

Quantitative frame analysis and gene discovery in prokaryote genomes

Luciano Brocchieri

University of Florida, USA

Abstract

With the deluge of prokaryote genome sequences, accurate computational prediction of coding regions is of fundamental importance for comparative genome analysis. Popular probabilistic gene prediction methods base their predictions on complex compositional rules derived from large, genome-specific learning sets of genes. An effective alternative way to identify coding regions in sequences of high GC content is the method of “frame-analysis”, based on graphical visualization of the contrasts in GC content among the three codon positions of GC-rich genes. To make frame analysis amenable to statistical characterization and to extend its applicability to sequences of any composition, we have developed new algorithms that quantify the principle of frame analysis allowing discovery of sequence segments of any length that exhibit statistically significant codon-position-specific associations of any nucleotide type. Applying our methods to a large set of prokaryote genomes, we identified a plethora of significant segments that were not explained by published annotations, among which many predicted genes similar in sequence and in length to gene sequences deposited in public databases. We identified a large number of ORFs, mostly of very short length (“mini-genes”), that were uniquely predicted by our methods. Some of these could be functionally characterized by conservation (e.g., transferases, transcription factors, leader peptides), but most were ORFs of unknown functionality. RNA-seq analysis of *Pseudomonas aeruginosa* revealed expression of many of the mini-genes identified in its genome by our methods. Our results suggest the potential of alternative gene-finding strategies to making new discoveries on the coding potential of prokaryote genomes.

Biography

Dr. Luciano Brocchieri has completed his Ph.D in theoretical population genetics from the University of Parma, Italy, and postdoctoral studies at Stanford University, Department of Mathematics. After working as Senior Scientist at Stanford University, Department of Mathematics, he is now Assistant Professor at the University of Florida, Department of Medicine. Dr. Brocchieri research is focused on the analysis of the evolution of protein families and on the development of bioinformatics methods for the analysis of nucleic acid and protein sequences. Dr. Brocchieri is the author of several papers, reviews and commentaries published in international reputed scientific journals.