Annual scientific meeting November 19 2012

Some properties of trinucleotides and tetranucleotides

Pirillo Giuseppe

Consiglio Nazionale delle Ricerche IASI CNR, Unità di Firenze Dipartimento di Matematica "U. Dini" Università di Firenze, Italy; Université de Marne-la-Vallée, France

Abstract

In 1994, with a statistical study of trinucleotide occurences per frame, D. Arquès and C. Michel identified the fol lowing set of 20 trin- ucleotides in the gene population of both eucaryotes E U K and procaryotes P RO: {AAC , AAT , AC C , AT C , AT T , C AG, C T C , C T G, GAA, GAC , GAG, GAT , GC C , GGC , GGT , GT A, GT C , GT T , T AC , T T C }. This set of words of length three on the genetic alphabet A_4 = {A, C, G, T } has remarkable properties: it is a self-complementary set, it is a circular code, it is maximal and it has the C^3-property. This and other identifications of trin- ucleotide circular codes in different genomes in the last twenty years raised interest in the concept of trinucleotide circular code for genetics. In 2003 we found an efficient algorithm for testing the circularity of a trinucleotide code and in 2005 we found the list of al l 528 maximal self-complementary circular codes. In 2008 we presented a hierarchy of the self-complementary circular codes and in 2012 we presented a hierarchy of al l circular codes. These circu- lar codes could permit the identification, either in paral lel with or substituting existing methods used by biologist, of as yet unknown coding regions of DNA. A hierarchy of tetranucleotide circular codes is one of our aims in the future. In this paper we begin the study of the unbordered tetranucleotides and of the "forbidden configurations" for tetranucleotides and we give a first result.

Keywords: Genetic alphabet, genetic code, circular code, trinucleotides, tetranucleotides.