

Abstract

Characterising RNA Secondary Structure Space:

With the advent of next-generation sequencing technologies and new methods in transcriptomics, an explosively growing amount of biological RNA data is available in public databases. Key to understanding function and regulatory effects of RNA, structure prediction is still difficult, despite many efforts. At the heart of the problem is unreliability in RNA secondary structure prediction, the best methods gaining about 60% sensitivity on single-stranded RNA sequences. What is more, many structural predictions are relied upon as correct in biological applications. Efforts must continue to be made to better understand the distribution of predictive accuracy and when structure prediction fails.

Here we explore the landscape of RNA secondary structure space, particularly with a view to predictive quality, considering where structural signals can be found. Novel phylo-SCFG entropy methods will be presented, and these compared with current PPfold reliability scores to demonstrate various failings in secondary structure prediction. Further factors will be explored including changes in alignment quality and various evolutionary distances, allowing for greater quantitative understanding of accuracy of resulting secondary structure predictions.