conferenceseries.com

International Conference on

Computational Biology and Bioinformatics

September 05-06, 2018   Tokyo, Japan

Feature extraction of short gene using recurrence quantification analysis

Saritha Namboodiri
University of Calicut, India

Computational prediction of short genes is an important field in bio-sequence analysis. In this work, we have developed a tool in python that discriminates short coding sequences from non-coding sequences in bacterial and archaea strains using recurrence quantification analysis. Our dataset comprises of 3,723 coding and 4,000 non-coding sequences belonging to two closely related *E. coli* strains K 12 MG 1655 and *E. coli* UT189 (UPEC)) and distantly related archaea strains *Halobacterium sp.* DL1 and *Natrinema pellirubrum* 157 obtained from Integrated Microbial Genome (IMG) database. The sequences were encoded into time series using 16 di-nucleotide properties obtained from the database of conformational and thermodynamic di-nucleotide properties. Recurrence Quantification Analysis (RQA) was applied to each of the di-nucleotide encoded sequence and quantified to RQA variables. RQA variables corresponding to di-nucleotide properties twist rise and GC Count together with CGT/CGG count and codon adaptation index of bacterial and archaea strains under study when subjected to ensemble classifier could distinguish coding from non-coding regions at about, 89% sensitivity, 81% specificity, 90% precision and 70% MCC. However, on considering bacterial strain alone, the ensemble classifier could distinguish coding from non-coding regions with 93% sensitivity, 88% specificity, 90% precision and 80% MCC. We observed that coding regions in the strains under study have higher Laminarity (RQA feature) for Twist_Rise and CGT/CGG Count than non-coding regions indicating repeated alternating cytosine and guanine residues (d(CG)n which have been demonstrated to form the left-handed Z DNA. This led us to conclude that short genes in Bacterial and Archaea strains may be enriched in Z-DNA.

Biography

Saritha Namboodiri has completed her PhD from University of Kerala, Department of Computational Biology and Bioinformatics. She is working as an Associate Professor in Department of Computer Science in an affiliated college of University of Calicut. She has held administrative position including Elected Academic Council Member and Syndicate Member of University of Calicut. She has served as Chair-Person of Board of Studies of University of Calicut and has many times been and is a Member of Board of Studies of University of Kerala and Kannur University. She delivers popular science lectures in and around Kerala.

saritha16.namboodiri@gmail.com

Notes: