Visualizing ncRNA Structural Evolution

Alyssa Tsiros
Worcester Polytechnic Institute

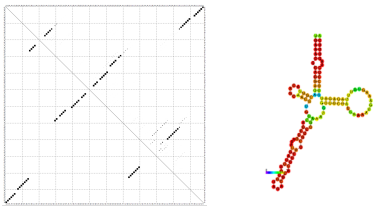
Abstract

This year's Biovis Symposium design challenge offered two challenges, the second of which is addressed here. The second design challenge is to create a visualization of noncoding RNA (ncRNA) structural evolution of the human accelerated region 1 (HAR1) gene in ancestral, denisovan, and human sequences. The Biovis Symposium design challenge #2 figure does not display the direct comparison base pairings among the three structures, whereas the new design "Relative distance vs. MFE base pairing for ancestral, denisovan, and human HAR1 genes" clearly highlights conserved or otherwise evolved base pairings. The direct comparisons of unique and common base pairings allows viewers to clearly identify differences in the structure predictions.

Keywords: Biovis Symposium, design challenge, structure prediction, noncoding RNA, structural evolution.

1. Introduction

Today, computational biologists are able to predict secondary structures from primary RNA sequences. These secondary structures are determined by a collection of predicted base pairs that minimize the free energy of the fold (Zuker, 1995), which is also known as the minimum free energy (MFE) structure. The visualization of this secondary structure typically comprises a "dot-plot", or a grid of possible base pairs (or non-pairs) at each nucleotide residue position (The W.C. Ray Lab, 2015), and a secondary structure graph of the MFE structure (Figure 1).



1. Dot-plot, or probabilities of base pair bindings, and secondary structure graph ("Design Contest", 2015).

There are several RNAs transcribed by the human genome known as non-coding RNAs, or ncRNAs, that do not code for proteins but still impact cellular processes. Until recent years, it has been understood that that most pertinent genetic information was obtained from proteins ("Non-coding RNA", 2006). Evidence suggests that the majority of mammal genomes is transcribed into ncRNAs (Makunin and Mattick, 2006). Such evidence has created a need to classify ncRNAs, where each class has been shown to have a characteristic secondary structure (Arora et al., 2014). Moreover, small changes in a primary RNA sequence can impact its structure and overall function, motivating molecular and structural biologists to compare secondary structures of evolved sequences.

atsiros@wpi.edu
774-275-0458

2. Methods

Several bioinformatics graduate students, a biovisualization professor, an evolutionary biologist professor, and a genetics professor were asked to comment on the Biovis Symposium design. The general consensus was that the new design should still clearly identify the different stems of the secondary structure prediction, but also provide a direct comparison of how the base pairings evolve from ancestral to human.

The presented redesign on the following page is titled "Relative distance versus minimum free energy base pairing for ancestral, denisovan, and human human accelerated region 1 genes". Each of the provided sequences have unique and common MFE structure base pairings. To highlight these pairings, a color was assigned to unique pairings, common pairings among ancestral and denisovan, common pairings among ancestral and human, common pairings among denisovan and human, and common among all three.

The first, middle, and last positions of the sequence are indicated in the figure. Moreover, relative distance was calculated in the following way, where k_1 and k_2 are the first and second positions of a MFE binding, respectively and L is the length of the sequence:

$$(| k_1 - k_2 | + 1) / L$$

The three graphs in the figure display half of an ellipse between two binding positions. Nothing is displayed if a position does not bind. The blue pairings are conserved among ancestra, denisovan, and human genes. The green pairings are conserved from ancestral to denisovan. The red pairings are common among ancestral and human, the fuscia pairings are conserved from denisovan to human, and finally, unique pairings are shown in black.

References

1. (2015). Design challenge. 5th Symposium on Biological Data Visualization. Retrieved from http://www.biovis.net/year/2015/design-contest
2. Arora, A., Panwar, B., Raghava, G. PS. (2014). Prediction and classification of ncRNAs using structural information. BMC Genomics, 2014 (15): 127. Retrieved from http://www.biomedcentral.com/1471-2164/15/127
3. Makunin, I. V., Mattick, J. S. (2006) Non-coding RNA. Human Molecular Genetics, 15 (suppl 1): R17 - R29. Retrieved from http://hmg.oxfordjournals.org/content/15/suppl_1/R17.short
4. The W.C. Ray Lab. (2015, January 22). Understanding RNA folding energy dot-plots. [Video file]. Retrieved from https://www.youtube.com/watch?v=v1UbIUZ8k9o
5. Zuker, M. (1995). Prediction of RNA secondary structure by energy minimization. Washington University Institute for Biomedical Computing. Retrieved from http://mfold.rna.albany.edu/doc/old-mfold-manual/index.php