BIOINFORMATICS APPLICATIONS NOTE Vol. 16 no. 11 2000
Pages 1042–1043

Circles: automating the comparative analysis of RNA secondary structure

Roderic D.M. Page

Division of Environmental and Evolutionary Biology, Institute of Biomedical and Life Sciences, Graham Kerr Building, University of Glasgow, Glasgow G12 8QQ, UK

Received on May 21, 2000; revised on July 4, 2000; accepted on July 7, 2000

Abstract

Summary: *Circles is a program for inferring RNA secondary structure using maximum weight matching. The program can read in an alignment in FASTA, ClustalW, or NEXUS format, compute a maximum weight matching, and export one or more secondary structures in various file formats.*

Availability: *The program is available at no cost from http://taxonomy.zoology.gla.ac.uk/rod/circles/ and requires Windows 95/98/NT.*

Contact: *r.page@bio.gla.ac.uk*

Comparative analysis is the preferred and most successful method of inferring RNA secondary structure (Gutell *et al.*, 1994; Konings and Gutell, 1995), however a major practical limitation of this technique is that it requires substantial manual effort. Readily available software programs simply output listings of possible pairings based on mutual information or other criteria (Brown, 1993; Hall, 1999), or display pairwise graphs of mutual information values (Gorodkin *et al.*, 1999). Converting this information into a structure is not a trivial task (Gutell *et al.*, 1992).

Maximum weight matching (MWM) is graph-theoretic approach to inferring RNA secondary structure that shows considerable promise (Cary and Stormo, 1995; Tabaska *et al.*, 1998). It takes as input a set of base pairing scores that can be derived from a range of sources, such as free energy considerations, mutual information, or experimental data. These data can be represented as a folding graph where the vertices are alignment positions, and the lines or edges connecting a pair of vertices are given a weight proportional to the amount of evidence for that pairing. A matching is a subgraph of the folding graph in which no vertex is connected to more than one other vertex. The matching with the greatest total edge weight is taken to be the best estimate of the RNA structure. MWM can greatly speed up the process of comparative structure analysis of RNA sequences, and has been tested successfully on tRNA, SRP RNA, and 16S rRNA sequences (Tabaska *et al.*, 1998). UNIX command line programs to compute the MWM have been made available by Tabaska *et al.* (1998) (ftp://ftp.cshl.org/pub/science/mzhanglab/tabaska/). However, these programs are limited to importing one sequence file format (FASTA), and do not provide tools to visualise the results, or easily extract secondary structures in standard formats.

Circles is designed to simplify the use of maximum weight matching by providing a graphical user interface to the MWM programs. Tabaska's have been ported to Windows, and incorporated into a graphical program that adds further capabilities. Circles can read and display alignments in FASTA, ClustalW and NEXUS formats. The user can compute folding graphs for either mutual information or helix plot scores (Tabaska *et al.*, 1998), or both, and can easily explore the effects of various options and weightings on the results.

The result of MWM analysis is a matching between base pairs. This matching can be visualised using a circle plot (Nussinov *et al.*, 1978) (Figure 1). By default, Circles displays only pairings involved in helices of at least two pairs (i.e. lone base pairs are ignored). However, the user can interactively edit the pairings. RNA secondary structure is typically represented as a planar graph (i.e. a graph where no pair of edges overlap). Circles implements Nussinov *et al.*'s (1978) maximum loop matching algorithm to extract the largest planar RNA structure from the matching. Circles can display the secondary structure in a separate window, and it can also write RNA secondary structure files in a range of commonly used formats (MFOLD .ct, LOOPSOL, DCSE alignment) for display using other programs (de Rijk and de Wachter, 1997; Matzura and Wennborg, 1996).

Acknowledgements

I thank Jack Tabaska for advice on using his programs, and Paul Lewis for his Nexus Class Library. This work was supported by NERC grant GR3/11075, and was completed while I was a University of Auckland 2000 Foundation Visitor to the School of Biological Sciences, University of Auckland.

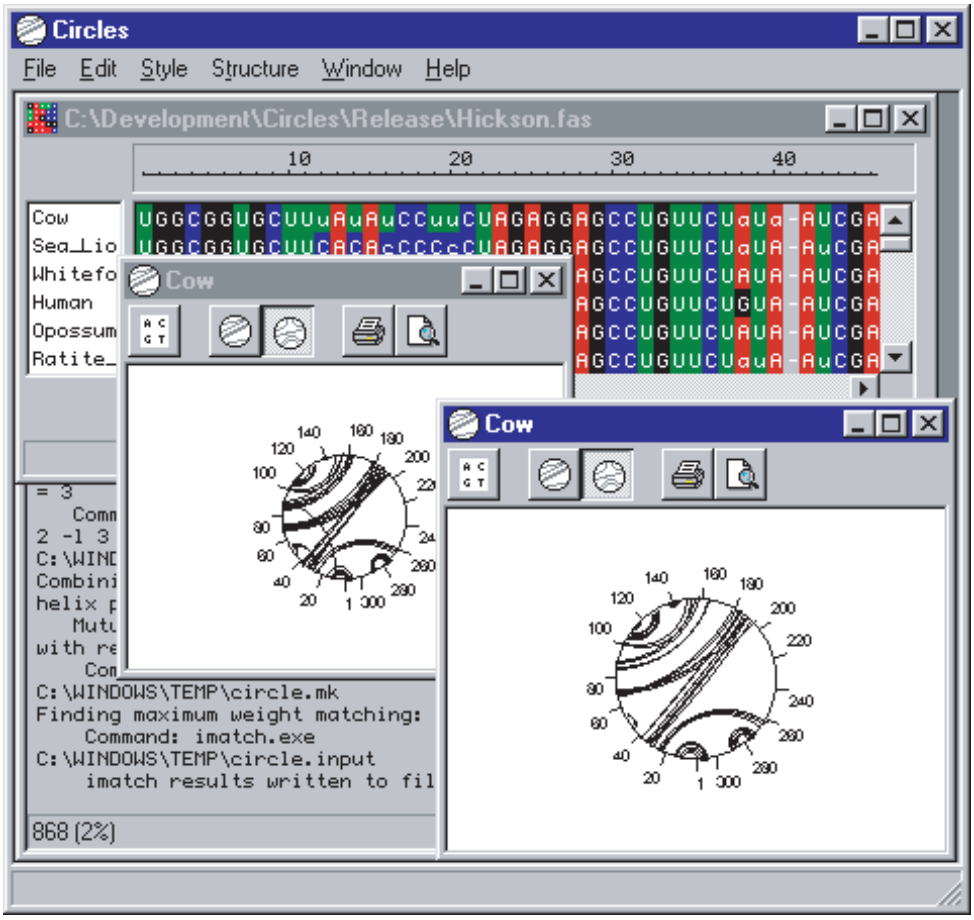

Fig. 1. Screen shot of the program Circles. Visible are the log window recording results, a sequence alignment display window, and two circle plots for the Cow sequence. The left circle plot displays the maximum weight matching, the right circle plot displays the largest planar structure that can be extracted from this matching.

References

Brown,J.W. (1993) A Macinthosh Hypercard compliation of small subunit ribosomal RNA sequences. *Comput. Appl. Biosci.*, **9**, 473.

Cary,R.B. and Stormo,G.D. (1995) Graph-theoretic approach to RNA modelling using comparative data. In Rawlings,C., Clark,D., Altman,R., Hunter,L., Lengauer,T. and Wodak,S. (eds), *Proceedings of the Third International Conference on Intelligent Systems for Molecular Biology* AAAI Press, Menlo Park, CA, pp. 75–80.

de Rijk,P. and de Wachter,R. (1997) RnaViz, a program for the visualisation of RNA secondary structure. *Nucleic Acids Res.*, **25**, 4679–4684.

Gorodkin,J., Stærfeldt,H.H., Lund,O. and Brunak,S. (1999) MatrixPlot: visualizing sequence constraints. *Bioinformatics*, **15**, 769–770.

Gutell,R.R., Power,A., Hertz,G.Z., Putz,E.J. and Stormo,G.D. (1992) Identifying constraints on the higher-order structure of RNA: continued development and application of comparative sequence analysis methods. *Nucleic Acids Res.*, **20**, 5785–5795.

Gutell,R.R., Larsen,L.N. and Woese,C.R. (1994) Lessons from an evolving rRNA: 16S and 23S rRNA structures from a comparative perspective. *Microbiol. Rev.*, **58**, 10–26.

Hall,T.A. (1999) BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/NT. *Nucleic Acids Symp. Ser.*, **41**, 95–98.

Konings,D.A.M. and Gutell,R.R. (1995) A comparison of thermodynamic foldings with comparatively derived structures of 16S and 16S-like rRNAs. *RNA*, **1**, 559–574.

Matzura,O. and Wennborg,A. (1996) RNAdraw: an integrated program for RNA secondary structure calculation and analysis under 32-bit Microsoft Windows. *Comput. Appl. Biosci.*, **12**, 247–249.

Nussinov,R., Pieczenik,G., Griggs,J.R. and Kleitman,D.J. (1978) Algorithms for loop matchings. *SIAM J. Appl. Math.*, **35**, 68–82.

Tabaska,J.E., Cary,R.E., Gabow,H.N. and Stormo,G.D. (1998) An RNA folding method capable of identifying pseudoknots and base triples. *Bioinformatics*, **14**, 691–699.