RNAspace: non-coding RNA annotation web platform

Philippe Bardou, Marie-Josée Cros, Christine Gaspin, Daniel Gautheret, Benjamin Grenier-Boley, Jean-Marc Larre, Jérôme J. Mariette, Antoine de Monte, Hélène Touzet

RNAspace: non-coding RNA annotation web platform

Philippe Bardou[1], Marie-Josée Cros[2], Christine Gaspin[2,3], Daniel Gautheret[4], Benjamin Grenier-Boley[5], Jean-Marc Larré[3], Jérôme Mariette[3], Antoine de Monte[5], Hélène Touzet[5]

[1] INRA, SIGENAE, UMR 444, F-31320 Castanet, [2] INRA, Unité de Biométrie et Intelligence Artificielle, UR 875, F-31320 Castanet, [3] INRA, Plateforme bioinformatique, F-31320 Castanet, [4] IGM UMR 8621 CNRS-U Paris sud, [5] LIFL, UMR CNRS 8022 Université Lille 1 and INRIA LNE

Contact: rnaspace-project@lists.sourceforge.net

General presentation

The increasing number of ncRNA discovered since 2000 and the lack of user friendly tools for finding and annotating them, have made necessary to propose to biologists an *in silico* environment allowing structural and functional annotations of these molecules. For the purpose, RNAspace is developed as a collaborative and open software allowing to:

- **run a variety of non-protein-coding RNA (ncRNA) gene finders in an integrated environment**,
- **explore computed results** with dedicated tools for comparison, visualization, alignment and edition of putative ncRNAs
- and **export** them in various formats (FASTA, GFF, RNAML).

The platform is developed both as a web site **rnaspace.org** (with limitations on analyzed sequence size and execution time), and for local installation with user authentication. It is written in Python language using CherryPy web framework (see cherrypy.org) and Cheetah template engine (see cheetahtemplate.org).

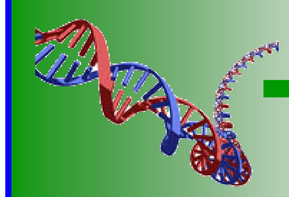
Acknowledgements: This work was founded by RNG (Réseau National des Génopoles), french network of Genomic Centers.

Main workflow

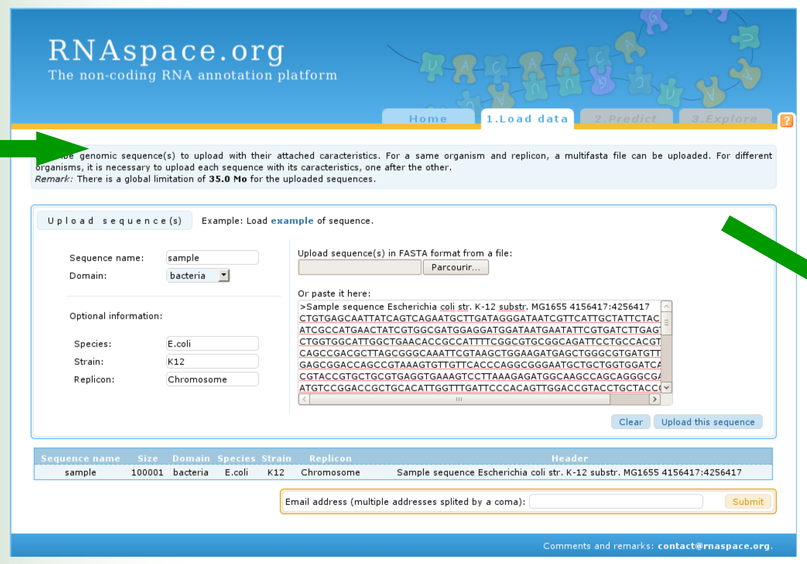
Gene finders are organized into 3 categories containing respectively:

- **known RNA based gene finders** including
  - (i) sequence homology search tools (BLAST[1], YASS[2]) on ncRNA databases (Rfam[3], fRNAdb[4]),
  - (ii) general purpose RNA motif search tools (darn![5], Erpin[6]),
  - (iii) specialized search tools (RNAmmer[7] for ribosomal RNAs, tRNAscan_SE[8] for transfer RNAs)
- **comparative analysis gene finders** (an *ad hoc* pipeline has been implemented based on BLAST or YASS for similarities search and caRNAc[9] or RNAz[10] for consensus structure inference),
- ***ab initio* gene finders** (just one based on detection of atypical GC% regions).

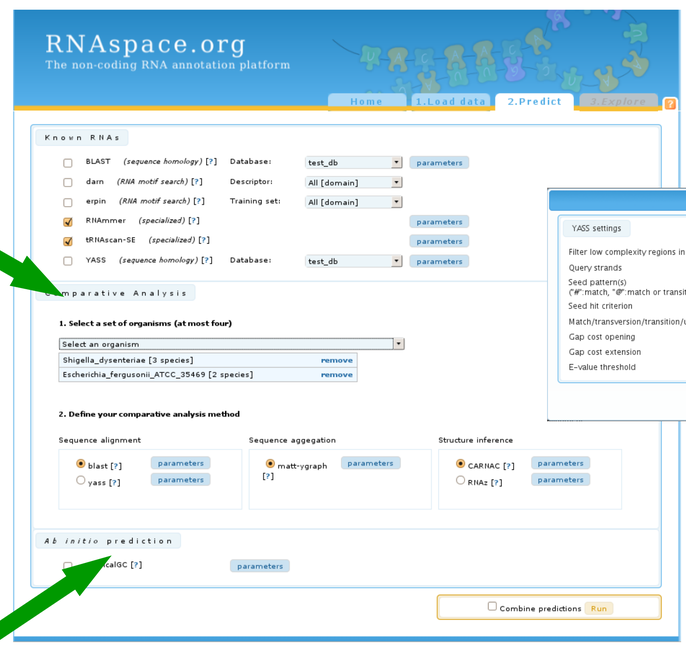
1. Enter genomic sequence(s)



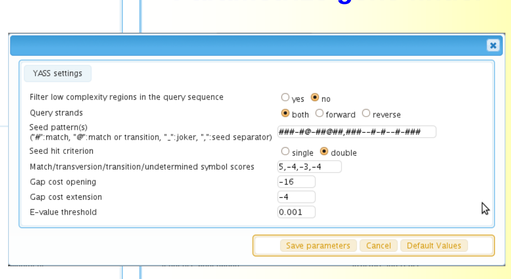
Genomic sequence

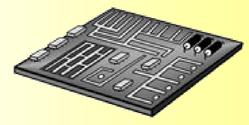


2. Predict ncRNAs



Parametrize gene finder

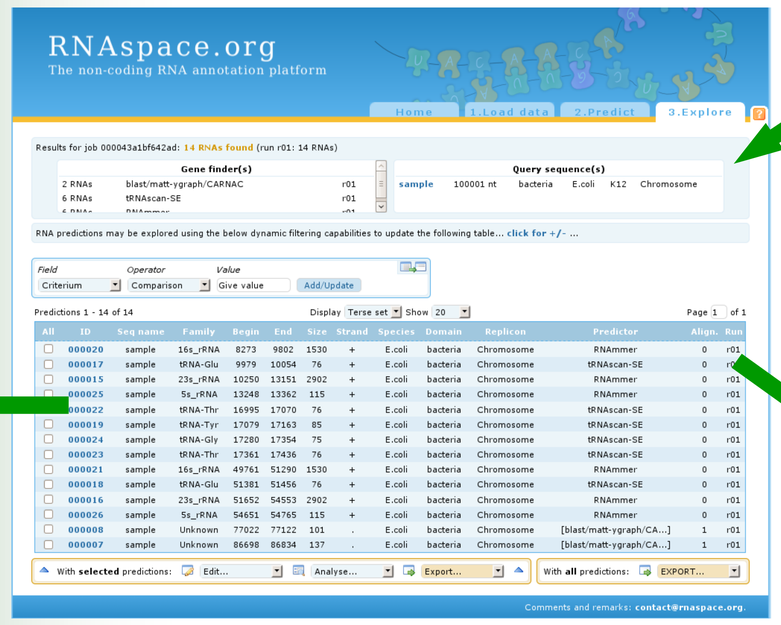


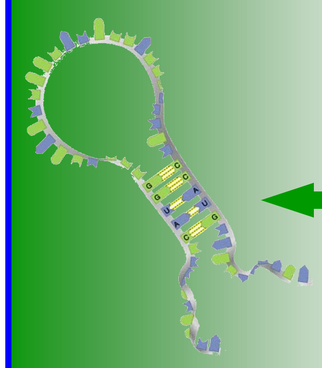


PC or cluster

BIOINFORMATIC'S PLATFORM

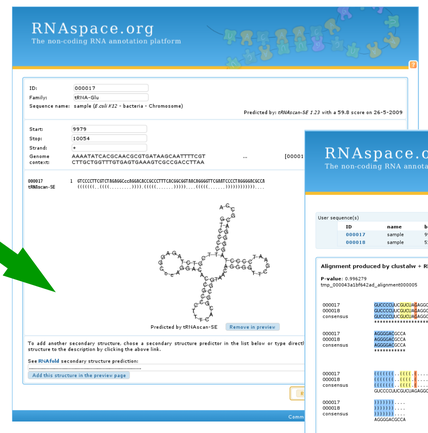
3. Explore and export putative ncRNAs



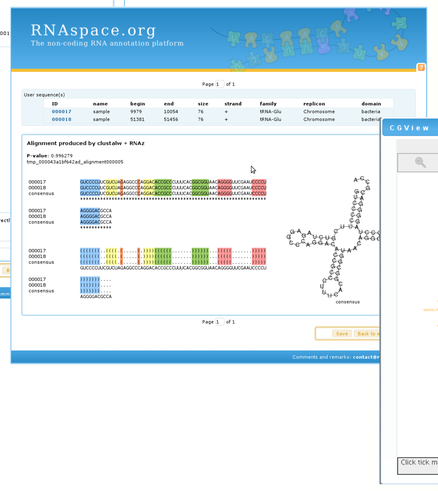


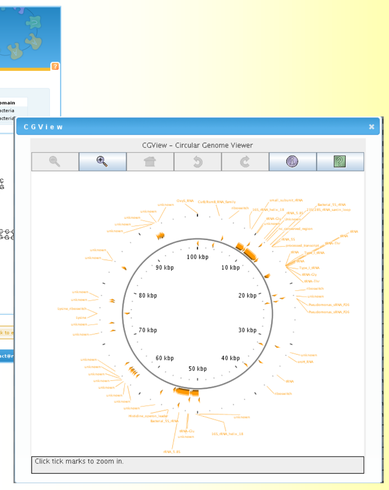
Putative ncRNA

Edit ncRNA



Generate and visualize ncRNAs alignment





Visualize ncRNAs on genomic sequence

(using CGview, see wishart.biology.ualberta.ca/cgview)

Dynamic exploration (sorting and filtering) of putative ncRNAs found on the genomic sequence. **Additional information can be computed and/or visualized on line** (*e.g.,* secondary structure, alignment of predictions), **saved and stored**.
It is also possible to edit and to delete any putative ncRNA.

References

[1] S.F. Altschul, W. Gish, W. Miller, E.W. Myers, D.J. Lipman, Basic local alignment search tool. J Mol Biol 215 (3): 403–410, 1990.
[2] L. Noe, G. Kucherov, YASS: enhancing the sensitivity of DNA similarity search. Nucleic Acids Research, 33(2), 2005:
[3] P.P. Gardner, J. Daub, J.G. Tate, E.P. Nawrocki, D.L. Kolbe, S. Lindgreen, A.C. Wilkinson, R.D. Finn, S. Griffiths-Jones, S.R. Eddy and A. Bateman, Rfam: updates to the RNA families database. Nucleic Acids Research, 37(Database Issue), 2009.
[4] T. Kin, K. Yamada, G. Terai, H. Okida, Y. Yoshinari, Y. Ono, A. Kojima, Y. Kimura, T. Komori, K. Asai, fRNAdb: a platform for mining/annotating functional RNA candidates from non-coding RNA sequences. Nucleic Acids Research, 35(Database Issue), 2007.
[5] M. Zytnicki, C. Gaspin, T. Schiex, DARN! A Weighted Constraint Solver for RNA Motif Localization. Constraints, Vol. 13, 2008.
[6] D. Gautheret, A. Lambert, Direct RNA Motif Definition and Identification from Multiple Sequence Alignments using Secondary Structure Profiles. J Mol Biol. 313:1003-11, 2001.
[7] K. Lagesen, P. Hallin, E.A. Rødland, H.-H. Stærfeldt, T. Rognes, D.W. Ussery, RNAmmer: consistent and rapid annotation of ribosomal RNA genes. Nucleic Acids Research, 35(9), 2003.
[8] T.M. Lowe, S.R. Eddy, tRNAscan-SE: A program for improved detection of transfer RNA genes in genomic sequence. Nucleic Acids Research, 25(5), 1997.
[9] H. Touzet, O. Perriquet, CARNAC: folding families of non coding RNAs. Nucleic Acids Research, 142(Web Server Issue), 2004.
[10] S. Washietl, I.L. Hofacker, P.F. Stadler, Fast and reliable prediction of noncoding RNAs. Proc. Natl. Acad. Sci. U.S.A. 102, 2454-2459, Feb. 2005.
[11] P. Stothard, D.S. Wishart, Circular genome visualization and exploration using CGView. Bioinformatics 21:537-539.