Andrew Hart, Center for Mathematical Modeling, University of Chile

An Entropy-Based Technique for Classifying Bacterial Chromosomes According to Synonymous Codon Usage

Abstract

Living organisms make use of molecular machinery called proteins that consist of chains of amino acids appropriately folded and fitted together. The instructions for building proteins are encoded in genes as sequences of nucleotide triplets called codons. In total, 61 of the 64 possible codons are used to represent 20 amino acids and the mapping from codons to amino acids is known as the genetic code. Codons that code for the same amino acid are called synonymous codons. Synonymous codon usage (SCU) is known to vary not only between organisms, but also between genes within the same genome and it is believed that SCU is implicated in a number of biological mechanisms.

We present a method for classifying bacterial chromosomes based on the degree to which they use synonymous codons uniformly or preferentially, that is, whether or not codons that code for the same amino acid appear with the same relative frequency. The method uses a measure of SCU we call the codon information bias (CIB) which is based on conditional entropy and the Dirichlet distribution. The application of the approach to a large collection of bacterial chromosomes will be discussed together with further possible applications.