Poster H-16

RASTA-Bacteria: a tool for the annotation of Toxin/Antitoxin modules in bacterial genomes

Authors:
Emeric SEVIN (*UMR6061-CNRS*)
Frédérique Barloy-Hubler (*UMR6061-CNRS*)

Short Abstract: Toxin-Antitoxin (TA) systems are major prokaryotic cell regulators during nutritional stress conditions, but they are poorly annotated. We developed RASTA-bacteria (Rapid Annotation for Systems of Toxin/Antitoxin in Bacteria), a tool that automatically identifies such systems in bacterial genomes, yielding a list of candidates ranked by expectation score.

Long Abstract:

Toxin-Antitoxin (TA) systems generally consist of two co-localized antagonist genes arranged in operon-like structure. The first gene encodes a stable toxin harmful to an essential cell process (the MazEF family for example inhibits translation by mRNA cleavage), while the second one encodes a labile antitoxin, which inhibits the aggressive behavior of the toxin by DNA- and/or protein-binding. Initially found on plasmids, TA systems were thought to solely act as plasmid maintenance mechanisms, and were called “plasmid addiction systems”. Later found on chromosomes as well thanks to homology searches, they turned out to be lethal only beyond a certain threshold of toxin expression. Moreover their effect showed reversibility if the antitoxin could again be expressed. It was then hypothesized that TA systems allowed amino-starved cells to enter a dormant state, before they could be “re-activated” in adequate conditions. TA systems are now seen as major actors in cell regulation during nutritional stress conditions, such as the stringent response, thereby assessing their importance in terms of prokaryotic biological processes. They are actually divided into two types, depending on the nature of the antitoxin (type I antitoxins are antisense-RNA while those of type II are translated proteins), and type II TAs are classified in 8 families, according to their modes of action or structural features. Nevertheless, they have mostly been neglected as genes of interest, whether during sequencing projects or later as subjects of study, and their annotation is far from optimal. This is mainly due to their relatively small sizes, as TAs uncovered to date range between 50 and 150 amino-acids in length.

Concerned by the lack of consideration in small Open Reading Frames (sORF), and basically in order to complete the annotation of Sinorhizobium meliloti, the bacterial model studied in our laboratory, we decided to correct this “TA issue”. We thus elaborated a method capable of easily discriminating eventual TA systems in a sequence. This method was automated, and led to the creation of RASTA-bacteria (Rapid Annotation for Systems of Toxin/Antitoxin in Bacteria), a bioinformatics tool allowing the automatic identification of such TA systems in bacterial genomes.

Presently limited to type II TAs, the prediction tool we designed is based on a thorough study of the genomic, physical, and structural features of toxins and antitoxins which are readily identified and characterized in prokaryotic organisms, and the tool is subdivided in the various corresponding modules. For each of these modules, the candidate sequences are

computed to verify if their characteristics fit the criteria defined as TA-relevant, and are then assigned a score, which summed up in the end to those of the other modules yields a ranking probability of being a TA gene. The tool also displays for each candidate a suggestion of family belonging, according to the above described admitted classification.
More precisely, the first step in RASTA-bacteria, which takes a raw DNA sequence as an input, is to locate all possible Open Reading Frames (ORFs). These are defined as lying between one of the four recognized bacterial start codons (ATG, CTG, GTG, TTG) and a stop codon, but with no former assumption except redundancy avoidance. This initial step is mandatory, and its evident goal is to recuperate genes missed during initial annotations. The extracted ORFs are then processed to identify TA features through the different modules. Namely, these include:
- A size-filter, so as to restrain the space of candidates to the coherent lengths.
- A “co-localisation” verifier that checks the presence of a close partner (the distance between a toxin and its antidote is seldom over twenty nucleotides)
- A Conserved Domain (CD) identifier, based on known TA domains extracted from academic databases.
- A tree builder and interpreter, for some candidates might not have a known CD but reveal TA homology if branched with a higher-confidence candidate.
- A codon utilization module is used as an additional indicator, as some TA genes demonstrate a particular use of codons.
RASTA-bacteria was developed on the alpha-proteobacterium Sinorhizobium meliloti, a gram-negative soil bacterium that can carry out symbiotic interaction with alfalfa. A former study characterized 110 possible TA genes among the 3 replicons of S. meliloti, all of them confirmed by RASTA. Thanks to its modular design, the tool is now under the process of generalization for use on all other rhizobiaceae, and will later be extended to all bacterial genomes.