A Simple Protocol for the Inference of RNA Global Pairwise Alignments

Recap

Similiar secondary structures are usually similiar in functionality.

SPS

Is called the Sum of Pairs Score, Similiarity is: 1 - (edit distance / unaligned length of shorter sequence)

Sequence Identity

The identity is the number of identical nucleotides divided by the shorter sequence length, so a measure for how many parts of the sequences are 'equal' or so to say.

needle

needle is an implementation of the sequence based Needleman-Wunsch algorithm, which is basically only the classical edit-distance (Levenstein-distance) algorithm.

Needleman-Wunsch match = 1 mismatch = -1 gap = -1U G С Α Т G С 0 -1 -2 -3 -4 -5 -6 -7 G -1 -0 -1 -2 -3 -4 -5 Α -2 0 0 1 0 -1 -2 -3 Т 0 2 -3 1 0 -1 -1 -1 Т -2 -1 0 -1 -4 -2 0 Α -5 -3 -1 0 ← -1 -3 0 -2 -2 -1 0 С -6 -4 -1 -7 Α -5 -3 -2 -2 0 0

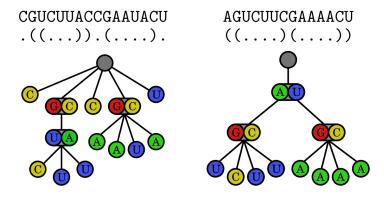
Tree-based Sequence Alignments

- gardenia
- RNA StrAT
- RNAdistance
- RNAforester

These are also extensively using the secondary structure for comparison, which is why they improve a lot with better curated secondary structures.

They are basically doing an edit-distance alignment with a metric that's also including the secondary structures and it's differences.

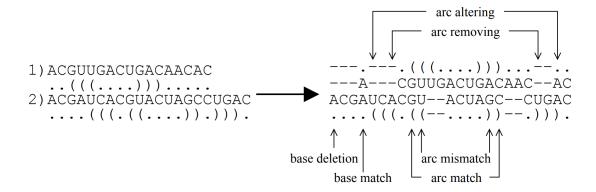
As edit operations on trees are basically edit operations on arc-annotated sequences, this is sometimes used to visualise it instead of some kind of trees.



Now there's more options for Edit-Distances, having the following possibilities:

- base match
- base indel
- arc match
- arc mismatch

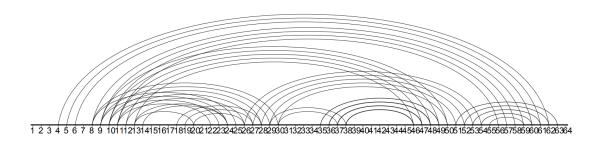
- arc breaking (missing arc)
- arc altering
- arc removing

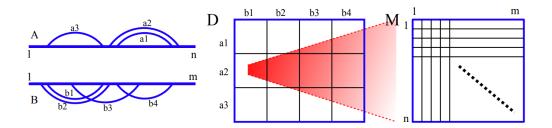


LocARNA

Is a folding and aligning sequence algorithm. First a base pair probability matrix is being created using RNAfold, this is then used to prune possible arcs. Only these with realistic chances are further considered.

$$p_{\rm cutoff}=0.1$$

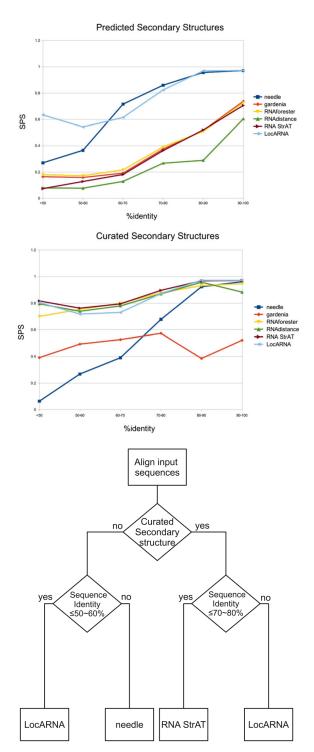




Then, Alignment scores for all of these realistic arcs are computed, caching smaller one's first. Meanwhile, the full path matrix is being filled out with values.



Comparison



 $\verb|https://github.com/fkarg/things-to-talk-about/tree/master/bioinfII|$