BMC Bioinformatics

SOFTWARE Open Access

StructRNAfinder: an automated pipeline and web server for RNA families prediction

Raúl Arias-Carrasco[1,2†], Yessenia Vásquez-Morán[1†], Helder I. Nakaya[3*] and Vinicius Maracaja-Coutinho[1,4,5,6*]

Abstract

Background: The function of many noncoding RNAs (ncRNAs) depend upon their secondary structures. Over the last decades, several methodologies have been developed to predict such structures or to use them to functionally annotate RNAs into RNA families. However, to fully perform this analysis, researchers should utilize multiple tools, which require the constant parsing and processing of several intermediate files. This makes the large-scale prediction and annotation of RNAs a daunting task even to researchers with good computational or bioinformatics skills.

Results: We present an automated pipeline named StructRNAfinder that predicts and annotates RNA families in transcript or genome sequences. This single tool not only displays the sequence/structural consensus alignments for each RNA family, according to Rfam database but also provides a taxonomic overview for each assigned functional RNA. Moreover, we implemented a user-friendly web service that allows researchers to upload their own nucleotide sequences in order to perform the whole analysis. Finally, we provided a stand-alone version of StructRNAfinder to be used in large-scale projects. The tool was developed under GNU General Public License (GPLv3) and is freely available at http://structrnafinder.integrativebioinformatics.me.

Conclusions: The main advantage of StructRNAfinder relies on the large-scale processing and integrating the data obtained by each tool and database employed along the workflow, of which several files are generated and displayed in user-friendly reports, useful for downstream analyses and data exploration.

Keywords: RNA family, RNA structure, Noncoding RNAs, Covariance models, Web server, Tool, Pipeline

Background

Noncoding RNAs (ncRNAs) are present in all domains of life, playing a critical role in the fine-tuning regulation of biological processes [1]. Their mode of action varies according to the RNA family it belongs. In 2005, the Rfam database created a limited type ontology to better represent the thousand of families identified so far and stored in the database [2]. Briefly, non-coding RNA genes (Gene) are composed by bona-fide RNAs with a recognised function (e.g. CRISPR, miRNAs, ribozymes, rRNAs, snoRNAs); structured *cis*-regulatory elements (Cis-reg), are represented by structural regulatory motifs available in RNA sequences (e.g. frameshift elements, riboswitches, thermoregulators); and Intron, composed by self-splicing RNAs. The prediction of RNA families in genome or transcriptome sequences often depends on its primary sequence conservation or secondary structural motifs. Thus, several bioinformatics workflows that use third-party software were created to predict or annotate different RNA classes using sequence or structure comparisons [2–5].

Secondary structure-based methods are critical for the annotation of specific regulatory RNAs [6, 7]. These approaches employ nucleotide sequences folding and minimum free energy calculation, in order to predict most of well-known RNA families [8]. For instance, microRNAs (miRNAs) have a characteristic structure of highly paired nucleotides, forming a double strand RNA structure from a single molecule of approximately 100 nt. Further, snoRNAs (small nucleolar RNAs) present a big loop associated to their binding with ribosomal RNAs (rRNAs). These structural features are of key importance for their ab-initio prediction and

* Correspondence: hnakaya@usp.br; viniciusmaracaja@integrativebioinformatics.me
†Equal contributors
[3]Faculdade de Ciências Farmacêuticas, Universidade de São Paulo, São Paulo 05508-900, Brazil
[1]Centro de Genómica y Bioinformática, Facultad de Ciencias, Universidad Mayor, 8580745 Santiago, Chile
Full list of author information is available at the end of the article

classification [9–11], sometimes coupled with other features, such as nucleotides conservation and covariation [12, 13]. Tools such as Infernal [14] utilize nucleotide sequences and/or secondary structure covariance models from known RNA families, like those available in Rfam and other databases [15, 16]. These tools directly compare known RNA families and nucleotide sequences, resulting in the identification of potential novel regulatory RNAs in genome or transcriptome sequences. However, large-scale prediction and annotation of regulatory noncoding RNAs can represent a daunting task, as multiple tools are required along the process. In addition, the constant parsing and processing of intermediate files needed to run these tools impose a great obstacle for researchers with lesser computational or bioinformatics backgrounds.

To avoid these technical bottlenecks, we developed an automated pipeline named StructRNAfinder. The tool was implemented using Perl scripts and third-party software and is focused on the identification and complete annotation of regulatory RNA families from transcriptome or genome sequences. We also implemented a user-friendly web server that allows users with no bioinformatics skills to perform all analyses. Both stand-alone and web server StructRNAfinder versions can be accessed at http://structrnafinder.integrativebioinformatics.me.

Implementation

StructRNAfinder automated pipeline

The stand-alone tool was developed to run in Linux and requires different third-party software (i.e. Infernal [14], RNAfold [3], Rfam database [15], Krona [17], different Perl libraries), automatically downloaded and installed together with the software. All the StructRNAfinder codes were developed using Perl scripts (Additional file 1), which creates HTML pages integrating JavaScript to generate dynamic tables and graphics. The usage is simple, only requiring the definition of two input files for the comparisons. The first is the database of covariance models, used as a reference for RNA families to be found in the second input file, which is the nucleotide sequences in FASTA format. StructRNAfinder was created under GNU General Public License (GPLv3).

Web server implementation

webStructRNAfinder server interface was implemented using PHP, HTML and Perl in an Apache2 server. We applied a FIFO pile (First In First Out) for the management of users jobs, to organise and manage the queue of submitted queries. When a user launches a job, PHP automatically creates a user-specific folder to save its results and retain the job process on the pile. When the process daemon detects a new job, it executes structRNAfinder with the FASTA sequence(s) and parameters provided by the user, updating automatically the PHP files inside the user folder. When the whole process is finished, the provided URL will be made available to the final user for 48 h. All StructRNAfinder generated HTML and output files are made available compressed in a zip file in the *Files* section.

Results

An automated pipeline named StructRNAfinder

To identify potential noncoding RNAs in genome or transcriptome sequences, research groups have to manually run several programs, which generate different formats of intermediate files. StructRNAfinder automates this laborious workflow, processing the data obtained by each employed tool, thus allowing non-bioinformaticians to identify and compare ncRNAs through the primary sequence and secondary structure inferences. All the files generated along the workflow are displayed in user-friendly reports and subsequently made available for downstream analyses. StructRNAfinder utilizes Infernal [14] to annotate genome/transcriptome-derived sequences to the corresponding RNA families. For data derived from next generation sequencing (NGS) studies, it is necessary to have the final sets of assembled sequences. Thus, all sequences are compared against covariance models, which represent the sequence/structural consensus alignments for each RNA family, reported to date in Rfam database [15].

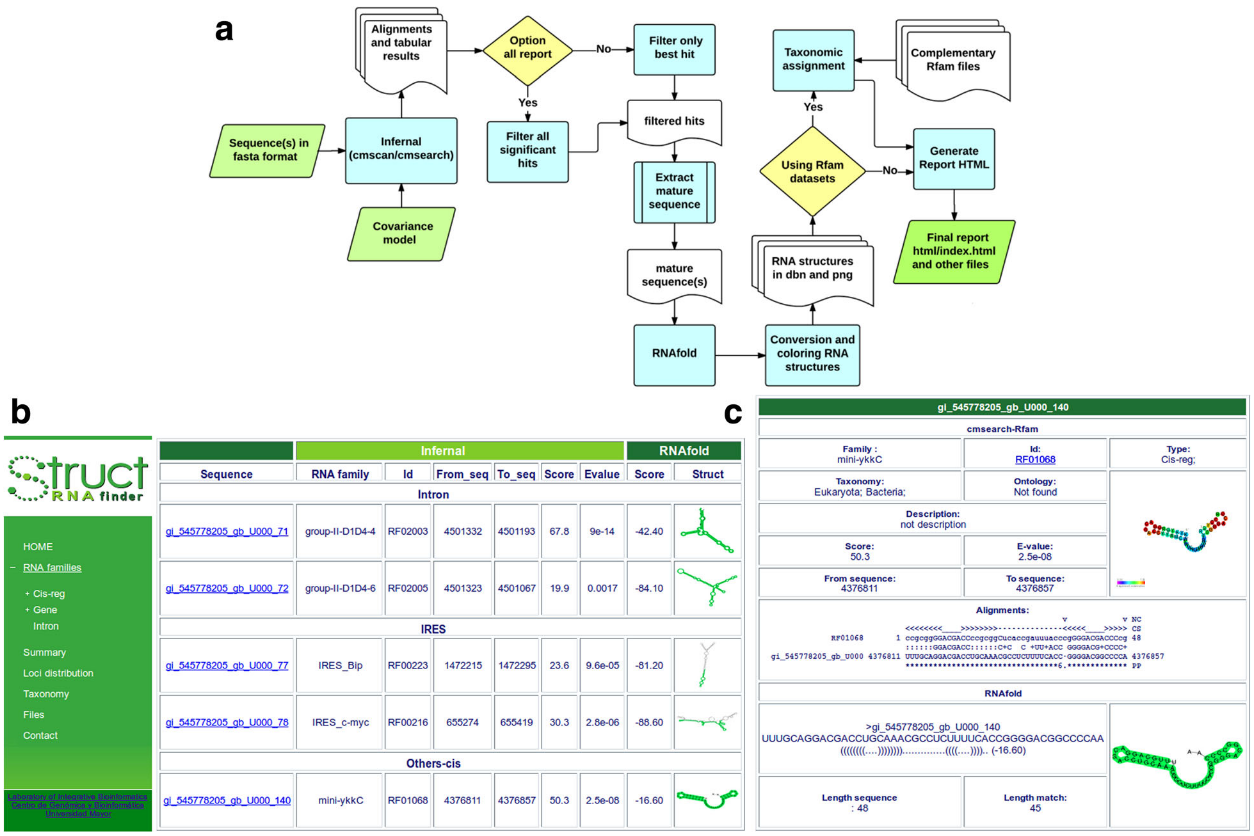
One issue that arises when comparing sequences and covariance models is that current tools only provide results in plain text outputs, which contains the sequence-structure alignments, positional coordinates and its statistics. No information is provided related to a potential annotation of the predicted RNA families, neither images with the potential RNA secondary structure, its description in the standard dot and bracket format, nor its functional description. StructRNAfinder automatically explores and parses Infernal alignments output, by filtering and extracting significant hits for each sequence/covariance model comparison. Based on its mapping coordinates, the tool calculates the length of alignment, the size of the input sequence and the size of the target RNA family. If necessary, the hit length from input sequence is expanded, in order to obtain a mature sequence with a similar size to that of the original Rfam secondary structure, which is used as input to RNAfold for secondary structure predictions. This tool is available in Vienna package [3], which is a widely-used suite of tools to analyse RNA structures. In the final structure, the region assigned to the alignment is highlighted in green. This procedure assures the length needed to estimate the optimal minimum energy, which is sequence- and length-dependent [18]. This secondary structure is a visual representation of the predicted structure, to be compared with those originally generated by Rfam, which is also available on the final report. The text representation of the structure generated by RNAfold and from the CM alignment are also reported. Once an RNA family is assigned, StructRNAfinder automatically retrieves all annotation information available in Rfam database for that particular RNA (i.e. family description, gene ontology, taxonomy, family secondary structure image). The general procedure performed by StructRNAfinder is explained in Fig. 1a.

Comprehensive reports

The reports generated by StructRNAfinder contain the annotation and statistics for all RNA families, secondary structures, alignments, functions and taxonomic assignations identified in the input sequence(s). These reports are provided in HTML format and contain tables and figures that can be used for further data exploration. For instance, the index.html file (available in the main folder of the stand-alone version) displays a table containing all significant hits obtained from the alignment between covariance models and input sequences (Fig. 1b). The menu on the left of the table is generated dynamically according to results and allows quick navigation through the different RNA families identified in the input sequences. If users click on the hyperlink associated with each identifier hit name, a new page is opened containing the complete information of the predicted RNA (Fig. 1c), such as the full description of RNA function (if available), associated gene ontology, covariance model alignment, secondary structure predicted by RNAfold [3] and reported secondary structure from the reference RNA available in Rfam. Briefly, this page contains information extracted from Infernal, RNAfold, Rfam database and generated by our in-house Perl scripts. A general overview of statistics and a graphic representation of predicted RNA families (Fig. 2a, b) are accessible in the *Summary* section.

Visual distribution of predicted RNA families

Users can visualize the localization of each predicted RNA along the nucleotide query sequence. This can be



	Infernal						RNAfold	
Sequence	RNA family	Id	From_seq	To_seq	Score	Evalue	Score	Struct
Intron								
gi_545778205_gb_U000_71	group-II-D1D4-4	RF02003	4501332	4501193	67.8	9e-14	-42.40	
gi_545778205_gb_U000_72	group-II-D1D4-6	RF02005	4501323	4501067	19.9	0.0017	-84.10	
IRES								
gi_545778205_gb_U000_77	IRES_Bip	RF00223	1472215	1472295	23.6	9.6e-05	-81.20	
gi_545778205_gb_U000_78	IRES_c-myc	RF00216	655274	655419	30.3	2.8e-06	-88.60	
Others-cis								
gi_545778205_gb_U000_140	mini-ykkC	RF01068	4376811	4376857	50.3	2.5e-08	-16.60	

gi_545778205_gb_U000_140

cmsearch-Rfam

Family : mini-ykkC	Id: RF01068	Type: Cis-reg;
Taxonomy: Eukaryota; Bacteria;	Ontology: Not found	
Description: not description		
Score: 50.3	E-value: 2.5e-08	
From sequence: 4376811	To sequence: 4376857	

Alignments:

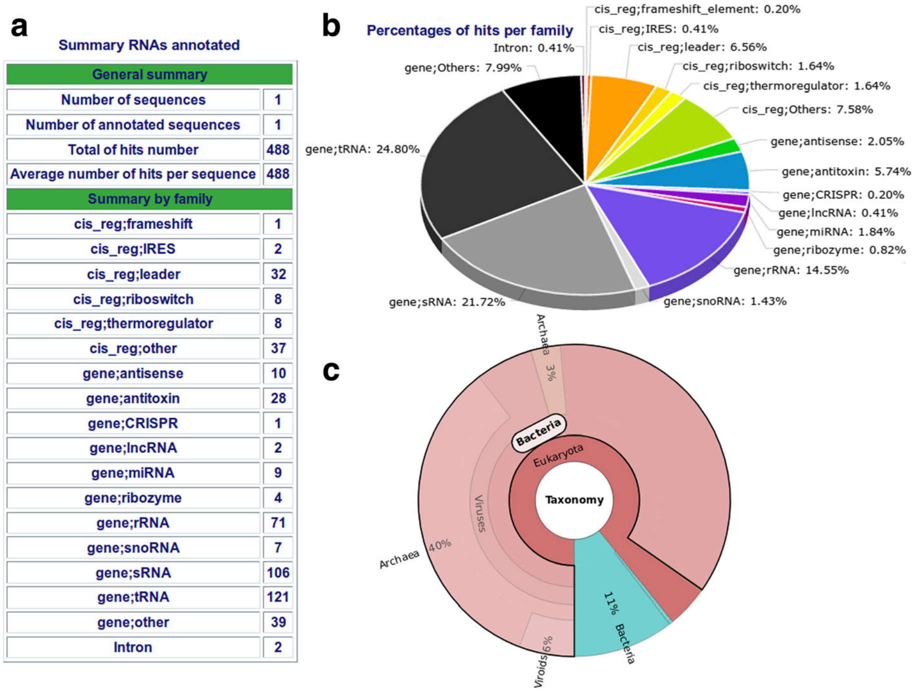
```
                                                              v           v NC
                               <<<<<<<<____>>>>>>>>-------------<<<<<____>>>>> CS
           RF01068           1 ccgcggGGACGACCccgcggCucaccgauuuacccgGGGACGACCCcg 48
                               ::::::GGACGACC::::::C+C  C +UU+ACC GGGGACG+CCCC+
gi_545778205_gb_U000 4376811 UUUGCAGGACGACCUGCAAACGCCUCUUUUCACC-GGGGACGGCCCCA 4376857
                               *********************************6.************* PP
```

RNAfold

```
>gi_545778205_gb_U000_140
UUUGCAGGACGACCUGCAAACGCCUCUUUUCACCGGGGACGGCCCCAA
((((((((....))))))))..............((((....)))).. (-16.60)
```

Length sequence : 48	Length match: 45

Fig. 1 **a** Whole pipeline implemented in StructRNAfinder. Input and output files are shown in green rhomboid; intermediate third-party or in-house scripts are shown in light blue squares; together with intermediate files generated along the process, shown in white shapes. Decisions taken along the process are shown in yellow diamonds. **b** General report containing the list of predicted RNA structures in *E. coli* strain K-12. **c** Detailed table for a particular predicted RNA of interest. All information related to the covariance model comparison, secondary structure inference and complete annotations are made available in this page, including a external link to Rfam database



Summary RNAs annotated

General summary	
Number of sequences	1
Number of annotated sequences	1
Total of hits number	488
Average number of hits per sequence	488
Summary by family	
cis_reg;frameshift	1
cis_reg;IRES	2
cis_reg;leader	32
cis_reg;riboswitch	8
cis_reg;thermoregulator	8
cis_reg;other	37
gene;antisense	10
gene;antitoxin	28
gene;CRISPR	1
gene;lncRNA	2
gene;miRNA	9
gene;ribozyme	4
gene;rRNA	71
gene;snoRNA	7
gene;sRNA	106
gene;tRNA	121
gene;other	39
Intron	2

Fig. 2 a, b Exemplary results of StructRNAfinder summary section. This section provides a general statistics of all identified RNA families in *E. coli* strain K-12 genome sequence. **a** Table showing the total numbers of each predicted RNA family according to Rfam nomenclature. **b** A pie-chart of the numbers shown in A. **c** A dynamic pie-chart with the taxonomic assignation of the 488 identified RNAs

useful to identify potential RNA clusters generated from a unique precursor RNA or to obtain a genome-wide visualization of predicted RNAs in a whole or partial genome sequence. The *Loci distribution* section provides a visual representation of all RNA families identified along the analysed nucleotide sequence. If more than one sequence is used as input, this page will provide one image for each analysed sequence with the general distribution of RNA families.

Taxonomic distribution and visualization

StructRNAfinder recovers the taxonomy of each predicted RNA family, according to Rfam annotation for each RNA family. We used Krona [17] to generate interactive graphics that show the abundance of all RNAs belonging to different taxonomic groups based on Rfam species annotations. In Rfam database, each RNA family is the result of multiple sequences alignments from different species. StructRNAfinder summarizes and plots the presence of all predicted RNAs according to three domains of life, plus viruses (Fig. 2c). For instance, the graphic on Fig. 2c shows the taxonomic distribution of 488 RNA families predicted on the *E. coli* strain K-12 substr. MG1655 genome sequence (accession number: U00096). This is a dynamic graphic, allowing the navigation within the number of predicted RNA available in each evolutionary branches. In this example, 53 RNA families (11% of the total) are present exclusively on the Bacteria domain (light blue in Fig. 2c); while 413 RNA families (85% of the total) are shared between Bacteria and other evolutionary branches (light red in Fig. 2c).

Output files for downstream analysis

StructRNAfinder generates several files in different formats. They can be useful to advance downstream analyses or can supplement information available from HTML reports, obtained after running the tool. All files can be accessed on the *Files* section. Standard outputs from Infernal and RNAfold tools are available, together with other files generated by StructRNAfinder. Output files include: (i) a BED format file containing the positional coordinates of predicted structures according to the nucleotide sequences used as input; (ii) a FASTA format file containing the nucleotide sequences of the predicted RNAs; (iii) an annotation tabular file comprised of extensive information generated by StructRNAfinder. This annotation file contains the RNA family name, the RNA class, Rfam database identifiers, scores and e-values from each prediction according to Infernal and folding energies according to RNAfold, the start and end positions of each prediction on the query sequence, and finally, a functional description of the predicted RNA.

webStructRNAfinder: An user friendly web server

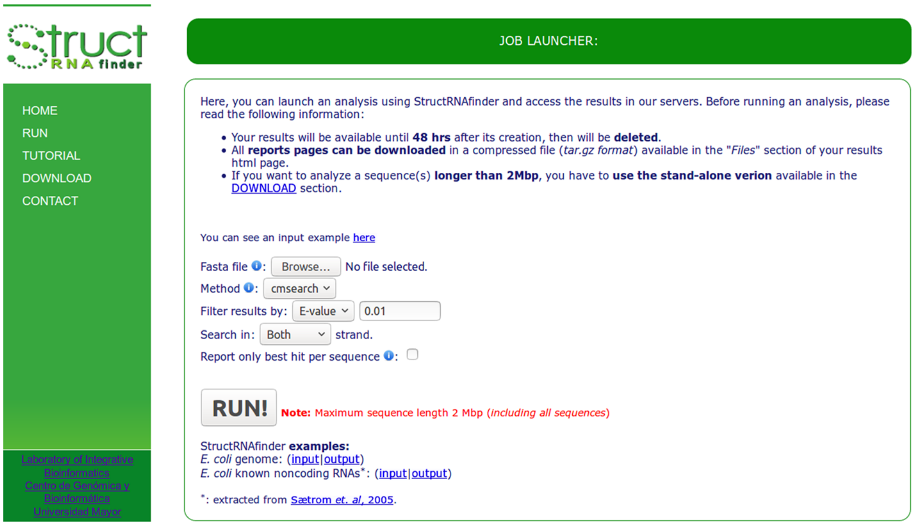
webStructRNAfinder server provides a job launcher interface (*RUN* section) where users can analyse sequences using different search methods according to their own criteria. Users are only required to provide a FASTA sequence(s) file, and to fill a small set of required parameters (Fig. 3). The parameters to choose are: (i) the Infernal search method (cmsearch, who searches the covariance models in a database composed by the input sequences; or cmscan, who searches the input sequences in a database composed by the covariance models. This difference influences the e-value calculation, due to this value mainly depend on the database size.); (ii) the cutoff filter to be used according to Infernal, based on: e-value, score, or one of the three covariance model-specific reporting thresholds (gathering, noise or trusted); (iii) the option to receive a report considering all significant hits according to selected e-value/score/CM-threshold or only the best one per sequence based on the lowest e-value; (iv) performs the search in both strands or only in one. As soon as StructRNAfinder finishes the whole analysis, the results will remain available on the provided hyperlink for 48 h. On the *Files* section, users can download a compressed file in zip format containing all output files and HTML web pages generated by the tool.

StructRNAfinder exemplary reports are made available in the *RUN* section. One with RNA families predicted in the genome of the eukaryotic human pathogen *Leishmania braziliensis* (Additional file 2); another in *E. coli* str k-12 genome (Fig. 2), both using the cmsearch Infernal method and filtered by an e-value of 0.01; and with RNA families predicted in a dataset of experimentally verified ncRNA sequences extracted from Sætrom and collaborators [19]. This last analysis predicted correctly 151 out of 154 (98.05% of the total) experimentally validated RNAs (Additional file 3).

Discussion

In this work, we described a new automated pipeline, named StructRNAfinder, which was developed to facilitate the identification and complete annotation of regulatory RNA families available in nucleotide sequences (DNA or RNA). When predicting/annotating RNAs in nucleotides sequences, the user needs to manipulate several plain text results generated by different tools used on the process, which are difficult to visualize and manipulate in typical text viewers rather than Linux command line. The advantage of StructRNAfinder relies on processing and integrating the data obtained by each tool and database employed along the workflow, of which several files are generated and displayed in user-friendly reports, useful for downstream analyses and data exploration.

We have successfully applied the stand-alone version of this automated pipeline on analyses of the noncoding RNA content in genomic sequences from the most diverse set of organisms, covering all three domains of life, plus viruses (unpublished and published data). For instance, on the development of LeishDB [20], a reference database for *Leishmania braziliensis* genomic information; we used this tool, together with other strategies, to obtain the most comprehensive characterization of non-coding RNAs for *Leishmania* species. Analyses using genomic and transcriptomic *E. coli* datasets allowed the prediction of 488 different RNAs (cmsearch e-value cutoff of 0.01), part of 184 different Rfam families, on the genomic sequence of *E. coli* str k-12; and the correct prediction of 98.05% of experimentally validated RNA


Struct
RNA finder
HOME
RUN
TUTORIAL
DOWNLOAD
CONTACT
Laboratory of Integrative Bioinformatics Centro de Genómica y Bioinformática Universidad Mayor
JOB LAUNCHER:
Browse...
RUN!

Here, you can launch an analysis using StructRNAfinder and access the results in our servers. Before running an analysis, please read the following information:

- Your results will be available until **48 hrs** after its creation, then will be **deleted**.
- All **reports pages can be downloaded** in a compressed file (*tar.gz format*) available in the "*Files*" section of your results html page.
- If you want to analyze a sequence(s) **longer than 2Mbp**, you have to **use the stand-alone verion** available in the DOWNLOAD section.

You can see an input example here

Fasta file: No file selected.
Method: cmsearch
Filter results by: E-value 0.01
Search in: Both strand.
Report only best hit per sequence: □

Note: Maximum sequence length 2 Mbp (*including all sequences*)

StructRNAfinder **examples:**
E. coli genome: (input|output)
E. coli known noncoding RNAs*: (input|output)

*: extracted from *Sætrom et. al*, 2005.

Fig. 3 Job launcher screen showing the different parameters that a user can use when running StructRNAfinder web server

transcripts (i.e. predicted with the same functional annotation as provided by the authors [19]). The remaining 2% are related to validated RNAs in which their covariance models are not yet made available in Rfam database.

StructRNAfinder is presented in both stand-alone and web server versions, facilitating the usage for all kind of users. The web-based version allows the user-friendly prediction of RNA families available in sequences of up to 10,000,000 nucleotides (10 Mb), which is enough to predict the repertoire of regulatory RNAs in the sequenced genomes of all Archaea and most of Bacteria available in NCBI. Indeed, Rfam allows a batch sequence search on their web server. However, it does not allow the usage of different filtering options available in StructRNAfinder (i.e. Infernal search method, cutoff filters, strand-specific search, best hit per sequence selection), and its search is limited to 200,000 nucleotides (200 Kb). A table comparing webStructRNAfinder, structRNAfinder and Rfam batch search is available in Additional file 4.

The advantages of StrucRNAfinder stand-alone compared to its web-based version are the possibility to analyse large genomes, e.g. eukaryotic organisms; to perform large-scale analyses using several genomes; and to use as input the covariance models generated by the user itself, instead of those generated by Rfam. Indeed, recently, Eggenhofer and collaborators developed an unsupervised tool that allows the generation of covariance models using a single sequence as input [21]. It collects potential RNA family members from a model generated based on multiple interactions of homology searches against nucleotide sequences of taxonomy related organisms available in NCBI, and structural alignments analyses, in order to generate the most suitable covariance model for that particular RNA sequence. For instance, this is useful for structural conservation analysis of new ncRNAs found in transcriptomes, in which users can easily generate a covariance model of one or a set of transcripts of interest and search for its presence in different sets of genome sequences using the stand-alone version of StructRNAfinder.

Conclusions

StructRNAfinder facilitates de prediction and complete functional annotation of RNA families in nucleotides sequences, by integrating different tools and databases commonly used on the prediction and functional annotation of regulatory RNAs. One main advantage over other existing tools relies on the large-scale processing and integration of the data obtained by each tool and database employed along the workflow, with useful files generated and displayed in user-friendly reports helpful for downstream analyses. The automatic generations of these reports avoid the time-consuming process of writing scripts for parsing the output and input files for the tools and databases employed, especially for users without any programming or bioinformatics skills. These features facilitate the genome-wide predictions of the complete repertoire of RNA families available in small and large genomes and also in assembled transcriptomes derived from NGS studies.

Availability and requirements

Project name: StructRNAfinder.

Project home page: http://structrnafinder.integrativebioinformatics.me

Operating system(s): LINUX.

Programming language: Perl, HTML5, PHP5 and JavaScript.

Other requirements: Linux Perl package libgd-perl version 2.53, Perl package Bio::Graphics 2.4, Infernal tool version 1.1, Vienna package version 2.1.8, Rfam covariance models version 12.

License: GNU GPL version 3.

Additional files

Additional file 1: Table in XLS format summarizing each developed Perl script part of StructRNAfinder tool. (XLS 18 kb)

Additional file 2: Exemplary results (figure in PNG format) of StructRNAfinder in *Leishmania braziliensis* genome. **(A)** A pie-chart of the total numbers of each predicted RNA family according to Rfam nomenclature. (**B**) Table showing the numbers shown in A. (**C**) A dynamic pie-chart with the taxonomic assignation of identified RNAs. (PNG 116 kb)

Additional file 3: (**A**) Exemplary results (figure in PNG format) of StructRNAfinder in *E. coli* validated transcripts from Sætromet al., 2005. **(A)** A pie-chart of the total numbers of each predicted RNA family according to Rfam nomenclature. (**B**) Table showing the numbers shown in A. (**C**) A dynamic pie-chart with the taxonomic assignation of identified RNAs. (PNG 120 kb)

Additional file 4: Table in XLS format comparing the features of StructRNAfinder, webStructRNAfinder and Rfam batch search. (XLS 21 kb)

Abbreviations

BED: Browser extensible data; Cis-reg: Cis-regulatory; CM: Covariance model; CRISPR: Clustered regularly interspaced short palindromic repeats; DNA: Deoxyribonucleic acid; *E. coli*: *Escherichia coli*; FIFO: First in first out; GNU GPL: GNU general public license; GNU: GNU is not Unix; HTML: Hypertext markup language; Kb: Kilobases; *L. braziliensis*: *Leishmania braziliensis*; Mb: Megabases; miRNA: MicroRNA; NCBI: National center for biotechnology information; ncRNA: Noncoding RNA; NGS: Next generation sequencing; PHP: Hypertext preprocessor; RNA: Ribonucleic acid; rRNA: Ribosomal RNA; snoRNA: Small nucleolar RNA; URL: Uniform resource locator

Acknowledgements

The authors would like to thank Dr.(c) Victor Aliaga-Tobar, Dr. Alvaro Orell, Dr. Paulo P. Amaral and Dr. Patricio Manque for the helpful discussions during the development of StructRNAfinder and the preparation of this manuscript; and Soledad Sandoval for designing the StructRNAfinder logo.

Funding

This work was supported by the Programa Becas Iberoamérica, Jóvenes Profesores e Investigadores, Santander Universidades, Chile; Fondecyt Iniciación, Comisión Nacional de Investigación Científica y Tecnológica

(CONICYT), Chile, grant 11161020; Programa Nacional de Inserción de Capital Humano Avanzado en la Academia, PAI-CONICYT, Chile, grant PAI79170021; Fondo de Financiamiento de Centro de Investigación en Áreas Prioritárias (FONDAP), CONICYT, grant 15130011; Programa de Bienes Públicos Estratégicos para la Competitividad, Corporación de Fomento de la Producción (CORFO), Chile, grant 16BPE-62321; Subsidio Semilla de Asignación Flexible (SSAF), CORFO, grant 14-SSAF-27061-9; and Fundação de Amparo à Pesquisa do Estado de São Paulo (FAPESP), Brazil, grant 12/19278–6. RAC received a PhD fellowship from Vicerrectoría de Investigación, Universidad Mayor, Chile..

Availability of data and materials

Some of the data analysed during this study were obtained from the article: Sætrom P, Sneve R, Kristiansen KI, Snøve O Jr., Grünfeld T, Rognes T, Seeberg E: Predicting non-coding RNA genes in Escherichia coli with boosted genetic programming. Nucleic Acids Res. 2005, 33:3263–3270. https://doi.org/10.1093/nar/gki644
Both stand-alone and web server StructRNAfinder versions can be accessed at http://structrnafinder.integrativebioinformatics.me.

Authors' contributions

RAC and YVM wrote the tool's scripts. RAC developed the web server. RAC, VMC and HN wrote and reviewed the manuscript. VMC conceived and supervised the research. All authors read and approved the final manuscript.

Ethics approval and consent to participate

Not applicable.

Consent for publication

Not applicable.

Competing interests

The authors declare that they have no competing interests.

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Author details

[1]Centro de Genómica y Bioinformática, Facultad de Ciencias, Universidad Mayor, 8580745 Santiago, Chile. [2]Programa de Doctorado en Genómica Integrativa, Vicerrectoría de Investigación, Universidad Mayor, 8580745 Santiago, Chile. [3]Faculdade de Ciências Farmacêuticas, Universidade de São Paulo, São Paulo 05508-900, Brazil. [4]Instituto Vandique, João Pessoa 58000-000, Brazil. [5]Beagle Bioinformatics, 8320000 Santiago, Chile. [6]Advanced Center for Chronic Diseases (ACCDiS), Facultad de Ciencias Químicas y Farmacéuticas, Universidad de Chile, 8380492 Santiago, Chile.

Received: 13 October 2017 Accepted: 2 February 2018
Published online: 17 February 2018

References

1. Mattick JS. The central role of RNA in the genetic programming of complex organisms. An Acad Bras Cienc. 2010;82:933–9.
2. Griffiths-Jones S, Moxon S, Marshall M, Khanna A, Eddy SR, Bateman A. Rfam: annotating non-coding RNAs in complete genomes. Nucleic Acids Res. 2005;33:D121–4.
3. Lorenz R, Bernhart SH, Höner Zu Siederdissen C, Tafer H, Flamm C, Stadler PF, et al. ViennaRNA Package 2.0. Algorithms Mol. Biol. 2011;6:26.
4. Mattei E, Pietrosanto M, Ferrè F, Helmer-Citterich M. Web-beagle: a web server for the alignment of RNA secondary structures. Nucleic Acids Res. 2015;43:W493–7.
5. Smith C, Heyne S, Richter AS, Will S, Backofen R. Freiburg RNA tools: a web server integrating INTARNA, EXPARNA and LOCARNA. Nucleic Acids Res. 2010;38:W373–7.
6. Cech TR, Bass BL. Biological catalysis by RNA. Annu Rev Biochem. 1986;55: 599–629.
7. Machado-Lima A, del Portillo HA, Durham AM. Computational methods in noncoding RNA research. J Math Biol. 2008;56:15–49.
8. Dandekar T, Hentze MW. Finding the hairpin in the haystack: searching for RNA motifs. Trends Genet. 1995;11:45–50.
9. Aguiar RR, Ambrosio LA, Sepúlveda-Hermosilla G, Maracaja-Coutinho V, Paschoal AR. miRQuest: integration of tools on a Web server for microRNA research. Genet Mol Res. 15(1):gmr.15016861.
10. Hertel J, Hofacker IL, Stadler PF. SnoReport: computational identification of snoRNAs with unknown targets. Bioinformatics. 2008;24:158–64.
11. Hertel J, Stadler PF. Hairpins in a haystack: recognizing microRNA precursors in comparative genomics data. Bioinformatics. 2006;22:e197–202.
12. An J, Lai J, Lehman ML, Nelson CC. miRDeep*: an integrated application tool for miRNA identification from RNA sequencing data. Nucleic Acids Res. 2013;41:727–37.
13. Washietl S, Hofacker IL, Stadler PF. Fast and reliable prediction of noncoding RNAs. Proc Natl Acad Sci. 2005;102:2454–59.
14. Nawrocki EP, Eddy SR. Infernal 1.1: 100-fold faster RNA homology searches. Bioinformatics. 2013;29:2933–5.
15. Nawrocki EP, Burge SW, Bateman A, Daub J, Eberhardt RY, Eddy SR, et al. Rfam 12.0: updates to the RNA families database. Nucleic Acids Res. 2015;43:D130–7.
16. Paschoal AR, Maracaja-Coutinho V, Setubal JC, Simões ZLP, Verjovski-Almeida S, Durham AM. Non-coding transcription characterization and annotation: a guide and web resource for non-coding RNA databases. RNA Biol. 2012;9:274–82.
17. Ondov BD, Bergman NH, Phillippy AM. Interactive metagenomic visualization in a web browser. BMC Bioinformatics. 2011;12:385.
18. Mathews DH, Sabina J, Zuker M, Turner DH. Expanded sequence dependence of thermodynamic parameters improves prediction of RNA secondary structure. J Mol Biol. 1999;288:911–40.
19. Sætrom P, Sneve R, Kristiansen KI, Snøve O Jr, Grünfeld T, Rognes T, et al. Predicting non-coding RNA genes in Escherichia Coli with boosted genetic programming. Nucleic Acids Res. 2005;33:3263–70.
20. Torres F, Arias-Carrasco R, Caris-Maldonado JC, Barral A, Maracaja-Coutinho V, De Queiroz ATL. LeishDB: a database of coding gene annotation and non-coding RNAs in Leishmania braziliensis. Database. 2017;2017:bax047.
21. Eggenhofer F, Hofacker IL, Höner Zu Siederdissen C. RNAlien - unsupervised RNA family model construction. Nucleic Acids Res. 2016;44:8433–41.