functionality of each module are introduced in the following.

Functional module 1:TVCurve

This module provides a visualization for the TV curve constructed from provided RNA sequences and secondary structures. The input should be RNA sequences in the FASTA format that are provided with or without secondary structure information. If the structures of some RNAs are not provided, the secondary structures are predicted with the minimum free energy algorithm. The output provides the TV-curve representation and the multi-scale decomposition of the TV-curve, and the secondary structure for each RNA. All the results including TV-curve representation, the multi-scale decomposition of the TV-curve, and the secondary structure can be freely downloaded. Detailed executive process and results of the example for this module are presented in Additional file 3: Figure S2. The example dataset uses nine virus RNA sequences.

Functional module 2:RNA mutation

This module is to accomplish the RNA single point structural mutation detection. The required input for this module is a single RNA sequence and structure or a sequence alone. If multiple RNA sequences are submitted, only the first sequence is calculated. For this functional module, there is an option “Compare with RNA distance and RNAdist” to benchmark the results with both tools. This option is not selected by default. The maximum distance in structures between the wild-type and those with mutations at each position is extracted into a structural deleteriousness profile. This module provides the deleteriousness profiles and their histograms for RNA-TVcurve. The structure with the maximum structure difference calculated by RNA-TVcurve is given the corresponding mutation types and mutation sites are also provided. In addition, in order to help users to visualize the structural difference between the wild-type and all the mutants, the secondary structures of the wild-type RNA and the mutated

RNA with the maximum structural distance are also displayed on the web page. All the structural distance scores of the single point mutations detected by RNA-TVcurve are ranked, which are included in the downloaded file. If users select the option “Compare with RNA distance and RNAdist”, this module provides the deleteriousness profiles and their histograms for RNA-TVcurve, RNAdistance and RNAdist, respectively. The structure with the maximum structure difference calculated by RNA-TVcurve, RNAdistance and RNAdist are given and the corresponding mutation types and mutation sites are also provided. We will rank all the structural differences of the single point mutations detected by the three methods respectively. The detailed executive process and results of the example for this module are presented in Additional file 4: Figure S3. The input example dataset is the *Leptomonas collosoma* sequence.

Functional module 3: RNA multiple

For Module 3, the input is a set of RNA sequences and their structures or RNA sequences and part of structures or primary sequences alone (at least three RNA sequences are required). The phylogenetic tree computed by RNA-TVcurve is displayed on the web page. All the related results include the phylogenetic tree, TV curves and distance matrix using RNA-TVcurve can be freely downloaded. Similar to the module of “RNA Mutation”, users have an option “Compare with RNA distance and RNAdist”. If this option is selected, the three phylogenetic trees computed by RNA-TVcurve, RNAdist and RNAdistance are all shown in an intuitive graphical manner on the web page, which can help users make comparison and further analysis. In this case, the downloaded file includes three phylogenetic trees and distance matrix computed by RNA-TVcurve, RNAdist and RNAdistance. When the number of the sequences is larger, the alignment-based RNAdist needs more time to compute. In addition, the labels of the leaves on the tree are hard to read when the number of multiple RNAs is large. We construct the tree at horizontal direction in order to more clearly display the labels of the leaves on the tree when the number of multiple RNAs is larger than 20.

The example of construction the phylogenetic trees for multiple RNAs is 100 sequences of four non-coding RNA families including miRNA, RNaseP_arch, 5S_rRNA and tRNA. In order to make the performance for each method clearly, four types of colors are marked for each family. The phylogenetic tree by RNA-TVcurve has four noticeable branches and only four RNAs are assigned at the wrong branch, where each branch represents one RNA non-coding family. The detailed executive process and results of the example for this module are presented in Additional file 5: Figure S4.

Functional module 4: RNA pairwise

This module is to accomplish RNA pairwise comparison based on RNA-TVcurve. The input of this module includes two sets of RNA sequences. One set is named as query set and the other one as target sets. Both of the two sequence sets include RNA sequences and their structures or RNA sequences and part of structures or primary sequences alone. They should be in the FASTA format and contains the same number of RNAs. In this module, the RNAs in the query set and the corresponding RNAs in the target set are compared by RNA-TVcurve. The distance score for each pair of RNAs are shown on the web server. In addition, the RNA TV curve and secondary structure of each sequence are also displayed on the web server to help users further understand the structure difference between a pair of RNAs. All the related results are packed in a zip file for download.

The detailed executive process and results of the example for this module is presented in Additional file 6: Figure S5. The pairwise comparison is conducted between query set and target set with 3 virus sequences respectively.

Results

We assessed the performance of RNA-TVcurve compared with the two popular alignment-based programs, RNAdistance and RNAdist. In order to systematically evaluate the performance of RNA-TVcurve, two types of experiments are designed, one of which is to test the capability to infer the evolution for different species and the other is to validate the performance to distinguish the different types of RNA families.

Type I study: 5S rRNAs, RNase P and RNase MRP

The sets of different types of RNAs from different species in [45] are designed to evaluate the performance of the constructed RNA phylogenetic tree in terms of its accuracy. In order to clearly identify the evolutionary relationship among the above sets, Additional file 7: Table S1 and Table S2 provided the detailed information of family and species for the following two sets: set I from 5S Ribosomal RNA Database [61]; and set II including RNase P and RNase MRP obtained from the RNase P Database [62] and NCBI. Set II is used to test distantly related sequences, which have limited homology information. The phylogenetic tree for set I of 5S rRNAs constructed by RNA-TVcurve, RNAdistance and RNAdist is shown in Fig. 2. The phylogenetic trees constructed by RNA-TVcurve has clearly two branches for the groups of *Archaea* and *Eukaryotes*, only *Dicyema misakiense* is wrongly placed on the *Archaea* branch, however the other two phylogenetic trees constructed by RNAdistance and RNAdist do not show such good property. In addition, on the phylogenetic trees constructed by RNA-TVcurve and RNAdist, the three groups of Fungi, Metazoa and

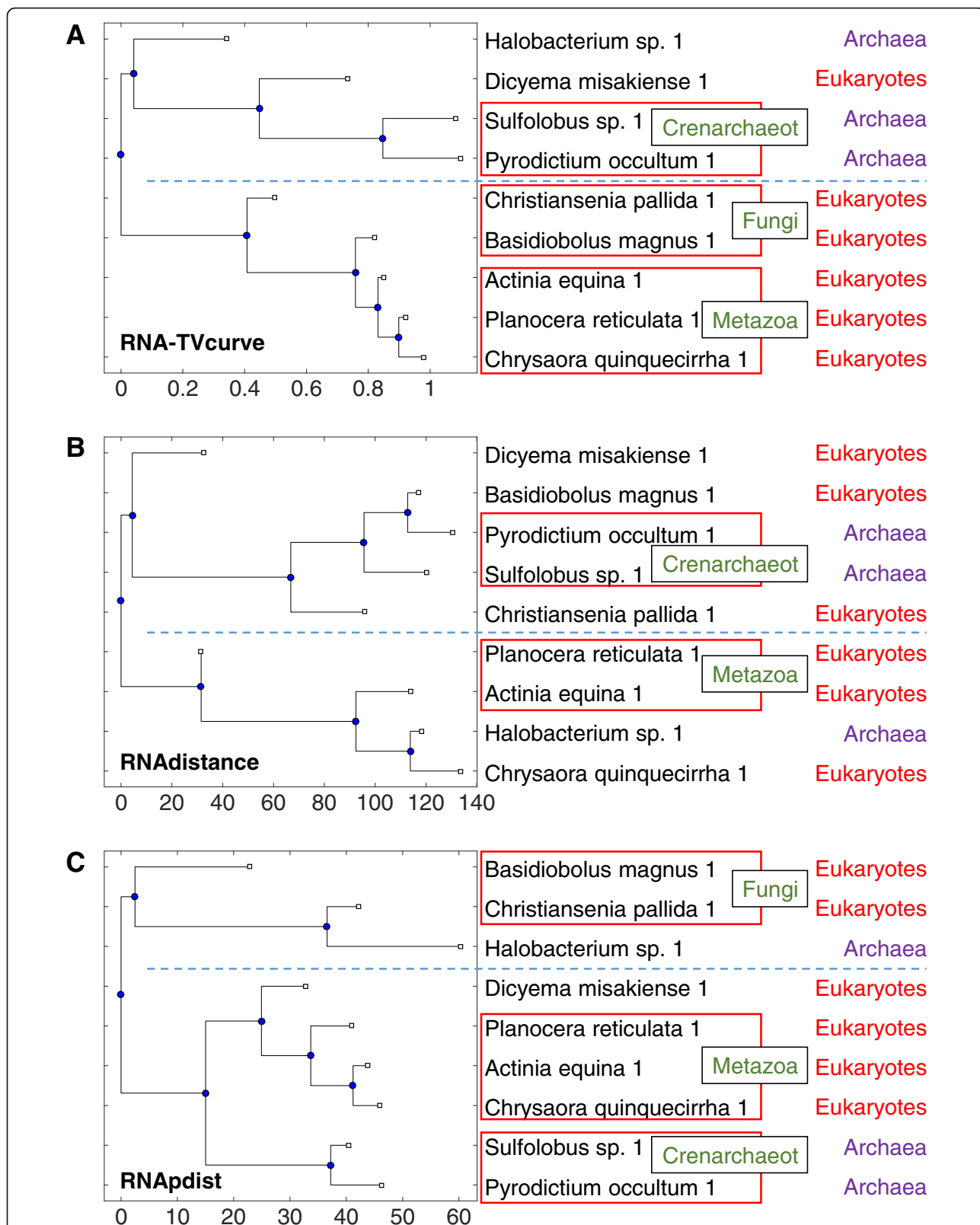


Fig. 2 The phylogenetic tree for the set I of 5S rRNAs. **a** the phylogenetic tree constructed by RNA-TVcurve, **b** the phylogenetic tree constructed by RNAdistance, **c** the phylogenetic tree constructed by RNApdist. In this figure RNAs in same the box mean they are consistent to the same known subgroup. The line is to mark the branch of the constructed tree

Crenarchaeot are all placed closely. The phylogenetic trees for set II of RNase P and RNase MRP constructed by the three programs are listed in Fig. 3. RNA-TVcurve also shows the best result compared with RNAdist and RNA-distance, in which the phylogenetic tree constructed by RNA-TVcurve has clearly two branches for the groups of

RNase MRP and RNase P, only one RNase P *M.jannaschii* is wrongly placed on the branch of RNase MRP. The groups of Alpha subdivision and Cyanobacterial are all grouped closely. The other two phylogenetic trees constructed by RNAdistance and RNAdist have limited power in distinguishing the RNase MRP and RNase P.

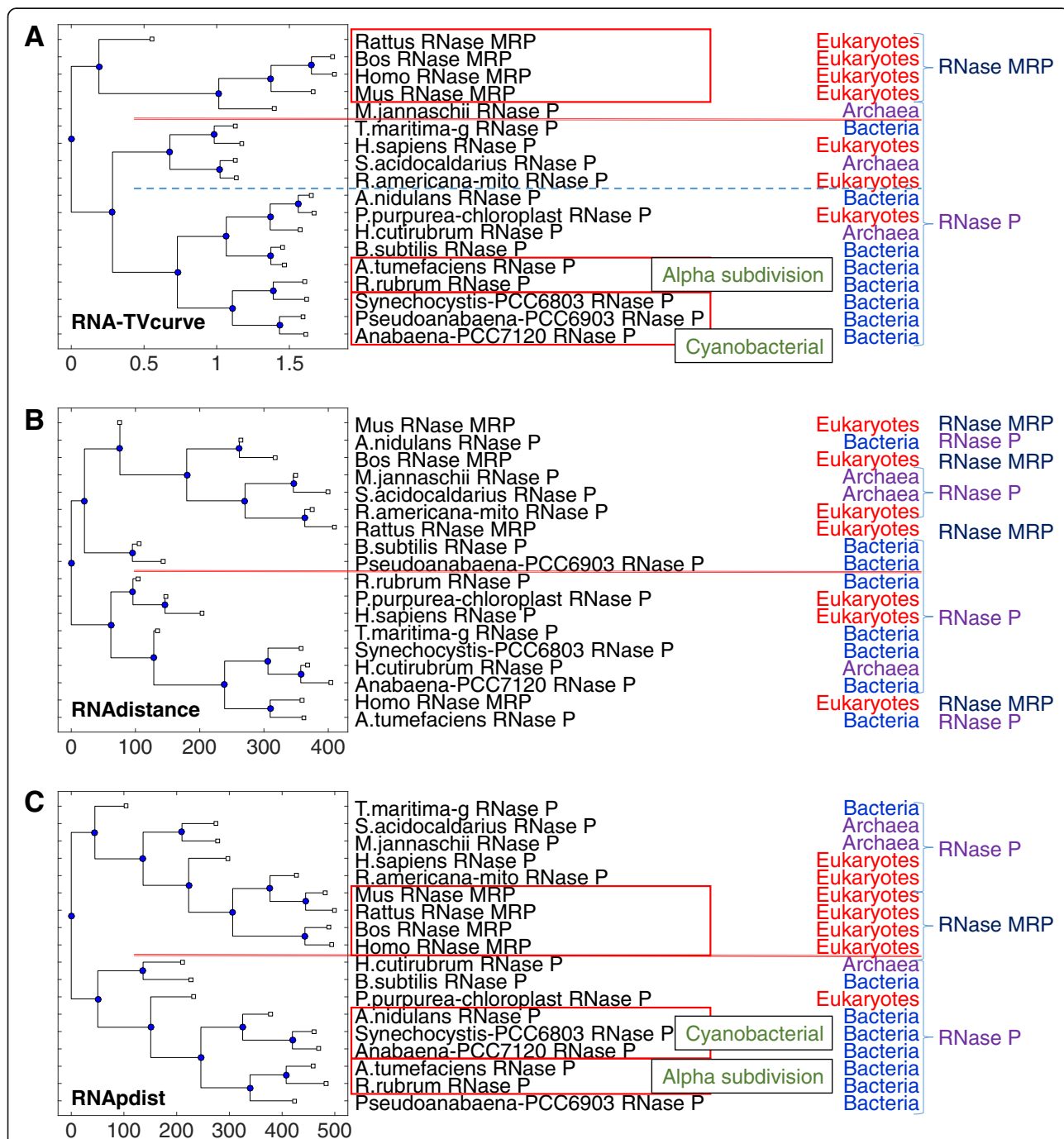
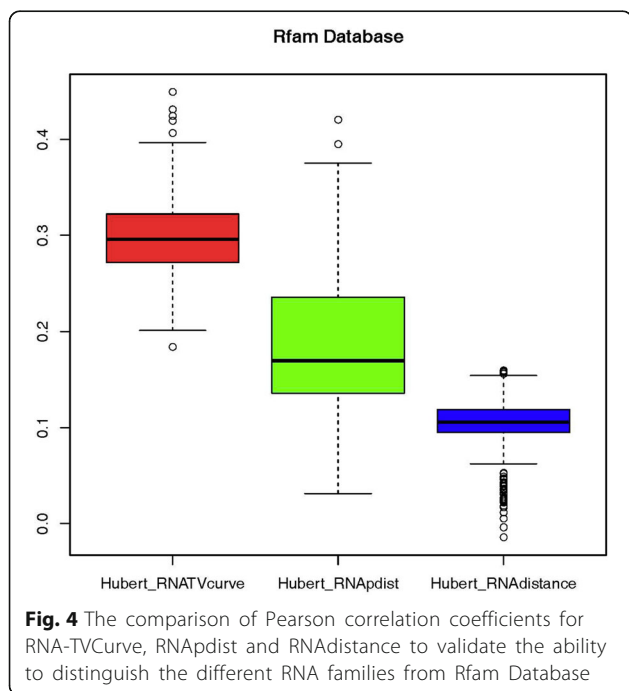


Fig. 3 The phylogenetic tree for the set II of RNase P and RNase MRP. **a** the phylogenetic tree constructed by RNA-TVcurve, **b** the phylogenetic tree constructed by RNAdistance, **c** the phylogenetic tree constructed by RNAdist. In this figure RNAs in same the box mean they are consistent to the same known subgroup. The line is to mark the branch of the constructed tree

Type II study: different RNA families from rfam database

Next, different RNA families from Rfam database are used to evaluate the ability of these three programs in distinguishing the different types of RNA families. The sequences of rRNA, tRNA, miRNA and RNase_P families are downloaded from the Rfam database. In order to directly compare the performance for distinguishing the different types of RNAs, Pearson correlation is used to measure the consistency between the similarity measure and the background co-membership matrix. It is actually the correlation between the similarity matrix computed and the background co-membership matrix [58]. Intuitively, a larger value of Pearson correlation means a higher consistency with the background matrix. Due to the large amount of sequences including 19,111 RNA sequences, it is not realistic to compute the similarities among them all. In this paper, we made 5,000 random samplings from the downloaded sequences, in which 100 sequences from each RNA family are randomly selected. Then the similarity matrices for all the families were computed by RNA-TVcurve, RNAdistance and RNApdist. The corresponding Pearson correlation coefficients were calculated for each sampling. The time complexity of RNAdistance and RNApdist are both $O(n^3)$. RNA-TVcurve dramatically reduce the computational time complexity to $O(n)$ [63], where n is the length of an RNA sequence. The comparison of Pearson correlation coefficients of RNA-TVcurve, RNAdistance and RNApdist for the 5,000 experiments is shown in Fig. 4. RNA-TVcurve is showcased significantly superior to RNAdistance and RNApdist.



Conclusions

The RNA-TVcurve server provides an access to our alignment-free RNA secondary structure comparison algorithms. It is the first integrated web server for RNA analysis based on alignment-free method, which provides a novel visualization of RNA secondary structure, mutation analysis, pairwise and multiple RNAs secondary structure comparison. This method can provide a novel view to visualize and analyze RNAs. Our evaluations suggest that the web server is a good visualization tool, and more importantly, it can be used to solve a wide range of RNA related analysis and mining problems. Particularly, it is very useful for making deleterious mutation prediction, RNA structural feature extraction and multiple RNA structural comparison. We believe the server will be a useful tool for RNA comparison and fill a void in this field.

A future improvement could be RNA-TVcurve capacity extension to handle multiple point mutations. In addition, the native secondary structure of an RNA is often a suboptimal structure not the predicted structure with minimum free energy due to limitations of thermodynamic models. Integrating multiple predicted suboptimal structures in the RNA structural similarity measurement is still very challenging and there is still a big room for improvement in its performance. In further study, we will focus on developing an RNA comparison tool based on integration of multiple suboptimal structural features. Since real structures of RNAs are very close to the structures with the minimum free energy, it is very important to consider a reasonable filter process. In addition, we will develop a new tool and a web server to search the public RNA datasets such as Human 3' UTR, Mouse 3' UTR, Human lncRNAs, Mouse lncRNAs, and Structured Rfam based on RNA-TVcurve. These computational techniques will be very useful and crucial for in-depth inference of RNA functions.

Additional files

Additional file 1: Figure S1. The web server of RNA-TVcurve including four functional modules. A) Main Home page, B) TVCurve, C) RNA Mutations, D) RNA Multiple, and E) RNA Pairwise. (PDF 361 kb)

Additional file 2: Supplementary_example_set: the fasta files of the used RNA sequences. (ZIP 8 kb)

Additional file 3: Figure S2. Schematic diagram of functional Module TVCurve: A) Select the example sequence and the operation button, B) Waiting interface: Job ID, C) Results of the module of TVCurve: results and download interface. (PDF 763 kb)

Additional file 4: Figure S3. Schematic diagram of functional Module RNA Mutation: A) Select the example sequence and the operation button, B) Waiting interface: Job ID, C) Results of the module of RNA Mutation: results and download interface. (PDF 645 kb)

Additional file 5: Figure S4. Schematic diagram of functional Module RNA Multiple: A) Select the example sequence and the operation button,

B) Waiting interface: Job ID, C) Results of the module of RNA Multiple: results and download interface. (PDF 934 kb)

Additional file 6: Figure S5. Schematic diagram of functional Module RNA Pairwise: A) Select the example sequence and the operation button, B) Results of the module of RNA Pairwise: results and download interface. (PDF 438 kb)

Additional file 7: Table S1. The information of families and species of Set 1 of 5S rRNA. **Table S2.** The information of families and species of Set 2 of RNase P and RNase MRP. (DOC 49 kb)

Abbreviation

RNA-TVcurve: RNA triple vector curve representation

Acknowledgements

We do appreciate the both anonymous reviewers and editor for valuable suggestions and constructive comments, which greatly help us to improve our manuscript and web server.

Funding

This work was supported by the National Natural Science Foundation of China (61272207, 61472158 and 61373050), and the Science-Technology Development Project from Jilin Province (20130522118JH, 20130101070JC). This work was also supported by the State of South Dakota Research Innovation Center and the Agriculture Experiment Station of South Dakota State University.

Availability of data and material

Project name: RNA-TVcurve

Project home page: <http://ml.jlu.edu.cn/tvcurve/>

Compatible browsers: Mozilla, Internet Explorer, Chrome, etc.

Operating system (s): Platform independent

Programming language: C#, HTML, Matlab

License: All copyrights of the components used belong to the legal copyright holder. Some rights reserved. For more information about this policy, please contact the authors. This work is licensed under a Creative Commons Attribution- NonCommercial- NoDerivatives 4.0 International License.

Any restriction to use by non-academics: None

Authors' contributions

YL carried out the program implementation, and participated in the computational analysis and wrote the draft manuscript. XS and YL helped the survey of related work. XJ collected the data. YZ designed and finished the web server. QM designed the project, and helped to write the manuscript. All authors read and approved the final manuscript.

Competing interests

The authors declare that they have no competing interests.

Consent for publication

Not applicable.

Ethics approval and consent to participate

Not applicable.

Author details

¹College of Computer Science and Technology, Jilin University, Changchun 130012, China. ²Key Laboratory of Symbolic Computation and Knowledge Engineering (Jilin University), Ministry of Education, Changchun 130012, China. ³Zhuhai Laboratory of Key Laboratory of Symbol Computation and Knowledge Engineering of Ministry of Education, Zhuhai College of Jilin University, Zhuhai 519041, China. ⁴Department of Mathematics and Statistics, South Dakota State University, Brookings, SD 57007, USA. ⁵Bioinformatics and Mathematical Biosciences Lab, Department of Agronomy, Horticulture and Plant Science, South Dakota State University, Brookings, SD 57007, USA. ⁶BioSNTR, Brookings, SD, USA.

Received: 25 August 2016 Accepted: 10 January 2017

Published online: 21 January 2017

References

- Gardner PP, Wilm A, Washietl S. A benchmark of multiple sequence alignment programs upon structural RNAs. *Nucleic Acids Res.* 2005;33(8):2433–9.
- Dowell RD, Eddy SR. Efficient pairwise RNA structure prediction and alignment using sequence alignment constraints. *BMC Bioinformatics.* 2006;7:400.
- Havgaard JH, Torarinsson E, Gorodkin J. Fast pairwise structural RNA alignments by pruning of the dynamical programming matrix. *PLoS Comput Biol.* 2007;3(10):1896–908.
- Shapiro BA, Zhang KZ. Comparing multiple RNA secondary structures using tree comparisons. *Comput Appl Biosci.* 1990;6(4):309–18.
- Allali J, Sagot MF. A new distance for high level RNA secondary structure comparison. *IEEE/ACM Trans Comput Biol Bioinform.* 2005;2(1):3–14.
- Sankoff D. Simultaneous solution of the RNA folding, alignment and protosequence problems. *SIAM J Appl Math.* 1985;45(5):810–25.
- Chatzou M, et al. Multiple sequence alignment modeling: methods and applications. *Brief Bioinform.* 2016;17(6):1009–23.
- Mathews DH, Turner DH. Dynalign: an algorithm for finding the secondary structure common to two RNA sequences. *J Mol Biol.* 2002;317(2):191–203.
- Mathews DH. Predicting a set of minimal free energy RNA secondary structures common to two sequences. *Bioinformatics.* 2005;21(10):2246–53.
- Hofacker IL, Bernhart SH, Stadler PF. Alignment of RNA base pairing probability matrices. *Bioinformatics.* 2004;20(14):2222–7.
- Holmes I. Accelerated probabilistic inference of RNA structure evolution. *BMC Bioinformatics.* 2005;6:73.
- Gorodkin J, Heyer LJ, Stormo GD. Finding the most significant common sequence and structure motifs in a set of RNA sequences. *Nucleic Acids Res.* 1997;25(18):3724–32.
- Havgaard JH, Lyngsø RB, Stormo GD, Gorodkin J. Pairwise local structural alignment of RNA sequences with sequence similarity less than 40%. *Bioinformatics.* 2005;21(9):1815–24.
- Will S, Reiche K, Hofacker IL, Stadler PF, Backofen R. Inferring noncoding RNA families and classes by means of genome-scale structure-based clustering. *PLoS Comput Biol.* 2007;3(4):e65.
- Will S, Otto C, Miladi M, Mohl M, Backofen R. SPARSE: quadratic time simultaneous alignment and folding of RNAs without sequence-based heuristics. *Bioinformatics.* 2015;31(15):2489–96.
- Siebert S, Backofen R. MARNA: multiple alignment and consensus structure prediction of RNAs based on sequence structure comparisons. *Bioinformatics.* 2005;21(16):3352–9.
- Torarinsson E, Havgaard JH, Gorodkin J. Multiple structural alignment and clustering of RNA sequences. *Bioinformatics.* 2007;23(8):926–32.
- Kiryu H, Tabei Y, Kin T, Asai K. Muret: a practical multiple alignment tool for structural RNA sequences. *Bioinformatics.* 2007;23(13):1588–98.
- Sorescu DA, Mohl M, Mann M, Backofen R, Will S. CARNAL—alignment of RNA structure ensembles. *Nucleic Acids Res.* 2012;40(Web Server issue):W49–53.
- Do CB, Foo CS, Batzoglou S. A max-margin model for efficient simultaneous alignment and folding of RNA sequences. *Bioinformatics.* 2008;24(13):i68–76.
- McCaskill JS. The equilibrium partition function and base pair binding probabilities for RNA secondary structure. *Biopolymers.* 1990;29(6–7):1105–19.
- Yao Z, Weinberg Z, Ruzzo WL. CMfinder—a covariance model based RNA motif finding algorithm. *Bioinformatics.* 2006;22(4):445–52.
- Bauer M, Klau GW, Reinert K. Accurate multiple sequence-structure alignment of RNA sequences using combinatorial optimization. *BMC Bioinformatics.* 2007;8:271.
- Guignon V, Chauve C, Hamel S. An edit distance between RNA stem-loops, in string processing and information retrieval: 12th International Conference, SPIRE 2005, Buenos Aires, Argentina, November 2–4, 2005. Proceedings. Edited by Consens M, Navarro G. Springer Berlin Heidelberg; 2005:335–47.
- Hochsmann M, Toller T, Giegerich R, Kurtz S. Local similarity in RNA secondary structures. *Proc IEEE Comput Soc Bioinform Conf.* 2003;2:159–68.
- Tabei Y, Tsuda K, Kin T, Asai K. SCARNA: fast and accurate structural alignment of RNA sequences by matching fixed-length stem fragments. *Bioinformatics.* 2006;22(14):1723–9.
- Blin G, Denise A, Dulucq S, Herrbach C, Touzet H. Alignments of RNA structures. *IEEE/ACM Trans Comput Biol Bioinform.* 2010;7(2):309–22.
- Zhong C, Zhang S. Efficient alignment of RNA secondary structures using sparse dynamic programming. *BMC Bioinformatics.* 2013;14:269.
- Mattei E, Pietrosanto M, Ferre F, Helmer-Citterich M. Web-Beagle: a web server for the alignment of RNA secondary structures. *Nucleic Acids Res.* 2015;43(W1):W493–7.

30. Hofacker IL, Fontana W, Stadler PF, Bonhoeffer LS, Tacker M, Schuster P. Fast folding and comparison of RNA secondary structures. *Monatshefte für Chemie/Chemical Monthly*. 1994;125:167–88.
31. Lorenz R, Bernhart SH, Honer Zu Siederdisen C, Tafer H, Flamm C, Stadler PF, Hofacker IL. ViennaRNA Package 2.0. *Algorithms Mol Biol*. 2011;6:26.
32. Mattei E, Ausiello G, Ferre F, Helmer-Citterich M. A novel approach to represent and compare RNA secondary structures. *Nucleic Acids Res*. 2014;42(10):6146–57.
33. Randić M, Zupan J, Balaban AT, Vikić-Topić D, Plavšić D. Graphical representation of proteins. *Chem Rev*. 2011;111(2):790–862.
34. Zhang R, Zhang CT. Z curves, an intuitive tool for visualizing and analyzing the DNA sequences. *J Biomol Struct Dyn*. 1994;11(4):767–82.
35. Hua ZG, Lin Y, Yuan YZ, Yang DC, Wei W, Guo FB. ZCURVE 3.0: identify prokaryotic genes with higher accuracy as well as automatically and accurately select essential genes. *Nucleic Acids Res*. 2015;43(W1):W85–90.
36. Wei W, Gao F, Du M-Z, Hua H-L, Wang J, Guo F-B. Zisland Explorer: detect genomic islands by combining homogeneity and heterogeneity properties. *Brief Bioinform*. 2016.
37. Zhang R, Zhang CT. A Brief Review: The Z-curve Theory and its Application in Genome Analysis. *Curr Genomics*. 2014;15(2):78–94.
38. Mattick JS, Makunin IV. Non-coding RNA. *Hum Mol Genet*. 2006;15 Spec No 1:R17–29.
39. Randić M, Basak SC. Characterization of DNA primary sequences based on the average distances between bases. *J Chem Inf Comput Sci*. 2001;41(3):561–8.
40. Randić M, Vrakoc M, Lers N, Plavšić D. Analysis of similarity/dissimilarity of DNA sequences based on novel 2-D graphical representation. *Chem Phys Lett*. 2003;371:202–7.
41. Guo XF, Nandy A. Numerical characterization of DNA sequences in a 2-D graphical representation scheme of low degeneracy. *Chem Phys Lett*. 2003;369(3–4):361–6.
42. Zupan J, Randić M. Algorithm for coding DNA sequences into “spectrum-like” and “zigzag” representations. *J Chem Inf Model*. 2005;45(2):309–13.
43. Liao B, Wang TM. 3-D graphical representation of DNA sequences and their numerical characterization. *Journal of Molecular Structure-Theochem*. 2004;681(1–3):209–12.
44. Gan HH, Pasquali S, Schlick T. Exploring the repertoire of RNA secondary motifs using graph theory; implications for RNA design. *Nucleic Acids Res*. 2003;31(11):2926–43.
45. Liu N, Wang T. A method for rapid similarity analysis of RNA secondary structures. *BMC Bioinformatics*. 2006;7:493.
46. Heyne S, Costa F, Rose D, Backofen R. GraphClust: alignment-free structural clustering of local RNA secondary structures. *Bioinformatics*. 2012;28(12):i224–32.
47. Almeida JS. Sequence analysis by iterated maps, a review. *Brief Bioinform*. 2014;15(3):369–75.
48. Hoksza D, Svozil D. Efficient RNA pairwise structure comparison by SETTER method. *Bioinformatics*. 2012;28(14):1858–64.
49. Cech P, Svozil D, Hoksza D. SETTER: web server for RNA structure comparison. *Nucleic Acids Res*. 2012;40(Web Server issue):W42–8.
50. Hoksza D, Svozil D. Multiple 3D RNA structure superposition using neighbor Jjoining. *IEEE/ACM IEEE/ACM Trans Comput Biol Bioinform*. 2015;12(3):520–30.
51. Cech P, Hoksza D, Svozil D. MultiSETTER: web server for multiple RNA structure comparison. *BMC Bioinformatics*. 2015;16:253.
52. Rahrig RR, Leontis NB, Zirbel CL. R3D Align: global pairwise alignment of RNA 3D structures using local superpositions. *Bioinformatics*. 2010;26(21):2689–97.
53. Rahrig RR, Petrov AI, Leontis NB, Zirbel CL. R3D Align web server for global nucleotide to nucleotide alignments of RNA 3D structures. *Nucleic Acids Res*. 2013;41(Web Server issue):W15–21.
54. Nguyen MN, Tan KP, Madhusudhan MS. CLICK—topology-independent comparison of biomolecular 3D structures. *Nucleic Acids Res*. 2011;39(Web Server issue):W24–8.
55. Nguyen MN, Verma C. Rclick: a web server for comparison of RNA 3D structures. *Bioinformatics*. 2015;31(6):966–8.
56. Liu YC, Yang CH, Chen KT, Wang JR, Cheng ML, Chung JC, Chiu HT, Lu CL. R3D-BLAST: a search tool for similar RNA 3D substructures. *Nucleic Acids Res*. 2011;39(Web Server issue):W45–9.
57. Cannone JJ, Sweeney BA, Petrov AI, Gutell RR, Zirbel CL, Leontis N. R3D-2-MSA: the RNA 3D structure-to-multiple sequence alignment server. *Nucleic Acids Res*. 2015;43(W1):W15–23.
58. Li Y, Duan M, Liang Y. Multi-scale RNA comparison based on RNA triple vector curve representation. *BMC Bioinformatics*. 2012;13:280.
59. Hofacker IL. Vienna RNA secondary structure server. *Nucleic Acids Res*. 2003;31(13):3429–31.
60. Unser M, Aldroubi A. A review of wavelets in biomedical applications. *Proc IEEE*. 1996;84(4):626–38.
61. Szymanski M, Barciszewska MZ, Erdmann VA, Barciszewski J. 5S Ribosomal RNA database. *Nucleic Acids Res*. 2002;30(1):176–8.
62. Brown JW. The Ribonuclease P Database. *Nucleic Acids Res*. 1999;27(1):314.
63. Mallat S. A Theory for multiresolution signal decomposition: the wavelet representation. *IEEE Trans Pattern Anal Mach Intell*. 1989;11(5):674–93.

Submit your next manuscript to BioMed Central and we will help you at every step:

- We accept pre-submission inquiries
- Our selector tool helps you to find the most relevant journal
- We provide round the clock customer support
- Convenient online submission
- Thorough peer review
- Inclusion in PubMed and all major indexing services
- Maximum visibility for your research

Submit your manuscript at
www.biomedcentral.com/submit

