Genome analysis

RNATOPS-W: a web server for RNA structure searches of genomes

Yingfeng Wang¹, Zhibin Huang¹, Yong Wu¹, Russell L. Malmberg²,³ and Liming Cai¹,³,∗

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

Received on November 5, 2008; revised on February 15, 2009; accepted on February 16, 2009
Advance Access publication March 5, 2009
Associate Editor: Ivo Hofacker

© The Author 2009. Published by Oxford University Press. All rights reserved. For Permissions, please email: journals.permissions@oxfordjournals.org

ABSTRACT
Summary: RNATOPS-W is a web server to search sequences for RNA secondary structures including pseudoknots. The server accepts an annotated RNA multiple structural alignment as a structural profile and genomic or other sequences to search. It is built upon RNATOPS, a command line C++ software package for the same purpose, in which filters to speed up search are manually selected. RNATOPS-W improves upon RNATOPS by adding the function of automatic selection of a hidden Markov model (HMM) filter and also a friendly user interface for selection of a substructure filter by the user. In addition, RNATOPS-W complements existing RNA secondary structure search web servers that either use built-in structure profiles or are not able to detect pseudoknots. RNATOPS-W inherits the efficiency of RNATOPS in detecting large, complex RNA structures.

Availability: The web server RNATOPS-W is available at the web site www.uga.edu/RNA-Informatics/?f=software&p=RNATOPS-W. The underlying search program RNATOPS can be downloaded at www.uga.edu/RNA-Informatics/?f=software&p=RNATOPS.

Contact: cai@cs.uga.edu

Supplementary information: Supplementary data are available at Bioinformatics online.

1 INTRODUCTION

Searching genomes using computational methods has become important for prediction and annotation of non-coding RNAs (Lowe and Eddy, 1997; Hofacker, 2006; Griffiths-Jones, 2007; Rivas and Eddy, 2003; Klein and Eddy, 2003; Washietl et al., 2005). Other available methods to identify conserved RNA or complex structures including pseudoknots are not available. This is due to the lack of appropriate RNA pseudoknot models that can permit efficient algorithms for structure–sequence alignment, a bottleneck task. Search programs can usually be speeded up with filtering methods that can quickly remove genome segments unlikely to contain the desired pattern in the profile (Bafna and Zhang, 2004; Lowe and Eddy, 1997; Weinberg and Ruzzo, 2006; Zhang et al., 2005), but even with a significant speed-up (e.g. with a 99% genome reduction), searching for complex RNA structures with a pseudoknot may still take hours, if not days, on a typical bacterial or yeast genome.

Our previous work (Song,Y. et al., 2005) introduced a graph-theoretic modeling method for profiling RNA secondary structures including pseudoknots. With this model, we were able to design a very efficient structure–sequence alignment algorithm, ideal for RNA pseudoknot search on genomes, and implemented it in an RNA structure search program called RNATOPS (Huang et al., 2008). One advantage of RNATOPS is its high efficiency searching for large RNA or complex structures including pseudoknots, while maintaining accuracy comparable with other search programs that are only capable of detecting pseudoknot-free structures. To further speed up searches, RNATOPS also executes the whole structure search on filtering results. However, filters (i.e. subsequence or substructure profiles) can only be manually selected. This article presents a web server version of RNATOPS, called RNATOPS-W with a new built-in function for automatic hidden Markov model (HMM) filter selection. The web server also allows an interactive selection of any substructure as a filter through a user-friendly interface.

2 PROGRAM FEATURES

This section presents the filtering functions of the web server RNATOPS-W and its interface features. We refer the reader to our previous work (Huang et al., 2008; Song et al., 2005) for detailed discussions on the search methods and algorithms used by RNATOPS.

2.1 Filtering method

RNATOPS-W incorporates a function of automatic HMM filter selection; the selected filter is used to speed up the search program. The filter selection chooses a conserved region as an HMM filter from the given RNA structural profile (a set of structurally aligned RNA sequences). Our filter selection method was built from two previous approaches used to identify conserved amino acids in protein sequences (Capra and Singh, 2007; Song,B. et al., 2005); it replaces the overall amino acid distribution in the BLOSUM62 matrix with a conserved subsequence from the profile.
To use RNATOPS-W for RNA structure search, the user is asked to submit an RNA structure profile (i.e. a set of structurally aligned training RNAs) in fasta format (Huang et al., 2008) and target genomes in fasta format. These data can be in either a file or an input text box to be uploaded in the start page. By default, RNATOPS-W automatically selects an HMM filter for the given structure profile. The user can also opt to select manually his/her own filter, by specifying the beginning and ending regions of any consecutive substructure from the given structure profile. After the submission of the input and an filter option, the server searches the target genomes with the filter and then searches the filtered hits for whole-structure matches. Each search request is given a ticket number with which the user can retrieve later a search result file from a provided link or from the start page.

For each search request, the result file contains information for all search hits that 'match' the structure profile. For each hit, the following information is produced: the name of the genome containing the hit, the hit sequence, its position in the genome, the score of the hit sequence, the fold conforming to the structure profile and the structural alignment between the hit sequence and the structure profile. The output also contains the parameter settings for each whole search request and the total time used in the search.

Additional options are provided for the user to redefine parameters pertinent to the search algorithm to achieve a desired search accuracy. The user, instead of choosing the 'All default' option, can select 'Adjust parameters'. These parameters mostly concern setting priors for stochastic modeling of individual stems and loops in the structure profile and improving the qualities of candidates found for individual stems. RNATOPS-W provides users a friendly web-interface to perform searches of genomes for RNAs on the basis of their structural profile, including pseudoknots. It adds functionality by automated selection of filters to speed up the search.

2.2 Interface features

To use RNATOPS-W for RNA structure search, the user is asked to submit an RNA structure profile (i.e. a set of structurally aligned training RNAs) in fasta format (Huang et al., 2008) and target genomes in fasta format. These data can be in either a file or an input text box to be uploaded in the start page. By default, RNATOPS-W automatically selects an HMM filter for the given structure profile. The user can also opt to select manually his/her own filter, by specifying the beginning and ending regions of any consecutive substructure from the given structure profile. After the submission of the input and an filter option, the server searches the target genomes with the filter and then searches the filtered hits for whole-structure matches. Each search request is given a ticket number with which the user can retrieve later a search result file from a provided link or from the start page.

For each search request, the result file contains information for all search hits that 'match' the structure profile. For each hit, the following information is produced: the name of the genome containing the hit, the hit sequence, its position in the genome, the score of the hit sequence, the fold conforming to the structure profile and the structural alignment between the hit sequence and the structure profile. The output also contains the parameter settings for each whole search request and the total time used in the search.

Additional options are provided for the user to redefine parameters pertinent to the search algorithm to achieve a desired search accuracy. The user, instead of choosing the 'All default' option, can select 'Adjust parameters'. These parameters mostly concern setting priors for stochastic modeling of individual stems and loops in the structure profile and improving the qualities of candidates found for individual stems. RNATOPS-W provides users a friendly web-interface to perform searches of genomes for RNAs on the basis of their structural profile, including pseudoknots. It adds functionality by automated selection of filters to speed up the search.

ACKNOWLEDGEMENTS

A part of the server interface was implemented with help from Mark Wilson.

Funding: NIH Biomedical Information Science and Technology Initiative (R01GM072080-01A1, in part).

Conflict of Interest: none declared.

REFERENCES


