

# Analyzing the Operational RNA Code for Amino Acids - Using R

Tal Galili<sup>1</sup>, Shaul Shaul<sup>2</sup>, Yoav Benjamini<sup>3</sup>

It has been suggested that the nucleotide sequence of the tRNA acceptor stems specify an operational RNA code for amino acids. In the last 20 years several attributes of the putative code have been elucidated for a small number of model organisms. To gain insight about the ensemble attributes of the code, we used R to analyze the 6350 tRNA sequences from 102 Bacterial and 55 Archaeal species.

Our study found that the RNA codes, in both Archaea and Bacteria, differs substantially from the genetic code in the degree of degeneracy and specificity. We found instances of taxon-specific alternative codes, i.e., identical acceptor stems encrypting different amino acids in different species, as well as instances of ambiguity, i.e., identical acceptor stems encrypting two or more amino acids in the same species. Further investigation revealed that the species-specific degree of difference in the RNA code holds useful information, allowing us to distinguish between Archaea originating from different ecological environments and conditions.

In our research we employed various existing R facilities, performing data importing (Biostrings), cleaning and preparation (reshape), classification and regression tree (rpart), clustering (cluster), visualization (lattice) etc.

Also, we self developed (or implemented), in R, algorithms for tasks such as the comparing of hierarchical clusters, the creating of a distance matrix between Archaeas based on the code in their tRNA acceptor stems, performing simulation studies testing how well the "uniqueness property" is held by the organisms, based on their tRNA code.

In this talk I will provide (a brief) biological background needed in order to understand the above discoveries, and present how we used R various packages and statistical methods, while devising and implementing new methods (in R), in order to support our investigation.

## Reference

Shaul S, Berel D, Benjamini Y, Graur D.(2010), "Revisiting the operational RNA code for amino acids: Ensemble attributes and their implications", *RNA*. 2010 Jan;16(1):141-53. Epub 2009 Dec 1.,

**Correspondence should be addressed to: Tal.Galili@gmail.com**

---

<sup>1</sup> Department of Statistics and Operations Research, Tel Aviv University, Israel.

<sup>2</sup> Department of Zoology, Tel Aviv University, Israel.

<sup>3</sup> Department of Statistics and Operations Research, Tel Aviv University, Israel.