

# Comparative genomics of R-body determinants

Gabriel Sánchez Hueck, Robson Francisco de Souza

## Abstract

Refractile (R) bodies are potential toxin-delivery proteic structures able to stretch and puncture cell membranes on a pH-dependent manner, whose presence have been verified in some free living and endosymbiont bacteria, but the role and genomic context of reb genes responsible for their polymerization is not yet understood. RebB domain genes appear on genomes of proteobacteria, bacteroidetes and acidobacteria, generally displaying several contiguous duplications, but operon configuration exhibits varied forms and distribution. R-bodies have been shown to require the presence of distinct reb family members, and sometimes even non-homologous genes, to ensure polymerization. Regulatory genes have been demonstrated, but their taxonomic distribution and domains have not been fully determined. The precise order of events guiding horizontal transfer and duplications in rebs and neighboring genes also remains undefined. We intend to explore the presence of reb homologs and identify new genes associated with R-body function, allowing the inference of their evolutionary history. Potential genes identified on our work will serve as candidates for experimental characterization in future essays, with special interest on of toxins and antitoxins. At the core of our strategy we make use of iterative sequence alignment tools such as jackhmmer and Psi-Blast for collection of homologs, domain recognition tools, such as hmmscan and hhsearch for classification and labeling, mmseqs as a tool for grouping sequences and reducing redundancy based on the e-value and identity of pairwise alignments, and FastTree and figtree for the construction and visualization of inferred phylogenies. We have also made use of tools for gene locus and neighborhood collection developed in our laboratory. Our main sources of data are the non-redundant RefSeq database for protein sequence collection and Pfam for models for domain description. We have found that reb family members have a conserved common core and variable regions on the N and C-terminal portions, which might account for the distinct polymerization events that occur during R-body assembly. We've created models based on the topology of the phylogenetic trees that the reb family displays, which has aided the visualization of reb loci and the distinct patterns displayed by them on genomes, and will be useful to describe events of horizontal transfer. We have additionally detected a high frequency of regulatory proteins neighboring rebs, and expect to identify, through the collection of homologs of known regulators as well as through mutation correlation measures, whether or not they might regulate the reb operons.

Funding: FAPESP - Iniciação Científica