**Abstract:**

In the last ten years, several tools have been proposed for RNA secondary structure pairwise comparison. These tools use different models (ordered tree or forest, arc annotated sequence, multi-level tree) and methods (edit distance, alignment). We present a first online benchmark for comparing these tools. For various RNA families, we built two sets of secondary structures. The second is called the reference set, composed of a small number of RNAs with their known structures. The second is composed of sequences folded using Mfold and RNAs shapes. Some of these sequences correspond to structural RNAs of the same family (true events), others correspond to noise. We studied the ability of each tool to find the true events using the reference set.

**Protocol:**

For each run, two sets of RNA secondary structures are built. The reference set is composed of 4 to 6 RNAs of a family using the structures provided in the literature. The data set is composed of structures obtained by folding sequences of RNAs of the same family (true events) and sequences of the same length as the references but supposed not to belong to that family (called noise or false events).

For each event, the best score obtained between the references and all its possible structures (optimal and suboptimal structures found by mfold or mashape) is retained. Given all events sorted by their best scores, a ROC curve (False Positive Rate / True Positive Rate) is plotted.

**Tools:**

- RNAtoReaver [1] is an ordered trees local/global alignment algorithm. It uses a special tree encoding that allows to break nucleotide pairings under certain conditions.

- MiGal [2] uses a multi-level representation of the secondary structure composed by four layers coded by rooted ordered trees. The layers model different structural levels from multi-loop network to the sequence of nucleotides composing the RNA. The algorithm is an adapted edit distance successively applied to each layer. (options: -harp -strict -indel -unc)

- TreeMatching [3] is based on a quotiented tree representation of the secondary structure which is a similar structure made of two rooted ordered trees at two different scales (nucleotides and structural elements). The core of the method relies on the comparison of both scales simultaneously: it computes an edit distance between quotiented trees at the macroscopic scale using edit costs defined as edit distances between subtrees at the microscopic scale.

- gardenia [4] and NestedAlign [5] use an arc-annotated based representation, that allows for complex edit operations, such as arc-breaking or arc- altering. They allow local and global alignment features. gardenia notably allows affine gap scores while NestedAlign implements an original local alignment algorithm.

- RNAMatch [6] performs the comparison in two steps. First, it compares stems of the two structures using an alignment algorithm with complex edit operations. Then it finds an optimal mapping between the different stems.

- RNAdistance [7] implements a classical edit distance on a tree representation of the structure. A particularity of RNAdistance is that it does not take into account the RNA sequences.

We also compute the score using blast [8] (b2seq -t blastn -W4).

1. M. Höchsmann, T. Töller, R. Giegerich, S. Kurtz
   Local Similarity in RNA Secondary Structures,
   Proceedings of the IEEE Bioinformatics Conference 2003
2. Julien Allali, Yves d’Aubenton-Carafa, Cédric Chauve, Alain Denise, Christine Drevet, Pascal Ferraro, Daniel Gautheret, Claire Herbrand, Fabrice Lexier, Antoine de Monte, Aida Ouangraoua, Marie-France Sagot, Cédric Saulé, Michel Tamaru, Claude Thomas, Hélène Touzet
   LaBRI, Bordeaux 1, CGM, Gil s’Yvette INRIA, LBIEE, Lyon 1, LRI, Orsay Univ. S. Fraser, Canada 1, IGMM, Orsay, LIFL, Lille "MAEM, Nancy
   This work was supported by the ANR project BRASERO ANR-06-BLAN-0045.

**Conclusion:**

We present a general protocol to evaluate the scoring capabilities of methods for comparing RNA secondary structures.

In particular, the data and the software used in this benchmark are freely available at: http://brasero.labri.fr.

These results are preliminary since most of these tools were run with their default parameters. Hence, this work represents only a starting point for a general benchmark. The impact of the various parameters common or specific to each tool (scoring function, matrices...) will be studied in future.

Currently, we considered data sets for three families: tRNA, SRP and 16S. We will add Intron Group I and II, RNAseP and 23S to the final benchmark.

Finally, another benchmark will be added to analyse the quality of the RNA alignments provided by the methods.

---

**REFERENCES**

1. Y. Ponty, M. Ternier and A. Denise
   GenRGenS: Software for generating random genomic sequences and structures, Bioinformatics, 2006

2. Bellench, H. Le Sot, C. Allali and M. F. Sagot
   A multiple layer model to compare RNA secondary structures
   Software: Practice and Experience 2007 (online)

3. A. Ouangraoua, P. Ferraro, L. Tichit, S. Dulucq
   Local similarity between quotiented ordered trees.

4. G. Blin and H. Tourret
   How to compare arc-annotated sequences: The alignment hierarchy.
   SPIRE 2006

5. C. Herbrand
   Elaboration algorithmique et du classement de la comparaison de structures secondaires d’ARN.
   Thesis 2007

6. V. Guignon, C. Chauve, S. Hamel
   An edit distance between RNA stem-loops.
   SPIRE 2005

   Fast Folding and Comparison of RNA Secondary Structures
   Monatshefte f. Chemie 1994

8. S.F. Altschul, W. Gish, W. Miller, E. W. Myers, D. J. Lipman
   Basic local alignment search tool
   Journal of Molecular Biology 1990.