

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 occurrences per frame, D. Arquès and C. Michel identified the following set of 20 trinucleotides in the gene population of both eucaryotes E U K and procaryotes P R O: {AAC, AAT, ACC, ATC, ATT, CAG, CTC, CTG, GAA, GAC, GAG, GAT, GCC, GGC, GGT, GTA, GTC, GTT, TAC, TTC}. 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 trinucleotide 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 all 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 all circular codes. These circular codes could permit the identification, either in parallel 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.