

# The rare lncRNA GOLLD is widespread and structurally conserved among *Mycobacterium* tRNA arrays

Sergio Mascarenhas Morgado, Deborah Antunes, Ernesto Raul Caffarena, Ana Carolina Paulo Vicente

*Oswaldo Cruz Institute*

## Abstract

Noncoding RNA (ncRNA) genes produce transcripts involved in catalytic or regulatory functions, some of them presenting highly complex structures. GOLLD RNA is the third-largest bacterial ncRNAs known (800 bp); however, its function is still unknown. The GOLLD RNA gene is generally found associated with tRNA genes and supposed to be chromosome- and phage-encoded in bacteria from Lactobacillales and Actinomycetales orders. Besides, the only inferred GOLLD RNA structure was mainly based on metagenomic sequences. To explore GOLLD in bacterial genomes, we mined GOLLD gene in thousands of *Mycobacterium* and virus genomes using Infernal software, identifying it in 350 mycobacteria (including two in megaplasms) and 39 virus genomes, mainly associated with tRNA arrays. *Mycobacterium* GOLLD genes are highly diverse and distributed in three clades: *Mycobacterium* exclusive; *Mycobacterium* and mycobacteriophages; and mycobacteriophage exclusive. We also determined the secondary structure of each clade using R2R software based on GOLLD alignments generated by Infernal software. All clades displayed a 3' half conserved structure including utter E-loops pseudoknots substructures, also shared by non-*Mycobacterium* GOLLD while the 5' half motif was different among the clades. In some cases, an ORF, coding a tRNA or transposase gene, was predicted in the 5' half motif. Moreover, in vitro assays determined the expression of GOLLD RNA gene present in a plasmid harbored by a *Mycobacterium* isolate from Atlantic Forest soil. Our study showed that the long ncRNA GOLLD is widespread within *Mycobacterium* in association with tRNA arrays, besides strengthening its structure, previously predicted based in metagenomic sequences.

Funding: Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq); Coordenação de Aperfeiçoamento de Pessoal de Nível Superior -Brasil (CAPES); Oswaldo Cruz Institute