RILogo and treeMI

Stefan E Seemann, Peter Menzel, Jan Gorodkin
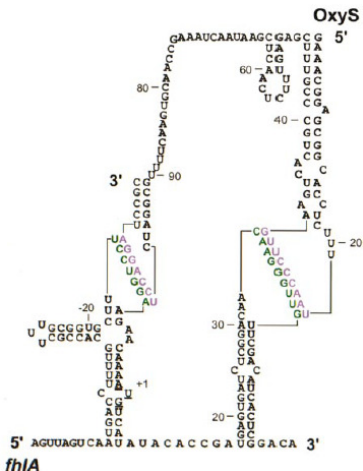
Center for non-coding RNA in Technology and Health (RTH),
University of Copenhagen

July 26, 2012

Motivation of RILogo

Visualize:

- interactions between RNA molecules
- sequence conservation
- structure conservation

(Argaman et al, JMB 2000)

RILogo

Plots

- arc representation RNA secondary structure and RNA-RNA interactions
- sequence logos
- measure of covariance as well as evolutionary-based (co-)variance measure

Standalone program:

https://github.com/pmenzel/RILogo/tarball/master

Webserver:

http://rth.dk/resources/rilogo

Measures of covariance

Standard measure of covariance of base pairs
(adaptation of the Kullback-Leibler divergence):

$$\mathrm{MI}_{ij} = O_{ij} \log_2 \frac{O_{ij}}{E_{ij}} + (1 - O_{ij}) \log_2 \frac{(1 - O_{ij})}{(1 - E_{ij})}$$

$$O_{ij} = \frac{1}{N} \sum_{s \in S} C_s \theta(s_i, s_j); \quad E_{ij} = \frac{1}{N^2} \sum_{s \in S} C_s \theta(s_i) \cdot \sum_{r \in S} C_r \theta(r_j)$$

$$C_s = 1$$

Adapted measure of covariance by Lindgreen et al. 2006:

$$\mathrm{MI\_WP} = O_{ij} \log_2 \frac{O_{ij}}{E_{ij}} - N_{ij}^{\mathrm{Gaps}} \times \beta$$

Evolutionary-based measure of (co-)variance

Drawback of MI and MI_WP:

- structurally neutral substitutions of one base within a base pair (e.g. G:C to G:U) do not contribute to the covariance measures
- phylogenetic bias in the sequence alignment can distort the individual contributions from each sequence
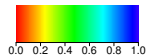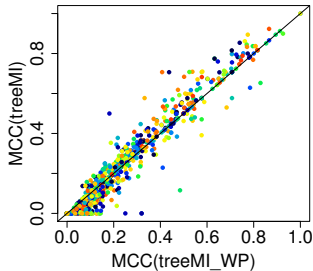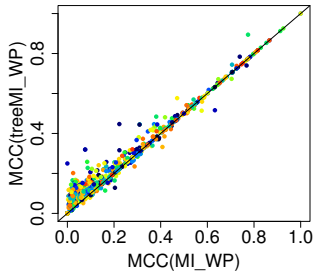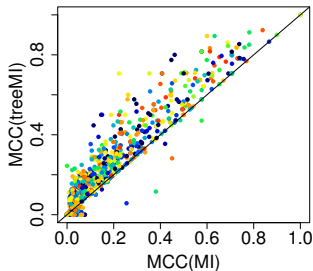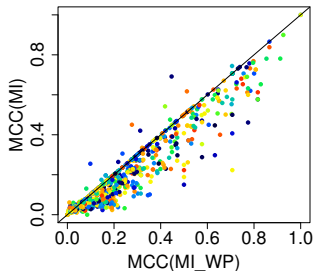
Solution:

- weigh base pair frequencies using pairwise phylogenetic distances between sequences.

treeMI: evolutionary-based (co-)variance measure

$$C_s = 1 - d_{\mathrm{avg}}(s)/N_d; \qquad N_d = \sum_s d_{\mathrm{avg}}(s)$$

Benchmark on 961 RNA alignments from Rfam 10.1

(Co-)variance of non-canonical basepairs

Consider non-canonical basepairs and their isostericity classes:

(Leontis et al, NAR 2002)

12 different basepair types with different classes of isosteric basepairs

Example basepair type 8:

```
tHH   I1   A-A   A-C   C-A   C-G   G-C
      I2   A-G   A-U   C-U   G-A   U-A   U-C
      I3   G-G
```

Bonus – Benchmark

average MCC:

	Vienna	MI	MIWP	treeMI	treeMIWP
Rfam 10.1	0.783	0.181	0.217	0.238	0.241

Bonus – RILogo

MIC