R-CHIE: A web server and R package for visualizing RNA secondary structure

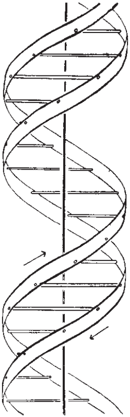
Irmtraud Meyer jointly with Daniel Lai, Jeff Proctor and Alice Zhu

Centre for High-Throughput Biology &
Department of Computer Science
University of British Columbia
meyer@chibi.ubc.ca

Benasque, summer 2012

A figure is worth a thousand words

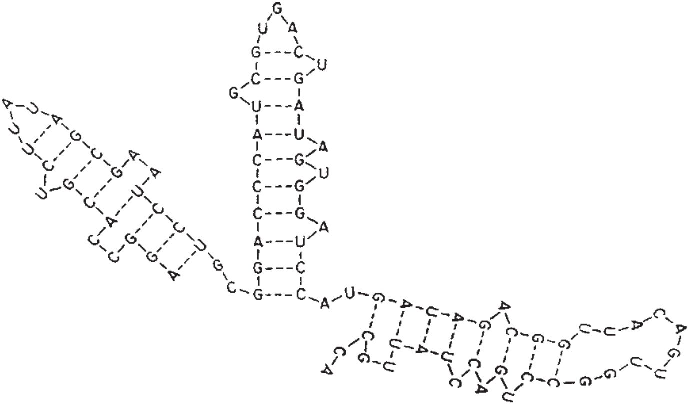
A figure is worth a thousand words



"Base-pairing"

Watson,J.D. and Crick,F.H. (1953)
Molecular structure of nucleic acids; a structure for deoxyribose nucleic acid. Nature, 171, 737-8.

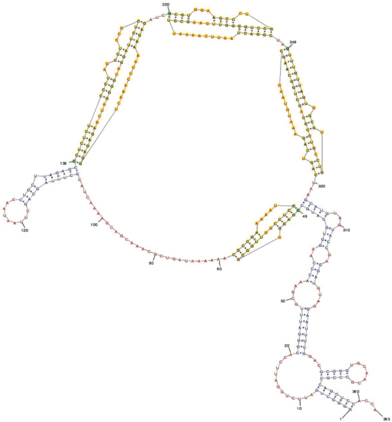
“Stem-loop” figures are the standard in RNA literature



One of the earliest “stems”

Fresco,J.R., Alberts,B.M. and Doty,P. (1960)
Some Molecular Details of the Secondary Structure of Ribonucleic Acid.
Nature, 188, 98-101.

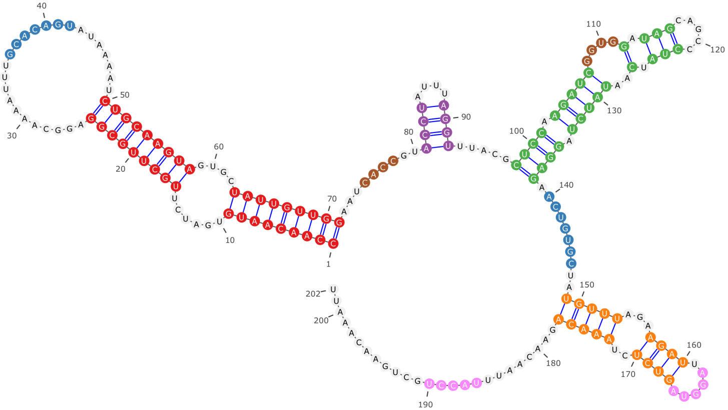
“Stem-loop” figures are the standard in RNA literature



The first automatic pseudoknotted stem-loop drawing program

Han,K., Lee,Y. and Kim,W. (2002)
PseudoViewer: automatic visualization of RNA pseudoknots.
Bioinformatics, 18 Suppl 1, S321-8.

“Stem-loop” figures are the standard in RNA literature

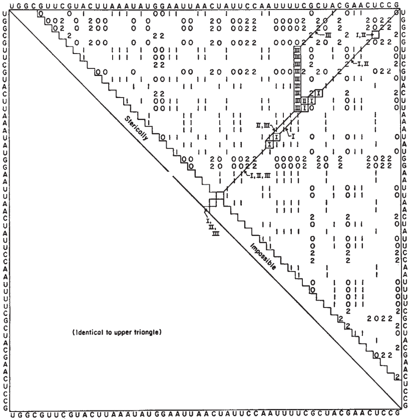


A recent lightweight Java Applet implementation

Darty,K., Denise,A. and Ponty,Y. (2009)
VARNA: Interactive drawing and editing of the RNA secondary structure.
Bioinformatics, 25, 1974-5.

Other less common diagrams exists...

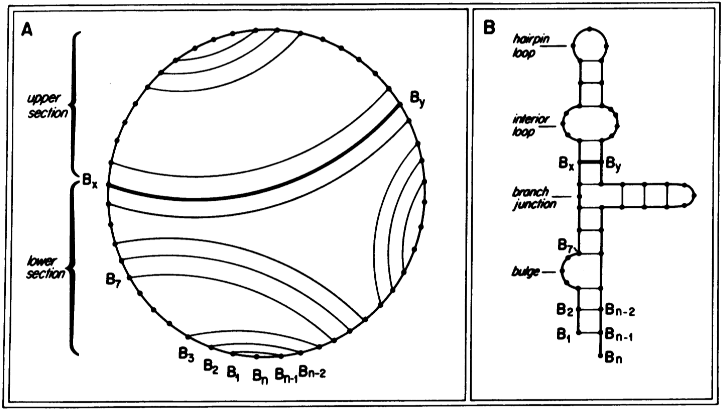
Basepairing matrices showing all possible pairings in cells



Old school RNA structure prediction by hand

Tinoco,I., Uhlenbeck,O.C. and Levine,M.D. (1971)
Estimation of secondary structure in ribonucleic acids. Nature, 230, 362-7.

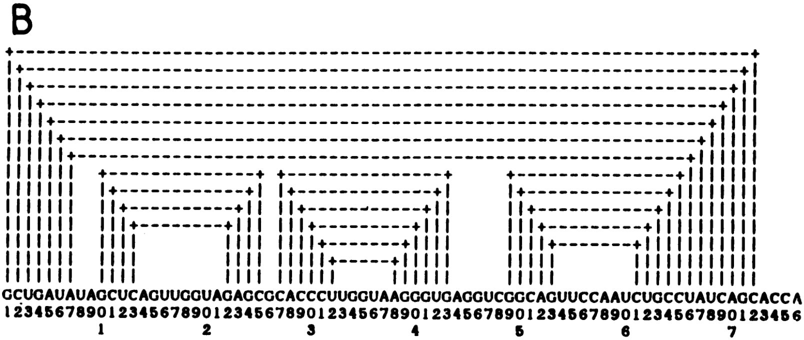
Circle diagrams showing basepairs as chords



Computer scientist + RNA == circle diagram

Nussinov,R. and Jacobson,A.B. (1980)
Fast algorithm for predicting the secondary structure of single-stranded RNA. PNAS, 77, 6309-13.

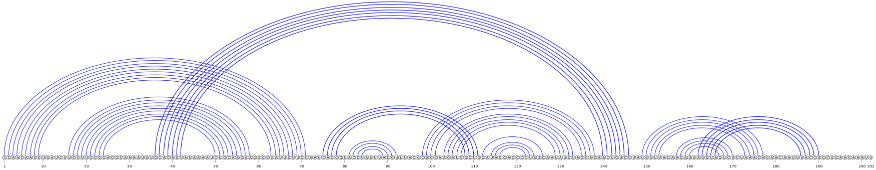
Linear digrams show basepairs as arcs



tRNA cloverleaf

Abrahams,J.P., van den Berg,M., van Batenburg,E. and Pleij,C. (1990) Prediction of RNA secondary structure, including pseudoknotting, by computer simulation. Nucleic Acids Research, 18, 3035-44.

Linear digrams show basepairs as arcs

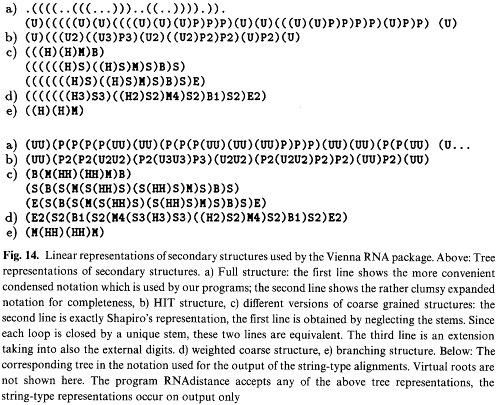


Pseudoknotted structure using VARNA

Darty,K., Denise,A. and Ponty,Y. (2009)
VARNA: Interactive drawing and editing of the RNA secondary structure.
Bioinformatics, 25, 1974–5.

Dot bracket “diagrams” showing basepairs

…with some other less known string representations

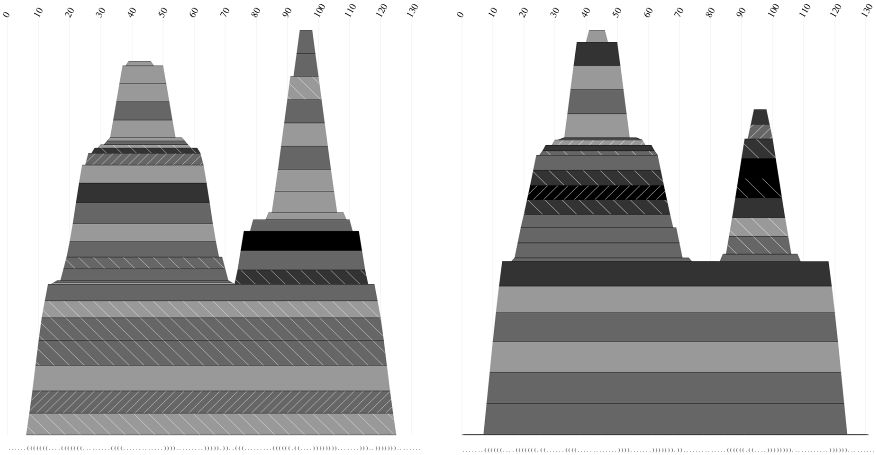


```
a) .((((..(((...)))..((..)))).)).
   (U)(((((U)(U)((((U)(U)(U)P)P)P)(U)(U)(((U)(U)P)P)P)P)(U)P)P) (U)
b) (U)(((U2)((U3)P3)(U2)((U2)P2)P2)(U)P2)(U)
c) (((H)(H)M)B)
   ((((((H)S)((H)S)M)S)B)S)
   (((((((H)S)((H)S)M)S)B)S)E)
d) (((((((H3)S3)((H2)S2)M4)S2)B1)S2)E2)
e) ((H)(H)M)

a) (UU)(P(P(P(P(UU)(UU)(P(P(P(UU)(UU)(UU)P)P)P)(UU)(UU)(P(P(UU) (U...
b) (UU)(P2(P2(U2U2)(P2(U3U3)P3)(U2U2)(P2(U2U2)P2)P2)(UU)P2)(UU)
c) (B(M(HH)(HH)M)B)
   (S(B(S(M(S(HH)S)(S(HH)S)M)S)B)S)
   (E(S(B(S(M(S(HH)S)(S(HH)S)M)S)B)S)E)
d) (E2(S2(B1(S2(M4(S3(H3)S3)((H2)S2)M4)S2)B1)S2)E2)
e) (M(HH)(HH)M)
```

Fig. 14. Linear representations of secondary structures used by the Vienna RNA package. Above: Tree representations of secondary structures. a) Full structure: the first line shows the more convenient condensed notation which is used by our programs; the second line shows the rather clumsy expanded notation for completeness, b) HIT structure, c) different versions of coarse grained structures: the second line is exactly Shapiro's representation, the first line is obtained by neglecting the stems. Since each loop is closed by a unique stem, these two lines are equivalent. The third line is an extension taking into also the external digits. d) weighted coarse structure, e) branching structure. Below: The corresponding tree in the notation used for the output of the string-type alignments. Virtual roots are not shown here. The program RNAdistance accepts any of the above tree representations, the string-type representations occur on output only

Hofacker,I.L., Fontana,W., Stadler,P.F., Bonhoeffer,L.S., Tacker,M. and Schuster,P. (1994)
Fast folding and comparison of RNA secondary structures. Monatshefte für Chemie Chemical Monthly, 125, 167-188.

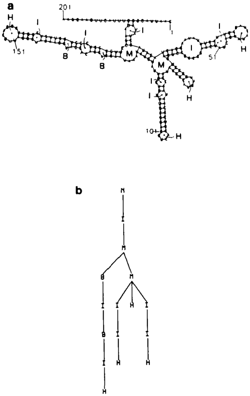
Moutain plots visualize dot bracket strings

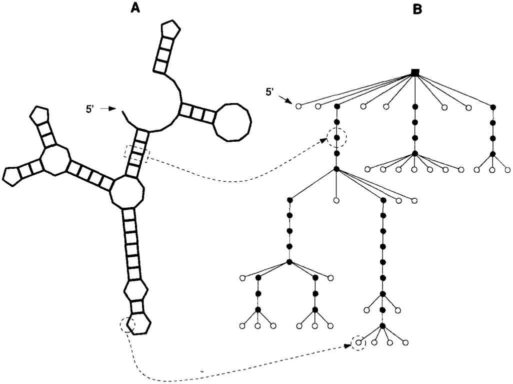


Hofacker,I.L. and Stadler,P.F. (1999)
Automatic detection of conserved base pairing patterns in RNA virus genomes. Computers & Chemistry, 23, 401-14.

Finally, some representations primarily used by Bioinformaticians . . .

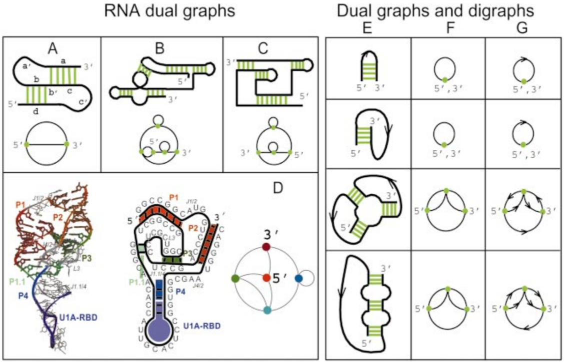
An unknotted structure is mathematically a tree





Shapiro,B.A. (1988)
An algorithm for comparing multiple RNA secondary structures.
CABIOS, 4, 387-93.
Hofacker,I.L. and Stadler,P.F. (1999)
Automatic detection of conserved base pairing patterns in RNA virus genomes. Computers & Chemistry, 23, 401-14.

Any structure can be expressed as a graph

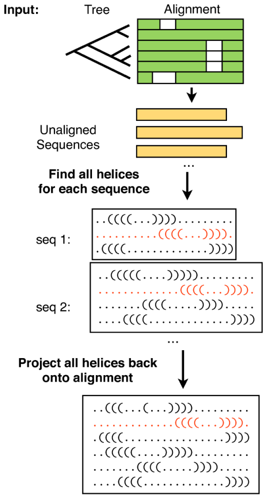


Gan,H.H. (2003)
Exploring the repertoire of RNA secondary motifs using graph theory; implications for RNA design. Nucleic Acids Research, 31, 2926-2943.

The woes of TRANSAT

The woes of TRANSAT

Find All Helices



Input: Tree Alignment

Unaligned Sequences

...

Find all helices for each sequence

seq 1:

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

seq 2:

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

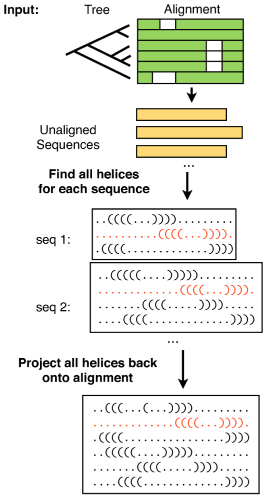
...

Project all helices back onto alignment

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

The woes of TRANSAT

Find All Helices



Input: Tree Alignment

Unaligned Sequences

...

Find all helices for each sequence

seq 1:

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

seq 2:

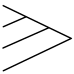
```
..(((((....))))).........
.............((((...)))).
.......((((.....)))).....
....((((.............))))
```

...

Project all helices back onto alignment

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

Score Helices

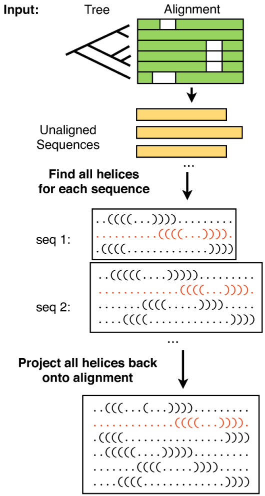


```
ACAGCAC-ACTGTTGT
ACAGCAT-ATTGTTGT
ACAACAT-ATTG-TGT
ACAACATTGTTG-TGT
```

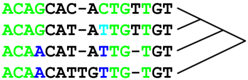
```
((((..))))......   -3.510
((((.....)))...    -1.911
((((.....))).)..   2.3017
(((((...)))).)..   1.1196
((((((....))))))   0.3987
...((((....).)))   1.8620
```

The woes of TRANSAT

Find All Helices



Score Helices

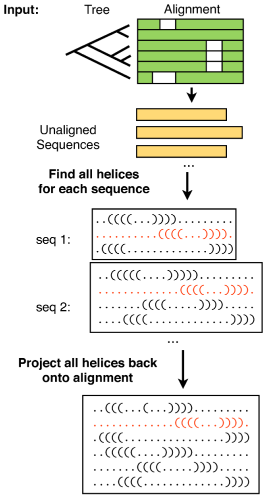


```
ACAGCAC-ACTGTTGT
ACAGCAT-ATTGTTGT
ACAACAT-ATTG-TGT
ACAACATTGTTG-TGT
```

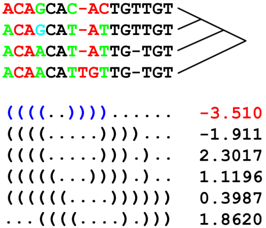
`((((..))))......`	-3.510
`((((.....))))...`	-1.911
`((((.....))).)..`	2.3017
`(((((...)))).)..`	1.1196
`((((((....))))))`	0.3987
`...((((....).)))`	1.8620

The woes of TRANSAT

Find All Helices



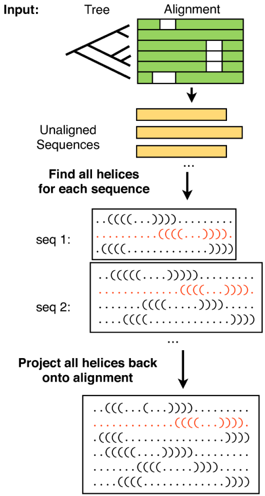
Score Helices



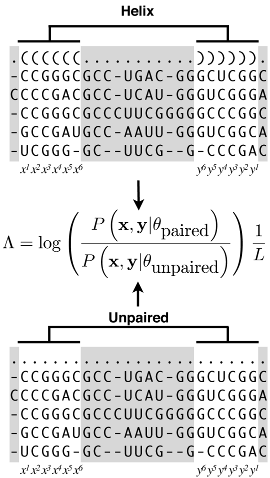
```
ACAGCAC-ACTGTTGT
ACAGCAT-ATTGTTGT
ACAACAT-ATTG-TGT
ACAACATTGTTG-TGT

((((..))))......   -3.510
((((.....))))...   -1.911
((((.....))).)..   2.3017
(((((...)))).)..   1.1196
((((((....))))))   0.3987
...((((....).)))   1.8620
```

The woes of TRANSAT

Find All Helices



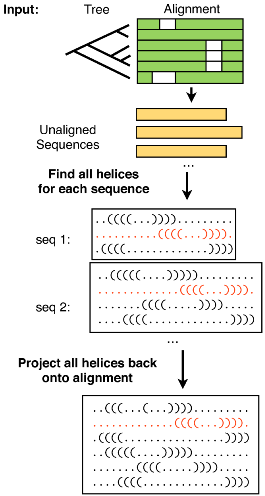
Score Helices



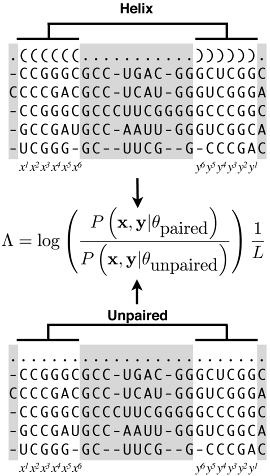
$$\Lambda = \log\left(\frac{P\left(\mathbf{x},\mathbf{y}|\theta_{\text{paired}}\right)}{P\left(\mathbf{x},\mathbf{y}|\theta_{\text{unpaired}}\right)}\right)\frac{1}{L}$$

The woes of TRANSAT

Find All Helices

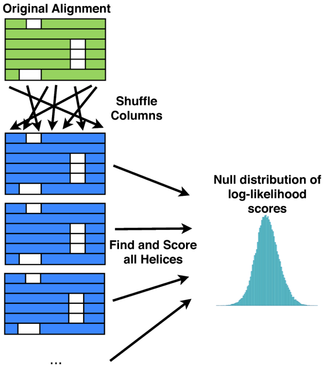


Score Helices



$$\Lambda = \log\left(\frac{P\left(\mathbf{x},\mathbf{y}|\theta_{\text{paired}}\right)}{P\left(\mathbf{x},\mathbf{y}|\theta_{\text{unpaired}}\right)}\right)\frac{1}{L}$$

Generate P-values

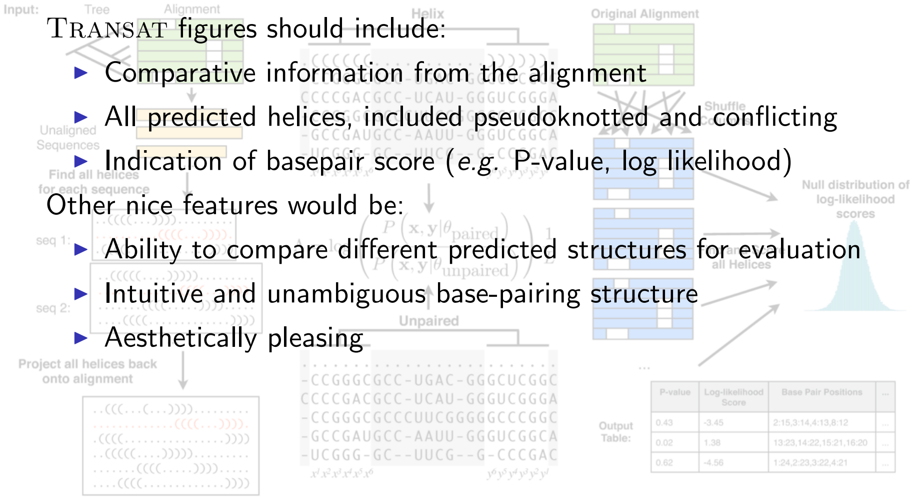


Output Table:

P-value	Log-likelihood Score	Base Pair Positions	...
0.43	-3.45	2:15,3:14,4:13,8:12	...
0.02	1.38	13:23,14:22,15:21,16:20	...
0.62	-4.56	1:24,2:23,3:22,4:21	...

Wiebe and Meyer, PLoS Comput Biol 2010

The woes of TRANSAT



TRANSAT figures should include:

- Comparative information from the alignment
- All predicted helices, included pseudoknotted and conflicting
- Indication of basepair score (*e.g.* P-value, log likelihood)

Other nice features would be:

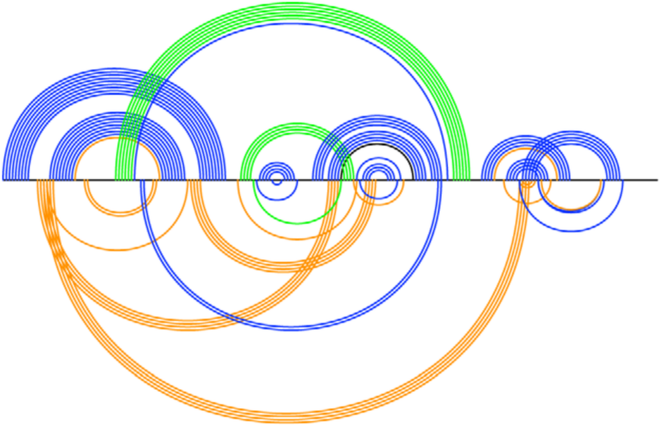
- Ability to compare different predicted structures for evaluation
- Intuitive and unambiguous base-pairing structure
- Aesthetically pleasing

Wiebe and Meyer, PLoS Comput Biol 2010

Linear diagrams seems to fulfill all requirements

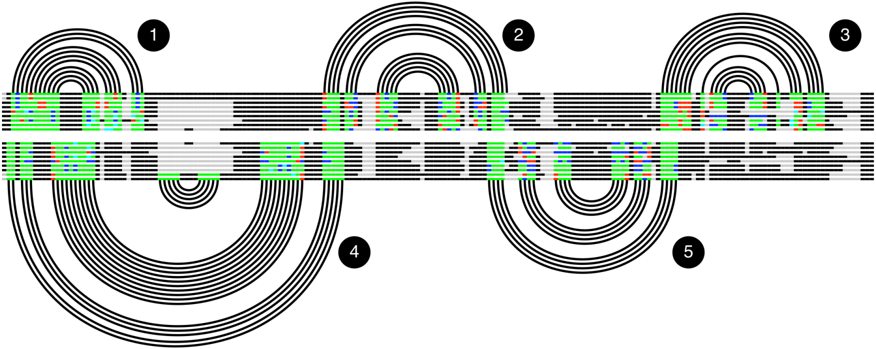
Feature	Stem-loop	Matrix	Circle	**Linear**	Bracket	Mountain	Tree	Graph
Comparative MSA	0	0	?	1	1	1	0	0
All basepairs	0	1	1	1	0	0	0	0
Basepair score	1	1	1	1	1	0	0	0
Easy comparison	0	0	1	1	1	1	1	1
Intuitive	1	0	1	1	1	0	0	0
Aesthetic	1	?	1	1	0	0	?	?

The prototype solution: “arc-plots”



Wiebe,N.J.P. and Meyer,I.M. (2010)
TRANSAT– method for detecting the conserved helices of functional RNA structures, including transient, pseudo-knotted and alternative structures.
PLoS Computational Biology, 6, e1000823.

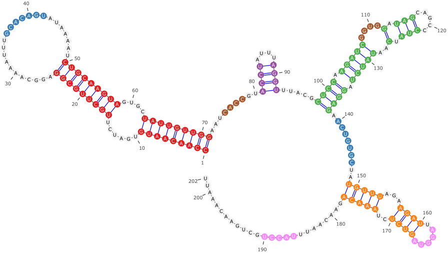
The prototype solution: “arc-plots”

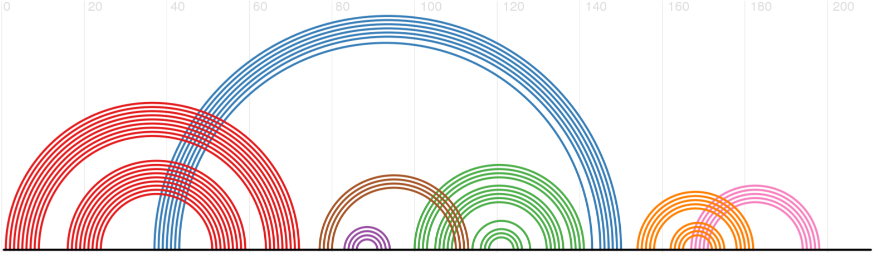


Wiebe,N.J.P. and Meyer,I.M. (2010)
TRANSAT– method for detecting the conserved helices of functional RNA structures, including transient, pseudo-knotted and alternative structures.
PLoS Computational Biology, 6, e1000823.

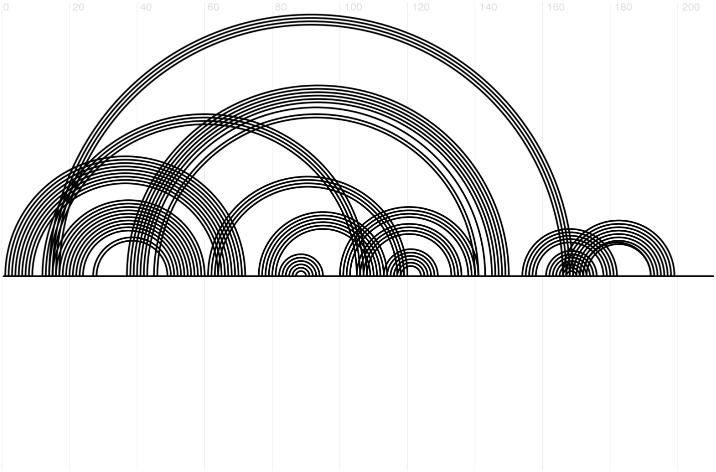
R-chie

A websever and R package for RNA secondary structure visualization



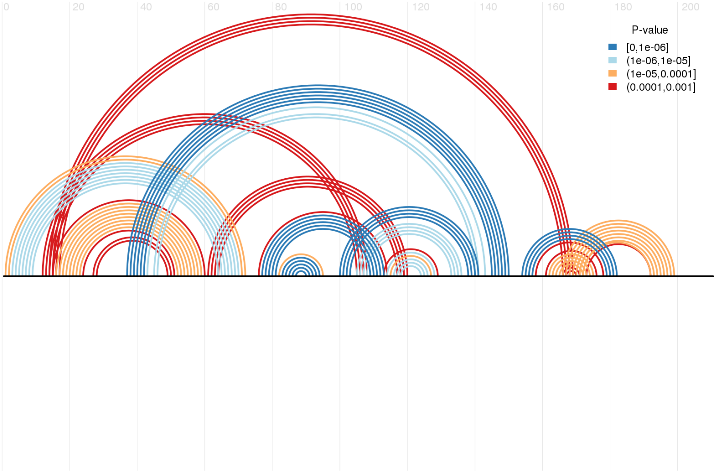


The base case is simply the linear RNA diagram



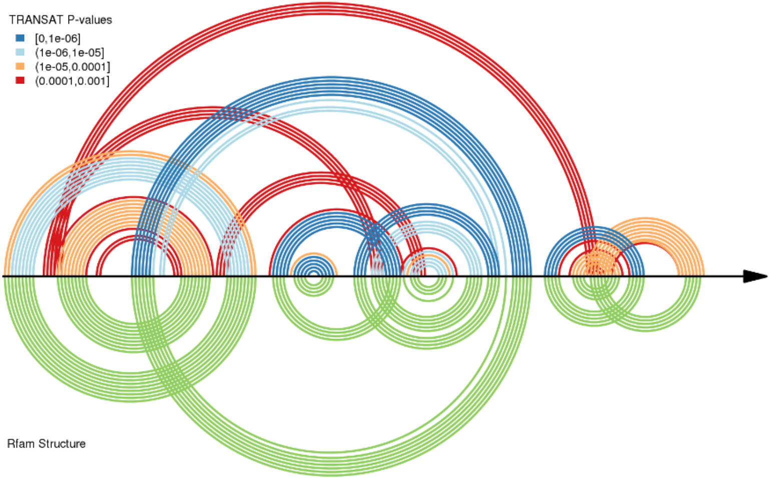
Pseudoknotted and conflicting basepairs can easily be drawn

R provides palettes and filters to colour basepairs



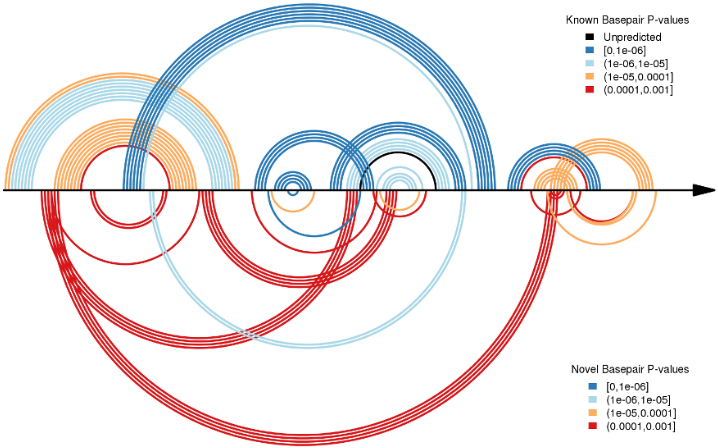
Arcs coloured by TRANSAT prediction P-values

Our package easily allows for two simultaneously structures



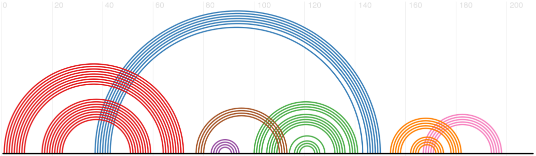
TRANSAT predictions vs. the known RFAM structure

Sensitivity/specificity diagrams are a novelty of our package



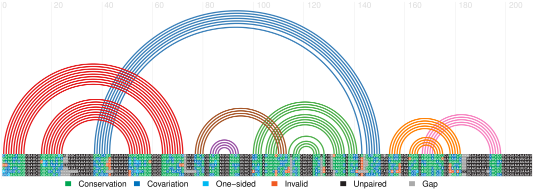
Sensitivity (top) and specificity (bottom) of the TRANSAT prediction

Arc diagram + MSA = Covariation plot

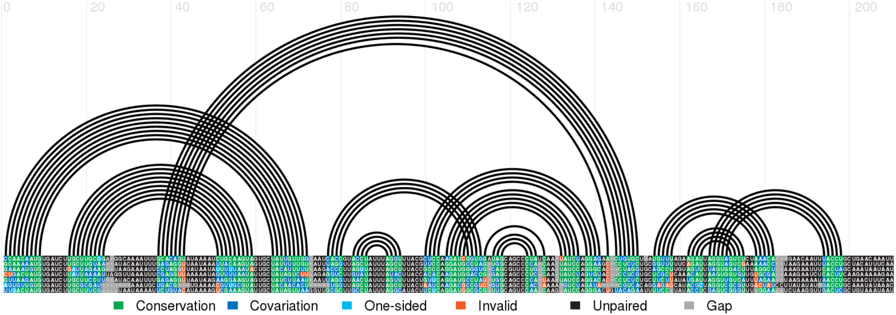


+

=

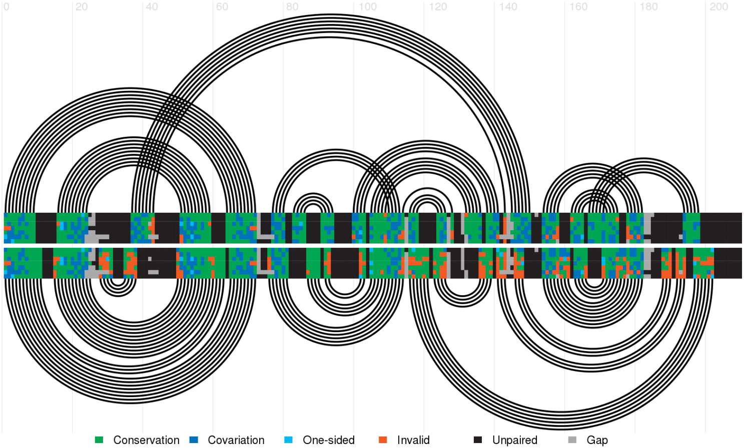


We can see how well a known structure is conserved



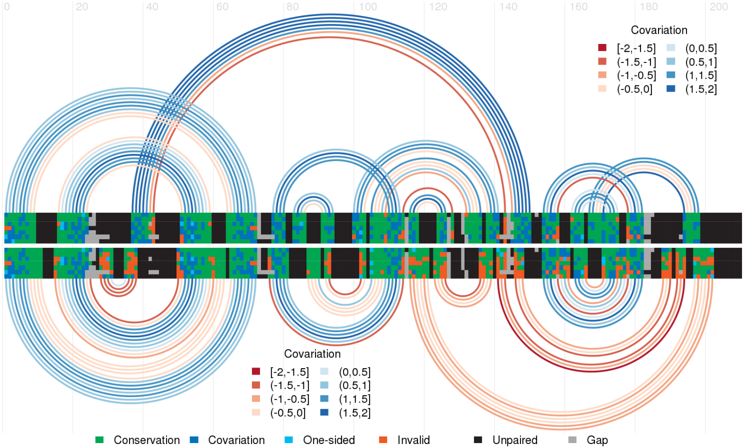
Conservation of the known structure of RF00458 from RFAM

We can see how well a prediction is supported by evolutionary evidence



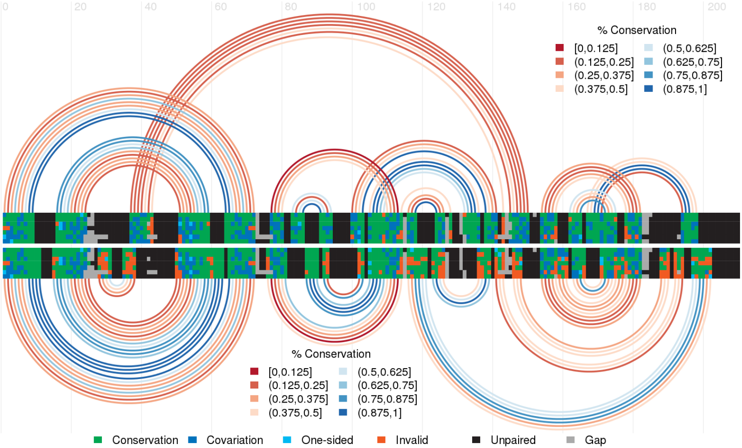
Known structure (top) vs. MFOLD prediction (bottom)

We can see how well a prediction is supported by evolutionary evidence



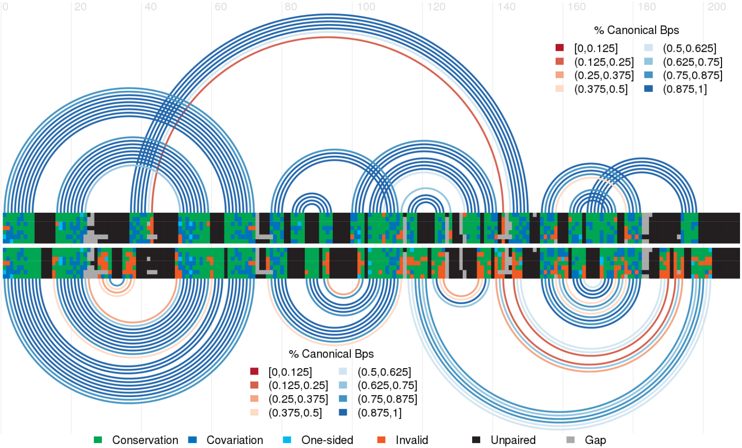
Known structure (top) vs. MFOLD prediction (bottom)

We can see how well a prediction is supported by evolutionary evidence



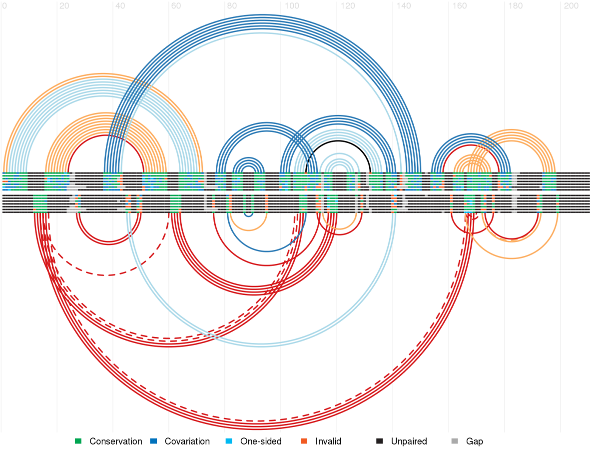
Known structure (top) vs. MFOLD prediction (bottom)

We can see how well a prediction is supported by evolutionary evidence



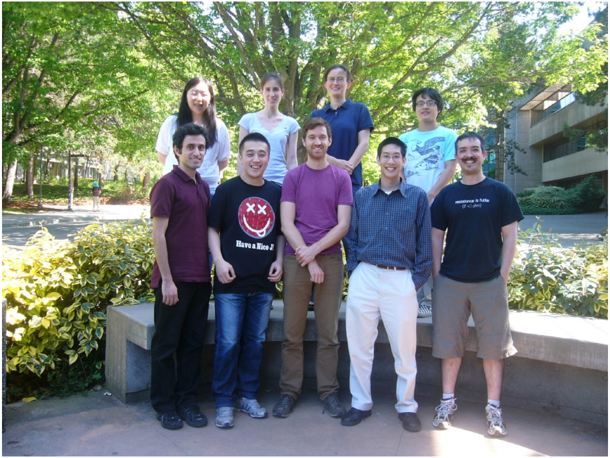
Known structure (top) vs. MFOLD prediction (bottom)

We can see how well false positives are conserved



Covariance + TRANSAT overlapping with a known structure (sens/spec plot)

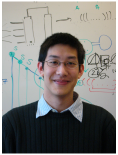
Acknowledgements:



My group (not shown is Evan, the photographer).

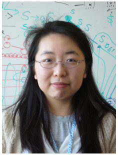
Acknowledgements:

Daniel Lai



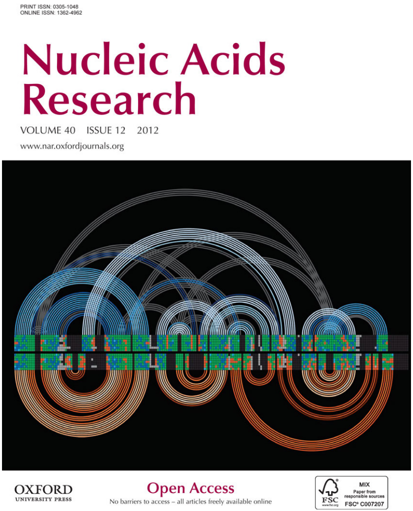
Jeff Proctor

Alice Zhu



Funding:

- Canadian Foundation for Innovation
- CIHR, Canada
- NSERC, Canada



PRINT ISSN: 0305-1048
ONLINE ISSN: 1362-4962

Nucleic Acids Research

VOLUME 40 ISSUE 12 2012

www.nar.oxfordjournals.org

OXFORD UNIVERSITY PRESS

Open Access

No barriers to access – all articles freely available online

MIX Paper from responsible sources FSC® C007207

http://www.e-rna.org/r-chie

- Nucl. Acids Res. (2012) 40 (12): e95. doi: 10.1093/nar/gks241
- R-package R4RNA is freely available for downloading
- plots for all Rfam families available on web-page