BIOINFORMATICS APPLICATIONS NOTE Vol. 28 no. 20 2012, pages 2696–2697
doi:10.1093/bioinformatics/bts506

*Structural bioinformatics*

Advance Access publication August 24, 2012

# TRFolder-W: a web server for telomerase RNA structure prediction in yeast genomes

Dong Zhang[1,†], Xingran Xue[2,†], Russell L. Malmberg[1,3] and Liming Cai[1,2,*]

[1]Institute of Bioinformatics and [2]Department of Computer Science and [3]Department of Plant Biology, University of Georgia, Athens, GA 30602, USA

Associate Editor: Anna Tramontano

## ABSTRACT

**Summary:** TRFolder-W is a web server capable of predicting core structures of telomerase RNA (TR) in yeast genomes. TRFolder is a command-line Python toolkit for TR-specific structure prediction. We developed a web-version built on the django web framework, leveraging the work done previously, to include enhancements to increase flexibility of usage. To date, there are five core sub-structures commonly found in TR of fungal species, which are the template region, downstream pseudoknot, boundary element, core-closing stem and triple helix. The aim of TRFolder-W is to use the five core structures as fundamental units to predict potential TR genes for yeast, and to provide a user-friendly interface. Moreover, the application of TRFolder-W can be extended to predict the characteristic structure on species other than fungal species.

**Availability:** The web server TRFolder-W is available at http://rna-informatics.uga.edu/?f=software&p=TRFolder-w.

**Contact**: cai@cs.uga.edu

**Supplementary information:** http://rna-informatics.uga.edu/?f=software&p=TRFolder-w-supplement.

Received on May 6, 2012; revised on August 8, 2012; accepted on August 9, 2012

## 1 INTRODUCTION

The ability of telomerase to repair telomeres increases the cell-division cycles of a lineage, each of which otherwise shortens the telomeres. Telomerase RNA (TR) is one fundamental component of telomerase. Secondary structure prediction facilitates identifying TR genes which are highly divergent in their sequences. Compared to the sequences, the secondary structures of TR are much more conserved. There are few existing computational tools that can accurately predict TR structures; the difficulties are mainly caused by the extensive variance in TR sequences and stem-loop length, and structures such as the pseudoknot.

In our previous work, TRFolder showed a high accuracy for known yeast TR structures folding, and gave predictions with high confidence for unknown yeast TR structures. The reported limitations of usage of TRFolder mainly focus on installation troubleshooting and the template position requirement. Therefore, we developed a web version of TRFolder capable of pseudoknot prediction without input of the template position.

## 2 FUNCTIONALITIES OF THE SYSTEM

TRFolder-W accepts a FASTA file as input which allows multiple sequences. The template position for each sequence can be provided in the identifier line. If the template position is unknown, TRFolder-W will predict just the pseudoknot. The system has the following functionalities.

(1) *Predicting pseudoknot structure*: 5′ pseudoknot structures close to the template sequence are extensively found in fungi. TRFolder-W starts with pseudoknot searching, and is able to target all the pseudoknot-like structures assembled from two potential stem-loops.

(2) *Predicting triple helix structure*: a triple helix motif within the pseudoknot, which is mainly U-A pairs, has been identified in *Kluyveromyces*. This is implemented in TRFolder as a filtering functionality; triple helix folding will significantly reduce the number of pseudoknot-like candidates.

(3) *Predicting the boundary element*: The boundary element is a stem-loop structure located at upstream of the template sequence. For a given position of the template, TRFolder-W will predict a boundary element.

(4) *Predicting core closing stem*: The core closing stem is a long-range base-pair region containing the bounding boundary element, template and pseudoknot. For each combination of boundary element and pseudoknot, its enclosing sequence will be further folded to shape a core closing stem.

(5) *Adjusting the stem-loop scan window size*: A pseudoknot structure is shaped by two crossing and non-overlapping stems. The window size for stem searching can be adjusted by users. The current window size was trained from known yeast pseudoknots.

(6) *Incorporating user's training data*: A logodds matrix, representing the frequency of pairing between each of two bases (Guo *et al.*, 2011), is applied to measure the likelihood of structure predictions. For each core structure, one $5 \times 5$ logodds matrix was trained on the basis of known yeast TR structures. TRFolder-W allows the user to use other scoring matrices instead.

*To whom correspondence should be addressed.

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

(7) *Restricting the hits shown number*: The number of top hits to be displayed can be selected by the user.

## 3 INTERFACE FEATURES

Users can use either text-based input or local file uploading to submit their dataset. Since users may not have training data, we organized parameter settings into two panels to avoid confusion. The basic panel provides a set of parameters optimized to predict all known TR structure in yeast, which could also be subsequently applied to other species. Since the template position is also usually unknown, the panel allows users to optionally search boundary element and core closing stem. The default stem-loop scan window size and logodds matrices were trained from known TR structures. The advanced panel can be used by users who have training data. Using modified settings, the prediction result can be significantly changed. After submission of data, TRFolder-W will display an estimate of the running time used in the searching process and redirect to a result page once the search is done. It also gives an explicit URL to the result page that the user can check back for search results. The folding structures make use of bracket view which is commonly used in RNA folding. For each query sequence, the top 10 scored candidates will be listed by default and the user can select a different number.

## 4 PERFORMANCE SUMMARY

TRFolder-W successfully confirmed TR structures of *Saccharomyces* and *Kluyveromyces* which have been published. To date, TRs of five *Saccharomyces and Kluyveromyces* (Lin *et al*., 2004; Shefer *et al*., 2007; Tzfati *et al*., 2000) that are *Kluyveromyces lactis*, *Kluyveromyces dobzhanskii*, *Kluyveromyces aestuarii*, *Kluyveromyces nonfermentans, Kluyveromyces wickerhamii*, *Saccharomyces cerevisiae*, *Saccharomyces kudriavzevii, Saccharomyces cariocanus*, *Saccharomyces paradoxus* and *Saccharomyces mikatas* have been proposed in multiple literatures. Except for *K. wickerhamii* which has a different stem 1 of the pseudoknot, TRFolder-W is capable of predicting highly similar TR core structures to published ones. The minor variances are mainly attributed to published structures having many non-canonical base pairs such as U-U and bulges. These limitations might be overcome by simply increasing the number of training data.

In addition to verifying published TRs, TRFolder-W was applied to predict TR structures for six novel yeast species, which are *Candida glabrata*, *Ashbya gossypii*, *Candida albicans*, *Candida guilliermondii*, *Pichia stipitis* and *Delphinium hansenii*. In most of these species, the TR genes recently identified by Gunisova *et al*. (2009) were also confirmed by our independent analysis. The typical signals of potentially correct TR structures consist of pseudoknot candidates close to putative templates, good stem-loop structures and TR core structures within the range of TR genes. Using TRFolder-W, we were able to identify template-close pseudoknot candidates in *C. albicans, A. gossypii, P. stipiti* and *D. hansenii*, all of which have long stem-loop for TR core structures. In *C. guilliermondii* and *C. albicans*, the structure predictions fall within the mapped genes. The pseudoknot and triple helix predicted in *C. glabrata* is highly similar to those recently proposed by Kachouri-Lafond *et al*. (2009).

TRFolder-W was used to predict TR core structures for a novel yeast species *Candida parapsilosis* which was not included in our previous studies (Guo *et al*., 2011). The telomerase RNA gene sequence and telomeric repeat sequence for *C. parapsilosis* have been available in the NCBI and Telomerase Database (http://telomerase.asu.edu/) individually. TRFolder-W is able to predict top ranked template-close pseudoknot, triple helix and boundary element candidates, all of which have the stem-loop structure and distance to template fitting within the length profile defined by published TR structures in *Saccharomyces* and *Kluyveromyces* (Supplementary Table S1). TRFolder-W also detected a core-closing stem candidate with 7 bp stem length outside *C. parapsilosis* RNA gene by searching its 2 kb surrounding genome sequence.

As a model organism, *Caenorhabditis elegans* telomerase RNA has not yet been reported. We used TRFolder-W on segments of *C. elegans* DNA and found 11 potential telomerase RNA structures for the further experimental validation. This indicates TRFolder-W may be applicable to species other than fungal species to reduce a set of candidate telomerase RNA regions to a smaller set.

TRFolder-W employs an exhaustive search approach in time $O(W^3N)$, where W is the size of the stem-loop scanning window and N is the length of the sequence, to predict TR core structures. W is a parameter that can be set with a known range, but it could be as long as N. Since this is computationally expensive and time consuming, for each species, a 4 kb RNA sequence cut from the genome surrounding the template was usually used to do test. TRFolder-W uses the features of TR structures to provide a means to identify TR genes and can be used as an alternative TR-specific structure prediction toolkit.

*Funding*: The National Institute of Health (BISTI R01GM072080-01A1, GM 61645) and in part by the National Science Foundation (IIS 0916250).

*Conflict of Interest*: none declared.

## REFERENCES

Gunisova,S. *et al*. (2009) Identification and comparative analysis of telomerase RNAs from Candida species reveal conservation of functional elements. *RNA*, **15**, 546–559.

Guo,L. *et al*. (2011) TRFolder: computational prediction of novel telomerase RNA structures in yeast genomes. *Int. J. Bioinform. Res. Appl*., **7**, 63–81.

Kachouri-Lafond,R. *et al*. (2009) Large telomerase RNA, telomere length heterogeneity and escape from senescence in *Candida glabrata*. *FEBS Lett*. **583**, 3605–3610.

Lin,J. *et al*. (2004) A universal telomerase RNA core structure includes structured motifs required for binding the telomerase reverse transcriptase protein. *Proc. Natl. Acad. Sci. USA*., **101**, 14713–14718.

Shefer,K. *et al*. (2007) A triple helix within a pseudoknot is a conserved and essential element of telomerase RNA. *Mol. Cell. Biol*., **27**, 2130–2143.

Tzfati,Y. *et al*. (2000) Template boundary in a yeast telomerase specified by RNA structure. *Science*, **288**, 863–867.