



RNA secondary structure visualization using tree edit distance

Richard Eliáš, David Hoksza

SIRET Research Group

Department of Software Engineering

Faculty of Mathematics and Physics

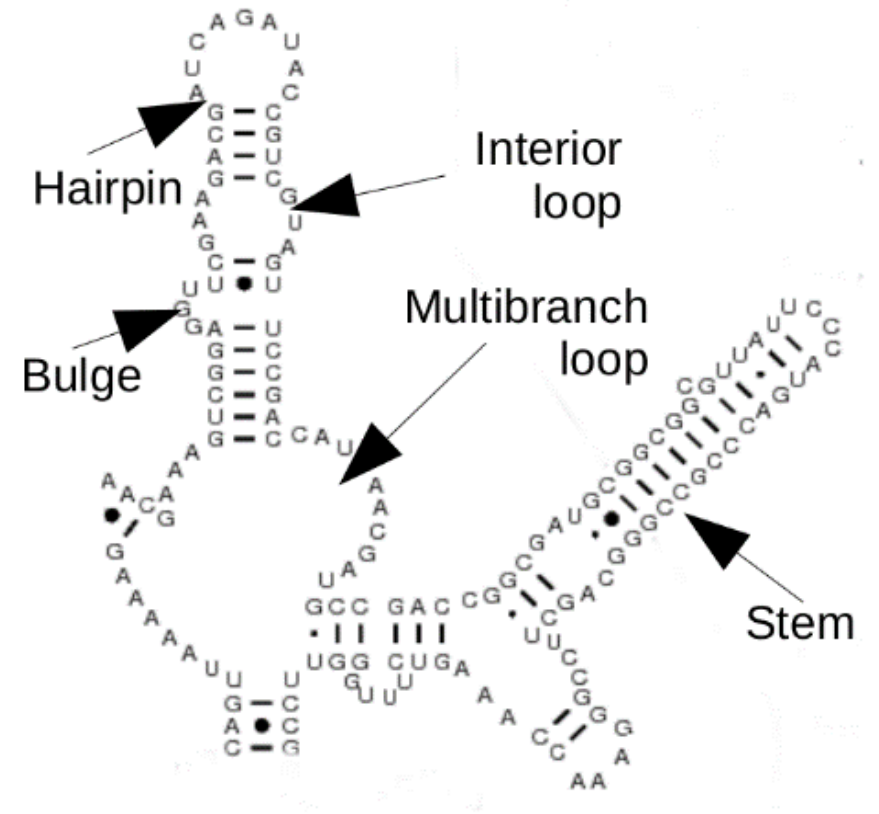
Charles University in Prague

Outline

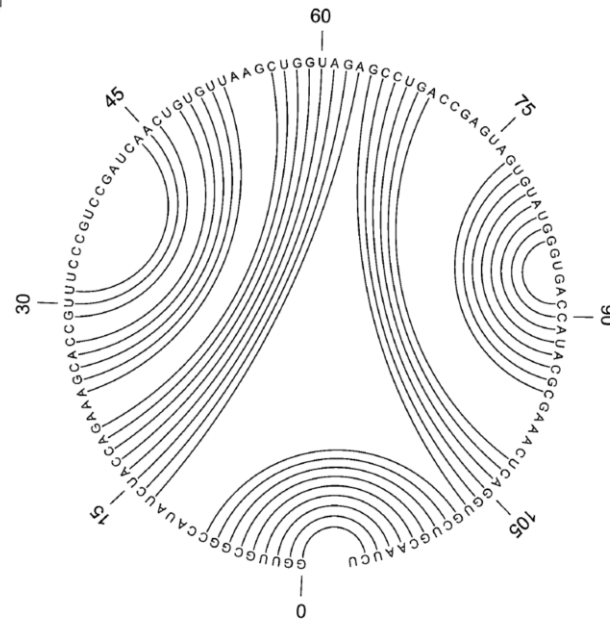
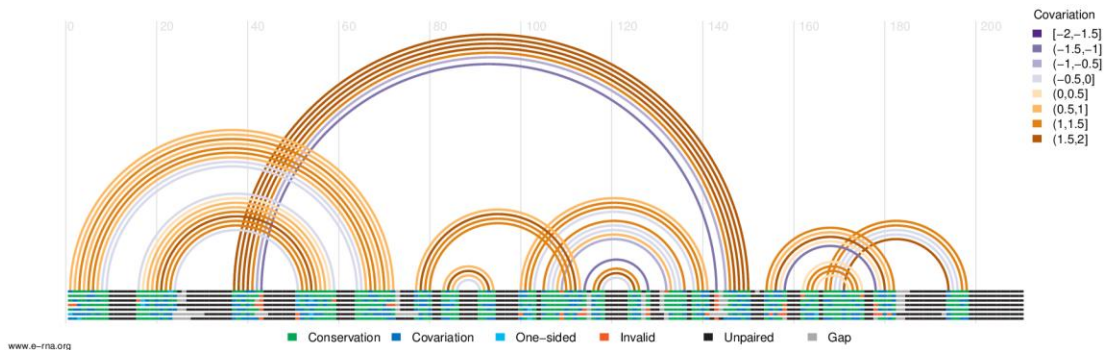
- Motivation
- Tree edit distance
- Template-based visualization algorithm
- Experimental results

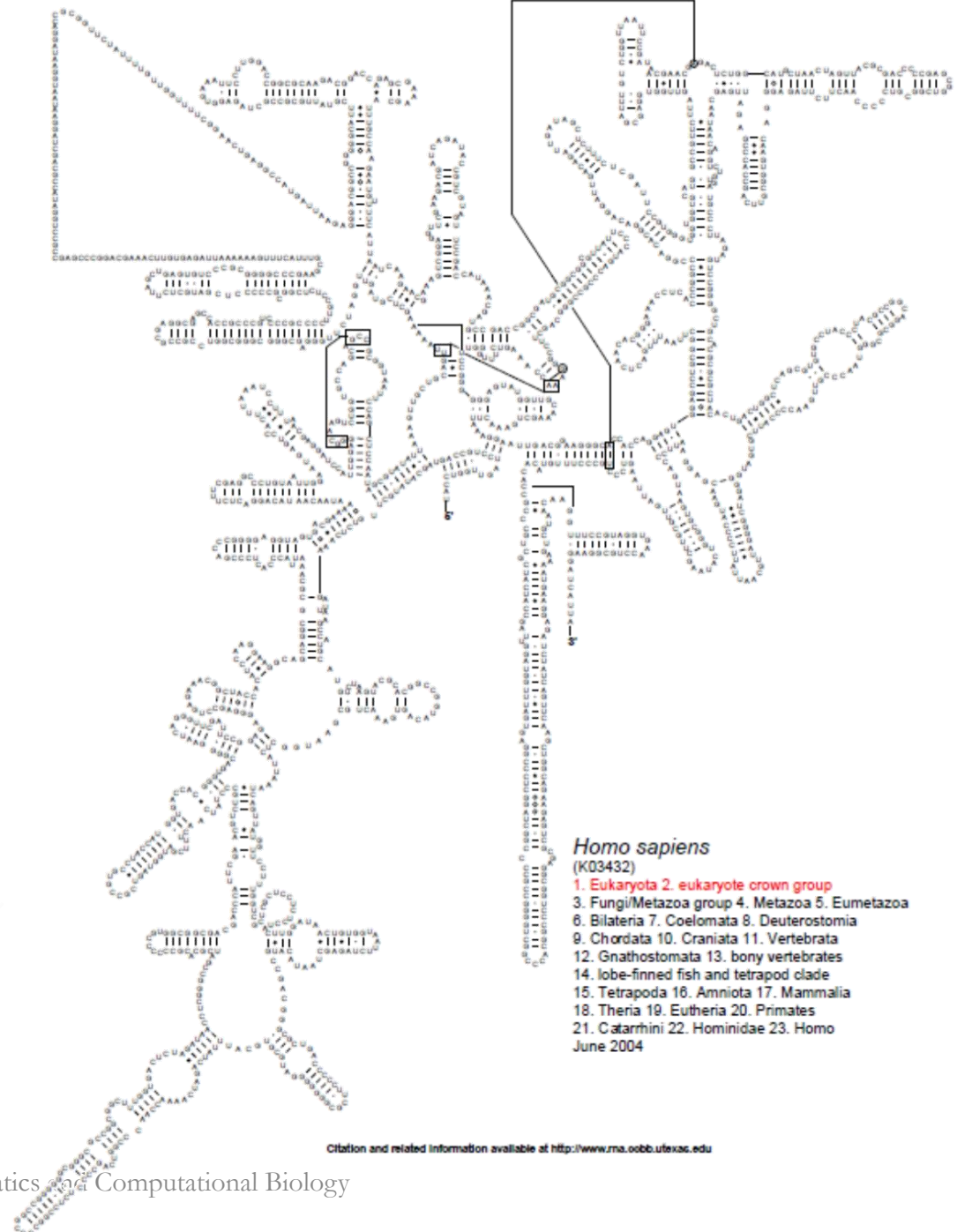
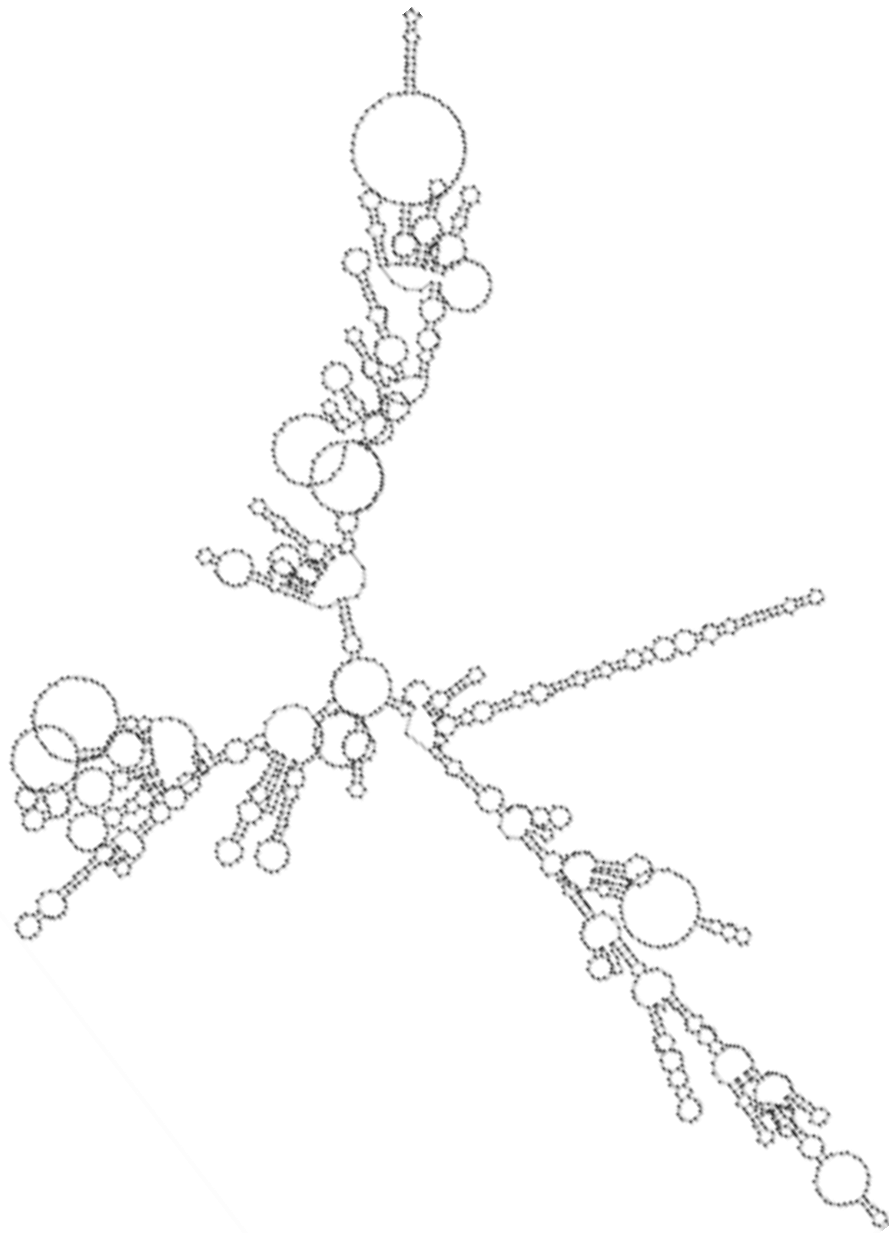
RNA secondary structure

- RNA pairs in the same way as DNA, but it is single-stranded
- Base-pairing interactions within a molecule
- Close approximation of the structure



RNA secondary structure visualization





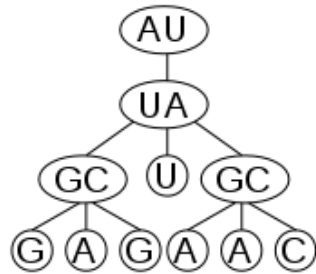
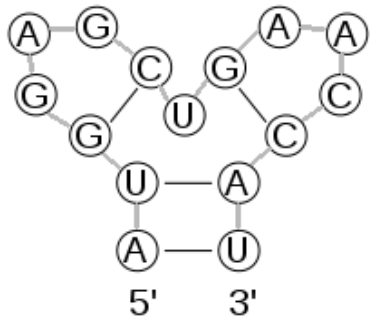
Citation and related information available at <http://www.rna.oobb.utexas.edu>

Algorithm outline

- Template-based visualization → preservation of **common motifs**
 - Template = homologous structure with known optimal layout
- 1. Convert input target and template structure into tree representation
- 2. Compute tree edit distance between template and target → sequence of tree edit operations
- 3. Map the tree edit operations to visual operations to convert template layout to target layout

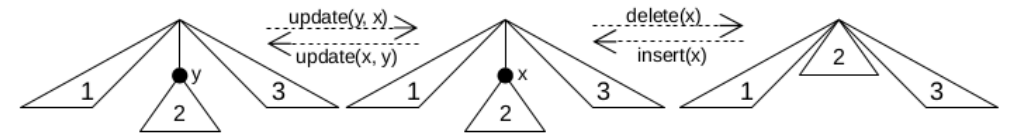
RNA Tree edit distance

- Structure \rightarrow tree
 - Base pairs \rightarrow inner nodes
 - Unpaired nucleotides \rightarrow leafs

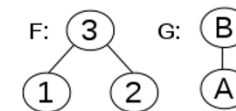


AUGGAGCUGGAACCAU

- Generalization of string edit distance
- Operations
 - **Update** – relabeling
 - **Delete** – deletion of a node and reconnection of children to the parent
 - **Insert** – insertion of a node between two connected nodes and reconnection of children



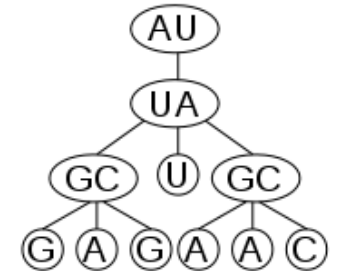
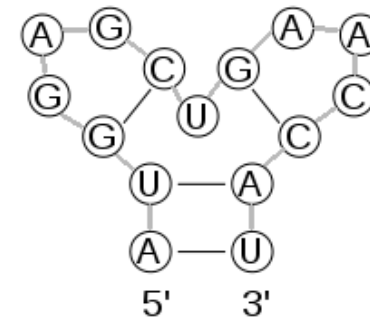
- Backtracking procedure resulting in sequence of operations



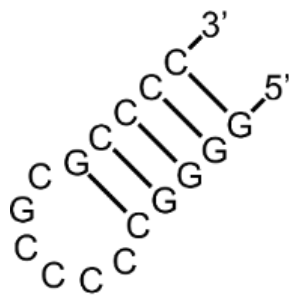
| F\G: | \emptyset | A | B |
|-------------|-------------|---|---|
| \emptyset | 0 | 1 | 2 |
| 1 | 1 | 0 | 1 |
| 1..2 | 2 | 1 | 2 |
| 3 | 3 | 2 | 1 |

Visual operations

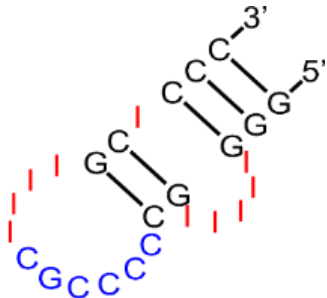
- Update
 - Relabeling
- Insert
 - Leaf node
 - No siblings → formation of a new loop
 - Existing siblings → loop extension → uniform distribution along a circle
 - Inner node
 - Insert base pair at given position
 - Shift all its “inner node” descendants
 - Recompute position of its possible leaf children on a circle
- Delete
- Multi-branch loops treated individually



Examples

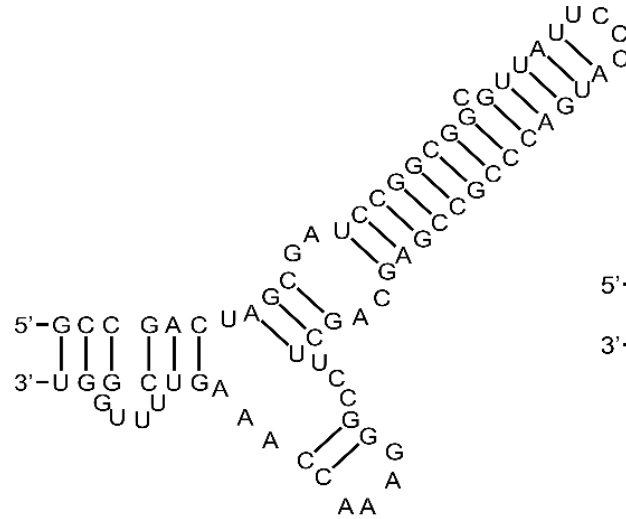


(a)

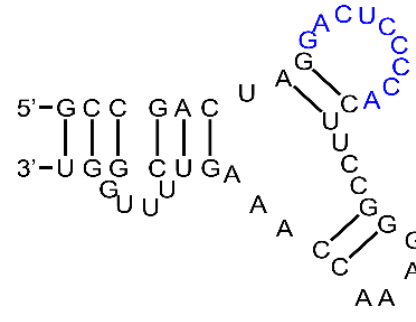


(b)

Insertion into both stem and loop parts of a hairpin

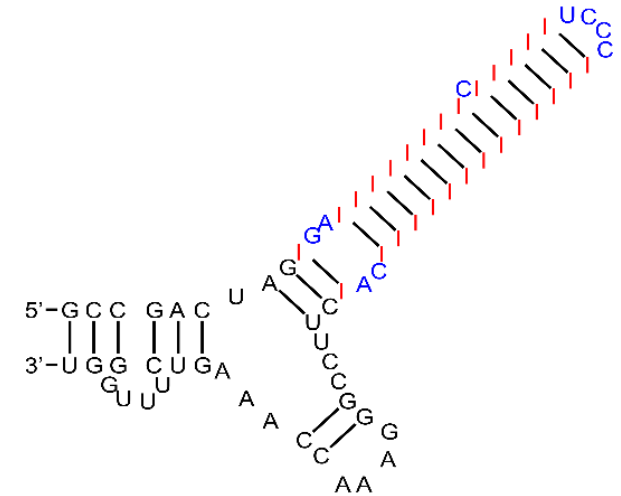


(a)



(b)

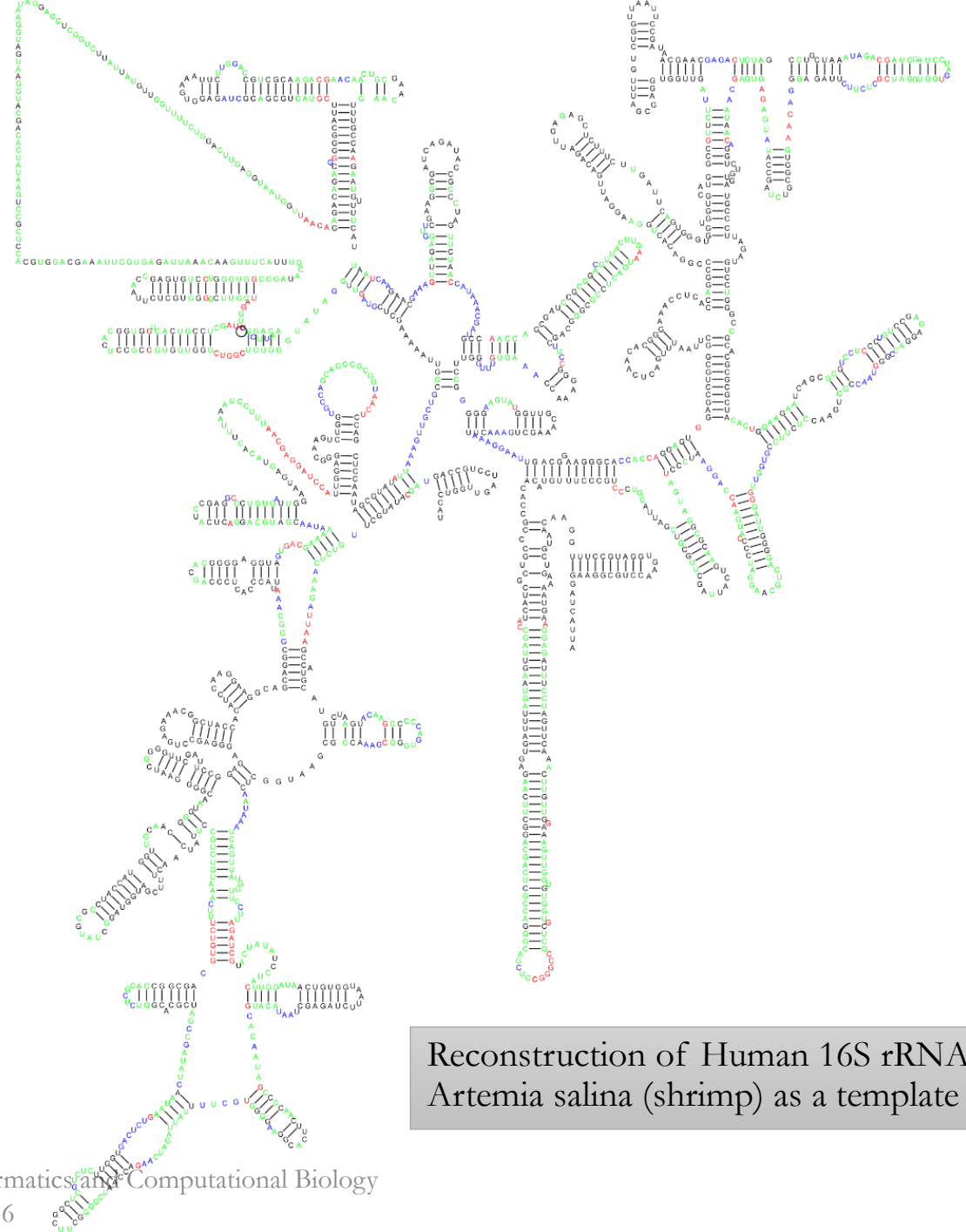
Substantial deletion and reinsert of one branch of a multibranch loop



(c)

Ribosomal RNA test

- Reconstruction of visualizations of known 16S ribosomal subunits from the Metazoa kingdom
- 16 organism
- Every pair of organisms tested → 272 layouts
- 3 crossings per layout on average



Reconstruction of Human 16S rRNA using *Artemia salina* (shrimp) as a template

Future work

- Recursion
- Web server
- Automatic selection of suitable templates

Questions?

