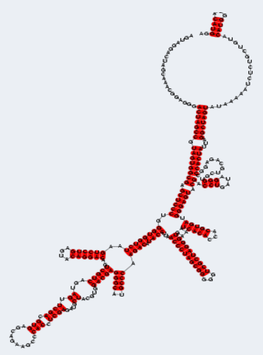
# Relationship between divergence of using synonymous codons in host-virus and the presence of microRNA

Franco Riberi[3], Laura Tardivo[3], Lucia Fazzi[2], Guillermo Biset[1], Daniel Gutson[1] and Daniel Rabinovich[1,2]

[1]FuDePAN - [2]INBIRS - [3] Universidad Nacional de Río Cuarto

## Background

MicroRNAs ($_{mi}$RNA) are small RNA that regulates the expression of $_{m}$RNA in the cells. They can interfere with viruses replication. In order to do this, it is necessary that the $_{mi}$RNA recognize genome target sites and that a pairing between the $_{mi}$RNA and a fragment of viral $_{mi}$RNA occurs. This recognition is more likely to happen if the fragment in the secondary structure of the viral miRNA is not paired (masked) [1]. It is known that the genome of some human viruses has a bias in the use of synonymous codons (different codons that encode for the same aminoacid) compared with the host even though its replication would be less efficient[2].



## Goal

The aim of the study is to determine if this bias could be the result of evolutionary pressure exerted by the $_{mi}$RNA. To achieve this goal massive comparisons will be made (in the order of 10e[7]) between the recognition of the virus natural genome and the "humanized" genome. The latter may be obtained by replacing codons in the viral genome, achieving a codon usage ratio similar to the host.

## Materials and Methods

For each $_{mi}$RNAs, the software to be developed will do parallel "sweeps" with the natural and humanized virus sequence. For each possible genome site, this program will determine the number of recognized nucleotides and whether these sites are available or masked by the secondary structure. When comparing results in homologous sites it can be determined whether $_{mi}$RNAs have a differential effect among different target $_{m}$RNAs (normal and humanized). The program will be coded using the C++ programming language and licensed under the GPLv3 software licence.
To calculate the matching score sequences the following formula is used:

$$\frac{(\#AT \times constAT + \#GC \times constGC)}{(totalAT \times constAT + totalGC \times constGC)} \quad (1)$$

where:

- **#AT:** amount of Adenine matching with Thymine, or vice versa.
- **#GC:** amount of Guanine matching with Cytosine, or vice versa.
- **constAT:** value for the pair A=T.
- **constGC:** value for the pair G=C.
- **total AT:** total Adenine and Thymine (paired or unpaired).
- **totalGC:** total Guanine and Cytosine (paired or unpaired).

**constAT** and **constGC** have different values depending on the score to be calculated. If the score of matching is calculated as a percentage, the constants have value 1, but if the score of matchig is calculated considering unions, values are used as mentioned in [**?**]. Currently, the program is under active development.
It is important to note that we use the software *geneDesign*[**?**] to obtain the humanized sequences.

## Results

For each $_{mi}$RNA and genome, a table should be produced that records, for each position, the matching $_{m}$RNA score, both in the original and the humanized sequence. Table 1 shows an example of a table estructure to be generated.

| | Original sequence | | | Humanized sequence | | | Score original sequence | | | | Score humanized sequence | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Position | Matching* | Masked+ | XYZ | Matching† | Masked‡ | XYZ | %const=1* | cFold* | %const=1+ | cFold+ | %const=1† | cFold† | %const=1‡ | cFold‡ |
| 1 | aaTTg<br>CacA | aaTTg<br>Maca | AaTTg<br>Xaca | ttAAC<br>Gtct | ttMAC<br>MtcM | ttYAC<br>YtcX | 0.44 | 0.45 | 0.22 | 0.24 | 0.55 | 0.54 | 0.11 | 0.21 |
| ... | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... |
| N | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... |

Table 1: Table structure to be generated. Where (cFold constAT = 1.25) and (cFold constGC = 0.95).

Also, it will be analized whether the results are in favor of the hypothesis of $_{mi}$RNAs selective pressure as a cause of bias in codon usage

## Some References

[1] Gareth M. Jenkins and Edward C. Holmes. **"The extent of codon usage bias in human RNA viruses and its evolutionary origin"**.

[2] Ulrike Muckstein, Hakim Tafer. **"Thermodynamics of RNA-RNA Interaction"**.

[3] Zuker, Michael. **"Computational Methods for RNA Secondary Structure"**.

[4] Sarah M. Richardson, Paul W. Nunley, Robert M. Yarrington, Jef D. Boeke and Joel S. Bader. **"GeneDesign 3.0 is an updated synthetic biology toolkit"**.

## At a glance

- **Problem:** determine the miRNA hibridization ability both in the viral secuence and in the humanized secuence.
- **Idea:** count the number of $_{mi}$RNA that hybridize to a viral and humanized secuence. For each genome secuence (viral and humanized), calculate two matching scores.
- **Results:** Table 1.

{friberi, lauratardivo}@dc.exa.unrc.edu.ar, lucia.fazzi@gmail.com, billybiset@gmail.com, daniel.gutson@fudepan.org.ar, rabinovichra@yahoo.com.ar