TITLE: Prediction of non-coding RNAs of *Bifidobacterium breve* from RNA-seq data

AUTHORS: O, R. S.[1]; S, C. E. [1]; VERAS, A. A. O.[2]; OLIVEIRA, M. S. [2]; ALVES, J. T. C.[3]; SÁ, P. H. C. G.[1]

INSTITUTION: [1]FEDERAL RURAL UNIVERSITY OF AMAZONIA (RODOVIA PA 140, CEP 68680-000, TOMÉ-AÇU, PARÁ, BRAZIL); [2]FEDERAL UNIVERSITY OF PARÁ (RUA ITAIPU, 36, CEP 68464-000, TUCURUÍ, PARÁ, BRAZIL); [3]PARÁ STATE UNIVERSITY (AV. HILÉIA, S/N, CEP 68502-100, MARABÁ, PARÁ, BRAZIL)

ABSTRACT:

Non-coding RNAs (ncRNAs) have several biological functions. Among the ncRNAs are transfer RNAs (tRNAs), ribosomal RNAs (rRNAs), and small RNAs (sRNAs), as well as snRNAs that are a class of non-coding RNAs found inside the cell nucleus, snoRNA which are located inside the nucleolus, miRNA which are small fragments of RNA, siRNA, piRNA and scaRNA. To predict ncRNAs computational tools generally use the conservation of the secondary structure of a given class as evidence. Currently the diversity of non-coding RNAs are the targets of numerous researches, many of the ncRNAs have not yet been related to their functions, unlike genes encoding proteins. The *Bifidobacterium breve* is an anaerobic and probiotic bacterium that favors the function and maintenance of intestinal flora. And it is related to the immune system, improvement of allergic symptoms and with cervical cancer.Thus the objective of this work is to perform the prediction of *Bifidobacterium breve* ncRNAs using RNA-seq data from whole transcriptome sequencing and bioinformatics approaches. The reference genome used was *B. breve* DSM 20213 (acession number: NZ_AP012324.1) and the whole transcriptome sequencing data performed by the MiSeq platform (acession number: SRR5799034). The reads were aligned against the reference using the Bowtie2 program. The alignment result and the reads were analyzed by the Rockhopper and sRNA-Detect programs, to predict the small transcripts. These predicted transcripts were used in the manual identification process using the Rfam (rfam.xfam.org/) and RNAcentral (rnacentral.org) databases. As a result they were predicted by Rockhopper programs and sRNA-Detect 987 and 1498 small transcripts, respectively. From an *in-house* script only transcripts smaller than 50 bases were selected, resulting in 149 for Rockhopper program and 178 for sRNA-Detect program. Both results were analyzed in Rfam and RNAcentral resulting in the identification of 47 ncRNAs of different classes. Finally, in addition to predicting possible *B. breve* ncRNAs, this work demonstrated that it is possible to identify ncRNAs from whole transcriptome sequencing data and bioinformatics tools.

Keywords: *Bifidobacterium breve;* NGS; Prediction; RNA-seq; small RNA.