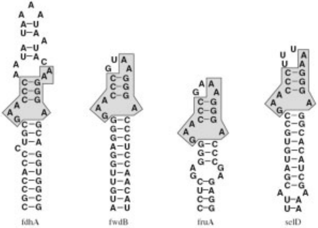
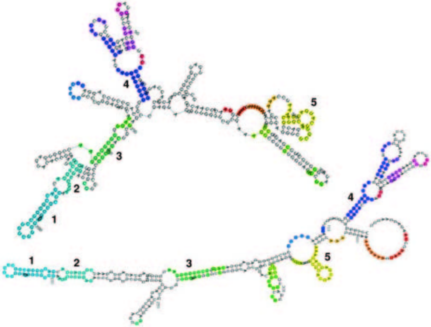
Maximal Exact Matchings

- similar to (local) alignment, but only identify parts that are exactly identical (no gaps)
- exact matches must be *connected* at sequence or structure level



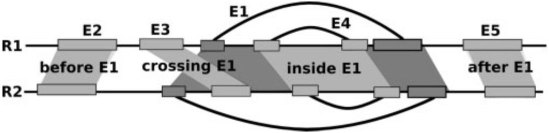


- faster than structure alignment: $O(n^2)$ time & space

Maximal Exact Matchings

expaRNA: compute alignments fast with help of exact matchings

- step 1: compute matchings
- step 2: chaining of matchings (select “chain” of *compatible* matchings, i.e. no overlap, no crossing)

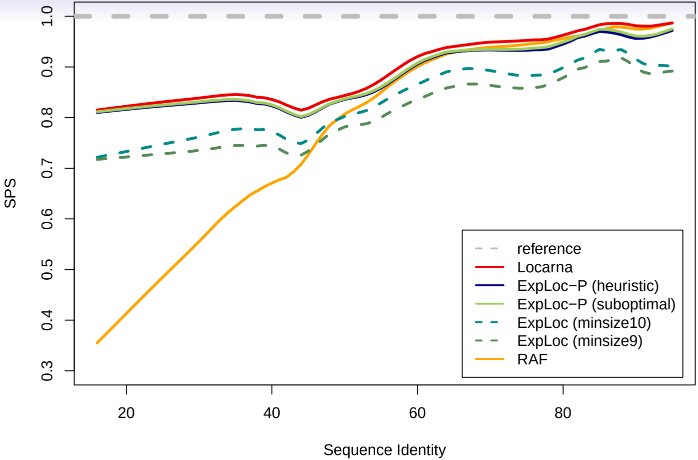


- step 3: compute alignment using chain of matchings as *anchor constraints* in LocARNA

→ speed-up over LocARNA

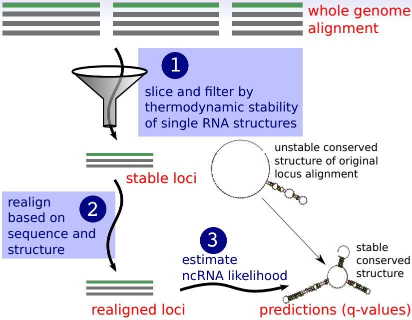
Steffen Heyne, Sebastian Will, Michael Beckstette, Rolf Backofen, Lightweight Comparison of RNAs Based on Exact Sequence-Structure Matches, Bioinformatics 2009

Bralibase 2.1-Benchmark



- ExpLoc: Exact matchings as anchors in LocARNA; 4.4, 5.4 times speed-up
- ExpLoc-P: Exact matchings *from structure ensembles* as anchors in LocARNA (submitted RECOMB'12; speed-up: 4.9, 6.0)
- RAF: Do et al., Bioinformatics 2008; speed-up 15.9

Whole Genome Realignment for ncRNA Prediction



whole genome alignment

1 slice and filter by thermodynamic stability of single RNA structures

stable loci

realign based on sequence and structure 2

realigned loci

3 estimate ncRNA likelihood

unstable conserved structure of original locus alignment

stable conserved structure

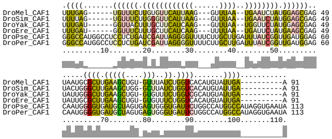
predictions (q-values)

```
             ............................................................
DroMel_CAF1  -----------------------TTTGAGTG-TTTCTTGTGTTCATTAAGGTTTAATGAA 36
DroSim_CAF1  -----------------------TTTTAGTG-TTTCTTGGGTTCATTAAGGTTTAATGAA 36
DroYak_CAF1  -----------------------TTTGATGG-TTACTTTGCTTCATCAAGGTTTAATGGT 36
DroEre_CAF1  -----------------------TTTGATGG-TTTCTTTGCTTCATCAAGTTTTAATGAT 36
DroPse_CAF1  GGGCCATGGCCTCCTCTGATCGATTAG-GGGTTTTCTTGCTTGATTTATCGGTTGATGGA 59
DroPer_CAF1  GGGCCATGGCCTCCTCTGATCGATTAG-GGGTTTTCTTGCTTGATTTATCGGTTGATGGA 59
             .........10........20........30........40........50.........

             ..((((..((.((....((((........))))....)).))..))))............
DroMel_CAF1  TCTATGGAGCGAGTAATGCGCTTGAAGCTGTGTTTATCTGGTCACATGTAT---TGA--A 91
DroSim_CAF1  TCTATGGAGCGAGTACTGGGCTTGAAGCTGGGCTTATCTGGTCACATGTAT---TGA--A 91
DroYak_CAF1  TCTATGGAGCGAGTATTGGGCTTGAAGCTGTGTGTTTCTGGTCGCATGTAT---TGA--A 91
DroEre_CAF1  TCTATGAAGCGAGTATTGCGCTTGAAGCTGTGTGTTTCTGGTCACATGTAT---TGA--A 91
DroPse_CAF1  GCAATGGGGTG----ATGCTAGTGA--GTGGGTGATTCTGGCCATGGCCATAGGTGAATA 113
DroPer_CAF1  GCAATGGGGTG----ATGCTAGTGA--GTGGGTGATTCTGGCCATGGCCATAGGTGAATA 113
             .........70........80........90.......100.......110........
```

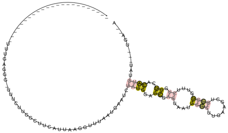
structure-based realignment

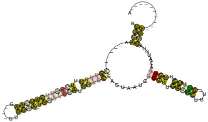
— — — — →



```
             .((((.......((((((.((((((((((((......))))..)))))))).))))))..
DroMel_CAF1  UUUGAG------UGUUUCUUGUGUUCAUUAAG---GUUUAA--UGAAUCUAUGGAGCGAG 49
DroSim_CAF1  UUUUAG------UGUUUCUUGGGUUCAUUAAG---GUUUAA--UGAAUCUAUGGAGCGAG 49
DroYak_CAF1  UUUGAU------GGUUACUUUGCUUCAUCAAG---GUUUAA--UGGUUCUAUGGAGCGAG 49
DroEre_CAF1  UUUGAU------GGUUUCUUUGCUUCAUCAAG---UUUUAA--UGAUUCUAUGAAGCGAG 49
DroPse_CAF1  GGGCCAUGGCCUCCUCUGAUCGAUUAGGGGUUUUCUUGCUUGAUUUAUCGGUUGAUGGAG 60
DroPer_CAF1  GGGCCAUGGCCUCCUCUGAUCGAUUAGGGGUUUUCUUGCUUGAUUUAUCGGUUGAUGGAG 60
             .........10........20........30........40........50.........

             .....((((.(((((....)))..)).)))).......))))...........
DroMel_CAF1  UAAUGCGCUUGAAGCUGU-GUUUAUCUGGUCACAUGUAUUGA----------A 91
DroSim_CAF1  UACUGGGCUUGAAGCUGG-GCUUAUCUGGUCACAUGUAUUGA----------A 91
DroYak_CAF1  UAUUGGGCUUGAAGCUGU-GUGUUUCUGGUCGCAUGUAUUGA----------A 91
DroEre_CAF1  UAUUGCGCUUGAAGCUGU-GUGUUUCUGGUCACAUGUAUUGA----------A 91
DroPse_CAF1  CAAUGGGGUGAUGCUAGUGAGUGGGUGAUUCUGGCCAUGGCCAUAGGUGAAUA 113
DroPer_CAF1  CAAUGGGGUGAUGCUAGUGAGUGGGUGAUUCUGGCCAUGGCCAUAGGUGAAUA 113
             .........70........80........90.......100.......110.
```

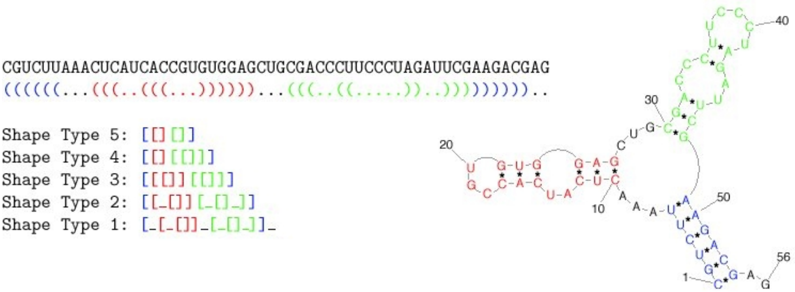




RNA Shapes : Idea

- A more coarse-grained look at RNA structure
- intuition: often general shape of RNA is more important for RNA function than “details”
- example: cloverleaf structure of tRNAs

Shape can be considered at different levels of abstraction

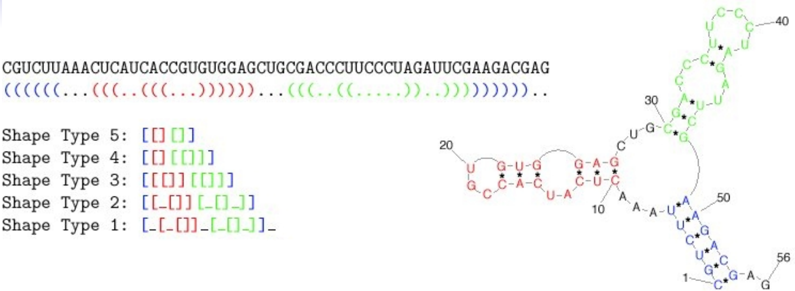


```
CGUCUUAAACUCAUCACCGUGUGGAGCUGCGACCCUUCCCUAGAUUCGAAGACGAG
((((((...(((..(((...))))))...(((..((.....))..)))))))))..

Shape Type 5: [[][]]
Shape Type 4: [[][[]]]
Shape Type 3: [[[]][[]]]
Shape Type 2: [[_[]][_[]_]]
Shape Type 1: [_[_[]]_[_[]_]]_
```

Robert Giegerich, Björn Voss, Marc Rehmsmeier, Abstract shapes of RNA, Nucleic Acids Research, 2004

RNA Shapes : different Shape Types



```
CGUCUUAAACUCAUCACCGUGUGGAGCUGCGACCCUUCCCUAGAUUCGAAGACGAG
((((((...(((..(((...))))))...(((..((.....))..)))))))))..

Shape Type 5: [[][]]
Shape Type 4: [[][[]]]
Shape Type 3: [[[]][[]]]
Shape Type 2: [[_[]][_[]_]]
Shape Type 1: [_[_[]]_[_[]_]]_
```

- 5 Most abstract - helix nesting pattern and no unpaired regions
- 4 Helix nesting pattern in internal loops and multiloops
- 3 Nesting pattern for all loop types but no unpaired regions
- 2 Nesting pattern for all loop types and unpaired regions in bulges, internal loops, and multiloops
- 1 Most accurate - all loops and all unpaired

RNAshapes: Computes shape probabilities for a sequence (+ Shrep = representative structure for each shape)