ISTITUTO DI ANALISI DEI SISTEMI ED INFORMATICA
"Antonio Ruberti"
CONSIGLIO NAZIONALE DELLE RICERCHE

L. Bussoli, C. J. Michel, G. Pirillo

ON CONJUGATION PARTITIONS OF SETS OF TRINUCLEOTIDES ON CONJUGATION PARTITIONS OF SETS OF TRINUCLEOTIDES

R. 29, 2011

L. Bussoli – Dipartimento di Matematica "U.Dini", Viale Morgagni 67/A, 50134 Firenze, Italia. Email: bussoli@math.unifi.it.

C. J. Michel – Equipe de Bioinformatique Théorique, BFO, LSIIT (UMR 7005), Université de Strasbourg, Pôle API, Boulevard Sébastien Brant, 67400 Illkirch, France. Email: michel@dpt-info.u-strasbg.fr.

G. Pirillo – Consiglio Nazionale delle Ricerche, Istituto di Analisi dei Sistemi ed Informatica "Antonio Ruberti", Unità di Firenze, Dipartimento di Matematica "U.Dini", Viale Morgagni 67/A, 50134 Firenze, Italia and Université de Marne-la-Vallée, 5 boulevard Descartes, 77454 Marne-la-Vallée Cedex 2, France. Email: pirillo@math.unifi.it.

ISSN: 1128–3378

Collana dei Rapporti dell'Istituto di Analisi dei Sistemi ed Informatica "Antonio Ruberti", CNR

viale Manzoni 30, 00185 ROMA, Italy

tel. ++39-06-77161
fax ++39-06-7716461
email: iasi@iasi.cnr.it
URL: http://www.iasi.cnr.it

Abstract

We prove that a trinucleotide circular code is self-complementary if and only if its two conjugated classes are complement of each other. Using only this proposition, we prove that if a circular code is self-complementary then either both its two conjugated classes are circular codes or none is a circular code.

Key words: trinucleotides, self-complementary circular codes.

1. Introduction

We continue our study of the combinatorial properties of trinucleotide circular codes. A trinucleotide is a word of three letters (triletter) on the genetic alphabet $\{A, C, G, T\}$. The set of 64 trinucleotides is a code in the sense of language theory, more precisely a uniform code, but not a circular code [4, 14]. In order to have an intuitive meaning of these notions, codes are written on a straight line while circular codes are written on a circle, but, in both cases, unique decipherability is required.

Comma free codes, a very particular case of circular codes, has been studied for a long time, e.g. [6, 9, 10]. After the discovery of a circular code in genes with important properties [1], circular codes are mathematical objects studied in combinatorics, theoretical computer science and theoretical biology, e.g. [13, 3, 2, 25, 11, 7, 19, 16, 8, 15, 20, 17, 21, 12, 22, 23, 5].

In particular, there are 528 self-complementary circular codes [1, 24, 18], and as proved here, they are naturally partitioned in two quite symmetric classes.

In this paper, we study some particular partitions of $\mathcal{A}_4^3 \setminus \{AAA, CCC, GGG, TTT\}$. Indeed, each circular code X_0 can be associated with two other subsets X_1 and X_2 of $\mathcal{A}_4^3 \setminus \{AAA, CCC, GGG, TTT\}$ simply by operating two circular permutations of one letter and two letters on the trinucleotides of X. Then, we prove our main result, i.e. *a circular code is self-complementary if and only if the remaining two classes are complement of each other*. Furthermore, we also show that *a subset of $A_4^3 \setminus \{AAA, CCC, GGG, TTT\}$ is a circular code if and only if the set consisting of all its complement is a circular code.*

As a consequence of these results, we also prove that *if a circular code is self-complementary then either both its two conjugated classes are circular codes or none is a circular code.*

2. Definitions

The classical notions of *alphabet, empty word, length, factor, proper factor, prefix, proper prefix, suffix, proper suffix, lexicographical order,* etc. are those of [4]. Let $\mathcal{A}_4 = \{A, C, G, T\}$ denote the genetic alphabet, lexicographically ordered with $A < C < G < T$. We use the following notation:

- $\mathcal{A}_4^*$ (respectively $\mathcal{A}_4^+$) is the set of words (respectively non-empty words) over $\mathcal{A}_4$,
- $\mathcal{A}_4^2$ is the set of the 16 words of length two (*diletters* or *dinucleotides*) and
- $\mathcal{A}_4^3$ is the set of the 64 words of length three (*triletters* or *trinucleotides*).

We now recall two important genetic maps, the definitions of code and circular code, and the property of C^3-self-complementary for a circular code, in particular [4, 1, 20, 24, 18].

Definition 2.1. *The complementarity map $\mathcal{C}$: $\mathcal{A}_4^+ \to \mathcal{A}_4^+$ is defined by $\mathcal{C}(A) = T$, $\mathcal{C}(T) = A$, $\mathcal{C}(C) = G$ and $\mathcal{C}(G) = C$, and by $\mathcal{C}(uv) = \mathcal{C}(v)\mathcal{C}(u)$ for all $u, v \in \mathcal{A}_4^+$, e.g., $\mathcal{C}(AAC) = GTT$.*

The map $\mathcal{C}$ on words is naturally extended to a word set X: its complementary trinucleotide set $\mathcal{C}(X)$ is obtained by applying the complementarity map $\mathcal{C}$ to all the trinucleotides of X.

Definition 2.2. *The circular permutation map $\mathcal{P}$: $\mathcal{A}_4^3 \to \mathcal{A}_4^3$ permutes circularly each trinucleotide $l_1l_2l_3$ as follows $\mathcal{P}(l_1l_2l_3) = l_2l_3l_1$.*

The map $\mathcal{P}$ on words is also naturally extended to a word set X: its permuted trinucleotide set $\mathcal{P}(X)$ is obtained by applying the circular permutation map $\mathcal{P}$ to all the trinucleotides of X. We shortly write $\mathcal{P}^2(X)$ for $\mathcal{P}(\mathcal{P}(X))$.

Definition 2.3. *A set X_0 of words is a code if, for each $x_1, \ldots, x_n, x'_1, \ldots, x'_m \in X$, $n, m \geq 1$, the condition $x_1 \cdots x_n = x'_1 \cdots x'_m$ implies $n = m$ and $x_i = x'_i$ for $i = 1, \ldots, n$.*

Definition 2.4. *A trinucleotide code X_0 is circular if, for each $x_1, \ldots, x_n, x'_1, \ldots, x'_m \in X$, $n, m \geq 1$, $p \in \mathcal{A}_4^*$, $s \in \mathcal{A}_4^+$, the conditions $sx_2 \cdots x_n p = x'_1 \cdots x'_m$ and $x_1 = ps$ imply $n = m$, $p = \varepsilon$ (empty word) and $x_i = x'_i$ for $i = 1, \ldots, n$.*

Definition 2.5. *A trinucleotide code X_0 is self-complementary if, for each $x \in X_0$, $\mathcal{C}(x) \in X_0$.*

Definition 2.6. *If X_0 is a subset of $\mathcal{A}_4^3 \setminus \{AAA, CCC, GGG, TTT\}$, we denote by X_1 the permuted trinucleotide set $\mathcal{P}(X_0)$ and by X_2 the permuted trinucleotide set $\mathcal{P}^2(X_0)$ and we call X_1 and X_2 the conjugated classes of X_0.*

Definition 2.7. *A trinucleotide circular code X_0 is C^3-self-complementary if X_0, X_1 and X_2 are circular codes satisfying the following properties: $X_0 = \mathcal{C}(X_0)$ (self-complementary), $\mathcal{C}(X_1) = X_2$ (and $\mathcal{C}(X_2) = X_1$).*

We have proved that there are exactly 528 trinucleotide self-complementary circular codes having 20 elements [1, 24, 18].

The concept of *necklace* was introduced by Pirillo for circular codes [20] in order to characterize the circular codes for an efficient algorithm development. Let $l_1, l_2, \ldots, l_{n-1}, l_n, \ldots$ be letters in $\mathcal{A}_4$, $d_1, d_2, \ldots, d_{n-1}, d_n, \ldots$ diletters in $\mathcal{A}_4^2$ and $n \geq 2$ an integer.

Definition 2.8. *Letter Diletter Continued Necklaces (LDCN): We say that the ordered sequence $l_1, d_1, l_2, d_2, \ldots, d_{n-1}, l_n, d_n, l_{n+1}$ is an $(n+1)LDCN$ for a subset $X \subset \mathcal{A}_4^3$ if $l_1d_1, l_2d_2, \ldots, l_nd_n \in X$ and $d_1l_2, d_2l_3, \ldots, d_{n-1}l_n, d_nl_{n+1} \in X$.*

Any trinucleotide set is a code (more precisely, a *uniform code* [4]) but only few of them are circular codes. We have the following proposition.

Proposition 2.9. *[20]. Let X be a trinucleotide code. The following conditions are equivalent:*
(i) X is a circular code;
(ii) X has no $5LDCN$.

The figure below explains the notion of $5LDCN$.

$\mathbf{l_1}$	$\mathbf{d_1}$	$\mathbf{l_2}$	$\mathbf{d_2}$	$\mathbf{l_3}$	$\mathbf{d_3}$	$\mathbf{l_4}$	$\mathbf{d_4}$	$\mathbf{l_5}$

3. Results

Proposition 3.1. *If X_0 is a trinucleotide circular code having 20 elements and X_1 and X_2 are its two conjugated classes then X_0, X_1 and X_2 constitute a partition of $\mathcal{A}_4^3 \setminus \{AAA, CCC, GGG, TTT\}$.*

Proof. It is enough to prove that $X_0 \cap X_1 = X_0 \cap X_2 = X_1 \cap X_2 = \emptyset$.
Suppose that the trinucleotide $l_1l_2l_3$ belongs both to the classes X_0 and X_1. Then $l_1l_2l_3$ and $l_3l_1l_2$ are both in class X_0. As no two conjugated trinucleotides can belong to a circular code, we are in contradiction.
Suppose that the trinucleotide $l_1l_2l_3$ belongs both to the classes X_0 and X_2. Then $l_1l_2l_3$ and $l_2l_3l_1$ are both in class X_0. As no two conjugated trinucleotides can belong to a circular code,

we are in contradiction.
Suppose that the trinucleotide $l_1l_2l_3$ belongs both to the classes X_1 and X_2. Then $l_3l_1l_2$ and $l_2l_3l_1$ are both in class X_0. As no two conjugated trinucleotides can belong to a circular code, we are in contradiction.
So, $X_0 \cap X_1 = X_0 \cap X_2 = X_1 \cap X_2 = \emptyset$. ∎

Proposition 3.2. *The class of self-complementary circular codes X_0 with both X_1 and X_2 in the class of circular codes is non-empty.*

Proof. Consider, for example, the following set of 20 trinucleotides $\{AAC, AAG, AAT, ACC, ACG, ACT, AGC, AGG, AGT, ATC, ATT, CCT, CGT, CTT, GAT, GCC, GCT, GGC, GGT, GTT\}$ and call it X_0. It is enough to prove that X_0 is a self-complementary circular code and that its two conjugated classes X_1 and X_2 are also circular codes. ∎

Proof. X_0 **is a self-complementary circular code.**
(i) X_0 is self-complementary. Obvious by inspection.
(ii) X_0 is a circular code. We use Proposition 1 [20]. By way of contradiction, suppose that X_0 admits a $5LDCN$. As l_2 can be A, C, G or T, it is enough to prove that each choice leads to a contradiction.
If $l_2 = A$ then there is no possible d_1 as A is not a suffix of any trinucleotide of X_0, contradiction;
If $l_2 = C$, there are three possible d_2:
- if $d_2 = CT$ (a) or $d_2 = GT$ (b) then $l_3 = T$ (c) but there is no possible d_3 as T is not a prefix of any trinucleotide of X_0, contradiction,
- if $d_2 = TT$ (d), there is a contradiction as no trinucleotide of X_0 has a prefix TT.

If $l_2 = G$, there are six possible d_2:
- if $d_2 = CT$ or $d_2 = GT$, contradiction (a) and (b),
- if $d_2 = CC$ then $l_3 = T$, contradiction (c),
- if $d_2 = GC$ or $d_2 = AT$ then $l_3 = C$ or $l_3 = T$:
  * if $l_3 = C$, there are three possible d_3: if $d_3 = CT$ or $d_3 = GT$ then $l_4 = T$, similarly to (c), contradiction, and if $d_3 = TT$, similarly to (d), contradiction,
  * if $l_3 = T$, contradiction (c),
- if $d_2 = TT$, contradiction (d);

If $l_2 = T$, similarly to (c), contradiction.
As, for each letter, we cannot complete the assumed $5LDCN$ necklace for X_0, we are in contradiction. Hence, X_0 is a circular code.

X_1 **is a circular code.** We have to prove that $X_1 = \mathcal{P}(X_0) = \{ACA, AGA, ATA, ATG, CCA, CCG, CGA, CTA, CTC, CTG, GCA, GCG, GGA, GTA, GTC, GTG, TCA, TTA, TTC, TTG\}$ is a circular code. By way of contradiction, assume that X_1 admits a $5LDCN$ necklace.
If $l_2 = A$, there are four possible d_2: CA, GA, TA and TG, but no possible l_3, contradiction.
If $l_2 = C$, there are three possible d_1: CT, GT and TT, but no possible l_1, contradiction.
If $l_2 = G$, there are five possible d_1: AT, CC and GC, and the cases CT, GT and TT already seen, but no possible l_1, contradiction.
If $l_2 = T$, there is no possible d_1, contradiction.
Hence, X_1 is also a circular code.

X_2 **is a circular code.** Finally, we have to prove that $X_2 = \mathcal{P}^2(X_0) = \{CAA, CAC, CAG, CAT, CGC, CGG, GAA, GAC, GAG, TAA, TAC, TAG, TAT, TCC, TCG, TCT, TGA, TGC, TGG, TGT\}$ is a circular code. By way of contradiction, assume that X_2 admits a $5LDCN$

necklace.
If $l_2 = A$, there is no possible d_2, contradiction.
If $l_2 = C$, there are six possible d_2: AA, AC, AG, AT, GC and GG, but no possible l_3, contradiction.
If $l_2 = G$, there are three possible d_2: AA, AC and AG which are cases already seen, contradiction.
If $l_2 = T$, there are four possible d_1: CA, TA, TC, and TG, but no possible l_1, contradiction.
Hence, as X_0 and X_1, X_2 is also a circular code. ∎

Proposition 3.3. *The class of self-complementary circular codes X_0 having 20 elements with neither X_1 nor X_2 in the class of circular codes is non-empty.*

Proof. Consider, for example, the following set of 20 trinucleotides $X_0 = \{AAC, AAG, AAT, ACC, ACG, ACT, AGC, AGT, ATC, ATT, CGT, CTT, GAT, GCC, GCT, GGA, GGC, GGT, GTT, TCC\}$ and call it X_0. It is enough to prove that X_0 is a self-complementary circular code and that neither its conjugated class X_1 nor its conjugated class X_2 are circular codes.

X_0 **is a self-complementary circular code.**
(i) X_0 is self-complementary. Obvious by inspection.
(ii) X_0 is a circular code. We use Proposition 1 [20]. By way of contradiction, assume that X_0 admits a $5LDCN$ necklace.
If $l_2 = A$ then there is one possible $d_1 = GG$ but no possible l_1, contradiction.
If $l_2 = C$, there are two possible d_2:
- if $d_2 = GT$ then $l_3 = T$ (a) and $d_3 = CC$ (b) but there is no possible l_4, contradiction;
- if $d_2 = TT$ (c) then there is no possible l_3, contradiction.

If $l_2 = G$ we have seven possible d_2:
- if $d_2 = AT$ then $l_3 = C$ or $l_3 = T$:
  - * if $l_3 = C$ (d) then $d_3 = GT$ or $d_3 = TT$:
    - \- if $d_3 = GT$ then $l_4 = T$ and $d_4 = CC$ but there is no possible l_5, contradiction,
    - \- if $d_3 = TT$ then there is no possible l_4, contradiction,
  - * if $l_3 = T$, contradiction (a);
- if $d_2 = CC$, similarly to (b), contradiction;
- if $d_2 = CT$, $d_2 = GA$ or $d_2 = GT$ then $l_3 = T$, contradiction (a);
- if $d_2 = GC$ then $l_3 = C$ or $l_3 = T$, contradiction (a) and (d),
- if $d_2 = TT$, contradiction (c);

If $l_2 = T$, similarly to (a), contradiction.
Hence, X_0 is a circular code.

X_1 **is not a circular code.** We have $X_1 = \mathcal{P}(X_0) = \{ACA, AGA, ATA, ATG, CCA, CCG, CCT, CGA, CTA, CTG, GAG, GCA, GCG, GTA, GTC, GTG, TCA, TTA, TTC, TTG\}$. We use a technique developed in [5]. Observe that X_1 contains $\{AGA, CCT, GAG, TTC\}$. So, $(l_1, d_1, l_2, d_2, l_3, d_3, l_4, d_4, l_5) = (A, GA, G, AG, A, GA, G, AG, A)$ is a $5LDCN$ for this 4-element subset of X_1 and, a fortiori, for X_1 itself which, consequently, is not a circular code.

X_2 **is not a circular code.** We have $X_2 = \mathcal{P}^2(X_0) = \{AGG, CAA, CAC, CAG, CAT, CGC, CGG, CTC, GAA, GAC, TAA, TAC, TAG, TAT, TCG, TCT, TGA, TGC, TGG, TGT\}$. We again use a technique developed in [5]. Remark that X_2 contains $\{GAA, CTC, AGG, TCT\}$. So, $(l_1, d_1, l_2, d_2, l_3, d_3, l_4, d_4, l_5) = (T, CT, C, TC, T, CT, C, TC, T)$ is a $5LDCN$ for this 4-element subset of X_2 and, a fortiori, for X_2 itself which, consequently, is not a circular code. ∎

We need the propositions hereafter and, in particular the following one which states a general property of the involutional antiisomorphisms ($\mathcal{C}$ is one of them).

Proposition 3.4. *A subset X of $\mathcal{A}_4^3 \setminus \{AAA, CCC, GGG, TTT\}$ is a circular code if and only if $\mathcal{C}(X)$ is a circular code.*

Proof. Suppose, first, that X is not a circular code and that $\mathcal{C}(X)$ is a circular code. So X has a $5LDCN$. This means that there are 13 nucleotides, say

$$b_1, b_2, b_3, b_4, b_5, b_6, b_7, b_8, b_9, b_{10}, b_{11}, b_{12}, b_{13}$$

such that the trinucleotides

$$b_1b_2b_3, b_4b_5b_6, b_7b_8b_9, b_{10}b_{11}b_{12} \in X$$

and

$$b_2b_3b_4, b_5b_6b_7, b_8b_9b_{10}, b_{11}b_{12}b_{13} \in X.$$

Now, consider the sequence

$$\mathcal{C}(b_{13}), \mathcal{C}(b_{12}), \mathcal{C}(b_{11}), \mathcal{C}(b_{10}), \mathcal{C}(b_9), \mathcal{C}(b_8), \mathcal{C}(b_7), \mathcal{C}(b_6), \mathcal{C}(b_5), \mathcal{C}(b_4), \mathcal{C}(b_3), \mathcal{C}(b_2), \mathcal{C}(b_1).$$

All the following trinucleotides belongs to $\mathcal{C}(X)$

$$\mathcal{C}(b_{13})\mathcal{C}(b_{12})\mathcal{C}(b_{11}), \mathcal{C}(b_{10})\mathcal{C}(b_9)\mathcal{C}(b_8), \mathcal{C}(b_7)\mathcal{C}(b_6)\mathcal{C}(b_5), \mathcal{C}(b_4)\mathcal{C}(b_3)\mathcal{C}(b_2) \in \mathcal{C}(X)$$

and

$$\mathcal{C}(b_{12})\mathcal{C}(b_{11})\mathcal{C}(b_{10}), \mathcal{C}(b_9)\mathcal{C}(b_8)\mathcal{C}(b_7), \mathcal{C}(b_6)\mathcal{C}(b_5)\mathcal{C}(b_4), \mathcal{C}(b_3)\mathcal{C}(b_2)\mathcal{C}(b_1) \in \mathcal{C}(X)$$

as they are the complement of trinucleotides in X. So, $\mathcal{C}(X)$ admits a $5LDCN$ and it cannot be a circular code. Contradiction.

The case X is a circular code and $\mathcal{C}(X)$ is not a circular code is similar. ∎

Proposition 3.5. *Let S be a self-complementary subset of $\mathcal{A}_4^3 \setminus \{AAA, CCC, GGG, TTT\}$. If S is partitioned into three classes such that two of them are the complement of each other then necessarily the third one is self-complementary.*

Proof. Let X, Y and Z be the three classes of an arbitrary partition of S and suppose that Y and Z are one complement of the other, i.e. Y and Z satisfy $\mathcal{C}(Y) = Z$. Let t be a trinucleotide of X. We claim that $\mathcal{C}(t) \notin Y$. Indeed, in the opposite case, Z should not be the complement of Y because $t \in X$. We also claim that $\mathcal{C}(t) \notin Z$. Indeed, in the opposite case, Y should not be the complement of Z because $t \in X$. It remains the case $\mathcal{C}(t) \in X$. So, X is self-complementary. ∎

Remark 1. *Clearly, if X, Y and Z constitutes an arbitrary partition of $\mathcal{A}_4^3 \setminus \{AAA, CCC, GGG, TTT\}$ then the self-complementarity of X is not enough to ensure that Y and Z are complementary of each other. This remark is again true if, in addition, X is a self-complementary circular code having 20 elements. Indeed in this case, it is easy to make a partition $\mathcal{A}_4^3 \setminus \{X \cup \{AAA, CCC, GGG, TTT\}\}$ in two classes Y and Z that are not self-complementary of each other. Any case, if we consider the partition of $\mathcal{A}_4^3 \setminus \{AAA, CCC, GGG, TTT\}$ in the three classes given by a self-complementary trinucleotide circular code X_0 having 20 elements and by its two conjugated classes X_1 and X_2 then the necessary and sufficient condition holds (Proposition 3.6 below).*

Proposition 3.6. *A trinucleotide circular code X_0 having 20 elements is self-complementary if and only if X_1 and X_2 are complement of each other.*

Proof. **"If " part.** It is a trivial consequence of Proposition 3.5.
"Only if " part. Suppose that X_0 is self-complementary and consider the partition X_0, X_1 and X_2 of $\mathcal{A}_4^3 \setminus \{AAA, CCC, GGG, TTT\}$. Suppose that the trinucleotide, say PQR, belongs to X_0. Then, also $\mathcal{C}(R)\mathcal{C}(Q)\mathcal{C}(P) \in X_0$. We have $QRP, \mathcal{C}(Q)\mathcal{C}(P)\mathcal{C}(R) \in X_1$ and $RPQ, \mathcal{C}(P)\mathcal{C}(R)\mathcal{C}(Q) \in X_2$. As PQR is a generic trinucleotide of X_0 and as

$$QRP \text{ is the complement of } \mathcal{C}(P)\mathcal{C}(R)\mathcal{C}(Q)$$

and

$$\mathcal{C}(Q)\mathcal{C}(P)\mathcal{C}(R) \text{ is the complement of } RPQ$$

then X_1 is the complement of X_2. ∎

As a consequence, we have the following proposition.

Proposition 3.7. *If a trinucleotide circular code X_0 having 20 elements is self-complementary then either*
(i) X_1 and X_2 are both circular codes
or
(ii) X_1 and X_2 admit both a necklace (and consequently, they are not circular codes).

Proof. We have four possibilities:
X_1 is a circular code and X_2 is a circular code;
X_1 is a circular code and X_2 is not a circular code;
X_1 is not a circular code and X_2 is a circular code;
X_1 is not a circular code and X_2 is not a circular code.
Now, by applying Propositions 3.2 and 3.3, we have that the first and the last possibilities can be effectively realized.
Suppose that, by way of contradiction, the second possibility is realized. So, X_1 is a circular code. By Proposition 3.6, we have $\mathcal{C}(X_1) = X_2$. So, by Proposition 3.4, X_2 must also be a circular code. Contradiction.
Suppose that, by way of contradiction, the third possibility is realized. So, X_2 is a circular code. By Proposition 3.6, we have $\mathcal{C}(X_2) = X_1$. So, by Proposition 3.4, X_1 must also be a circular code. Contradiction.
So, only the first and the last possibilities are true. ∎

Hence, our proposition holds.

Proposition 3.8. *The 528 self-complementary circular codes having 20 elements are partitioned in two classes: one class contains codes with the two permuted sets X_1 and X_2 which are both circular codes while the other class contains codes with the two permuted sets X_1 and X_2 which both are not circular codes.*

Proof. It is enough to apply Proposition 3.7 to each of the 528 trinucleotide circular codes having 20 elements. ∎

Acknowledgement. We thank Jacques Justin for his advices. The second author thanks the Dipartimento di matematica "U. Dini" for giving him a friendly hospitality.

References

[1] D.G. Arquès, C.J. Michel. A complementary circular code in the protein coding genes. *J. Theor. Biol.* **182**, 45-58 (1996).

[2] F. Bassino. Generating function of circular codes. *Adv. Appl. Math.* **22**, 1-24 (1999).

[3] M.-P. Béal, J. Senellart. On the bound of the synchronization delay of a local automaton. *Theoret. Comput. Sci.* **205**, 297-306 (1998).

[4] J. Berstel, D. Perrin. *Theory of Codes*. Academic Press, London, 1985.

[5] L. Bussoli, C.J. Michel, G. Pirillo. On some forbidden configurations for self-complementary trinucleotide circular codes. *J. Algebra Number Theory Academia*, in press (2011).

[6] F.H.C. Crick, J.S. Griffith, L.E. Orgel. Codes without commas. *Proc. Natl. Acad. Sci.* **43**, 416-421 (1957).

[7] G. Frey, C.J. Michel. Circular codes in archaeal genomes. *J. Theor. Biol.* **223**, 413-431 (2003).

[8] G. Frey, C.J. Michel. Identification of circular codes in bacterial genomes and their use in a factorization method for retrieving the reading frames of genes. *J. Comput. Biol. Chem.* **30**, 87-101 (2006).

[9] S.W. Golomb, B. Gordon, L.R. Welch. Comma-free codes. *Canad. J. Math.* **10**, 202-209 (1958).

[10] S.W. Golomb, L.R. Welch, M. Delbrück. Construction and properties of comma-free codes. *Biol. Medd. Dan. Vid. Selsk.* **23** (1958).

[11] R. Jolivet, F. Rothen. Peculiar symmetry of DNA sequences and evidence suggesting its evolutionary origin in a primeval genetic code. Proceedings of the First European Workshop in Exo-/astro-biology. Eds.: P. Ehrenfreund, O. Angerer & B. Battrick. ESA SP-496, Noordwijk, 173-176 (2001).

[12] M.V. José, T. Govezensky, J.A. García, J.R. Bobadilla. On the evolution of the standard genetic code: vestiges of critical scale invariance from the RNA world in current prokaryote genomes. *PLoS ONE*, **4(2)**, e4340 (2009).

[13] A.J. Koch, J. Lehman. About a symmetry of the genetic code. *J. Theor. Biol.* **189**, 171-174 (1997).

[14] J.-L. Lassez. Circular codes and synchronization. *Int. J. Comput. Syst. Sciences* **5**, 201-208 (1976).

[15] J.-L. Lassez, R.A. Rossi, A.E. Bernal. Crick's hypothesis revisited: the existence of a universal coding frame. *IEEE AINAW'07* (2007).

[16] E.E. May, M.A. Vouk, D.L. Bitzer, D.I. Rosnick. An error-correcting framework for genetic sequence analysis. *J. Franklin Inst.* **341**, 89-109 (2004).

[17] C.J. Michel, G. Pirillo, M.A. Pirillo. Varieties of comma-free codes. *Comput. Math. Appl.* **55**, 989-996 (2008).

[18] C.J. Michel, G. Pirillo, M.A. Pirillo. A relation between trinucleotide comma-free codes and trinucleotide circular codes. *Theoret. Comput. Sci.* **401**, 17-25 (2008).

[19] C. Nikolaou, Y. Almirantis. Mutually symmetric and complementary triplets: difference in their use distinguish systematically between coding and non-coding genomic sequences. *J. Theor. Biol.* **223**, 477-487 (2003).

[20] G. Pirillo. A characterization for a set of trinucleotides to be a circular code. In *Determinism, Holism, and Complexity* (Edited by C. Pellegrini, P. Cerrai, P. Freguglia, V. Benci and G. Israel), Kluwer (2003).

[21] G. Pirillo. A hierarchy for circular codes. *RAIRO-Theor. Inf. Appl.* **42**, 717-728 (2008).

[22] G. Pirillo. Some remarks on prefix and suffix codes. *Pure Math. Appl.* **19**, 53-60 (2008).

[23] G. Pirillo. Non sharing border codes. *Adv. Appl. Math. Sci.* **3**, 215-223 (2010).

[24] G. Pirillo, M.A. Pirillo. Growth function of self-complementary circular codes. *Biology Forum* **98**, 97-110 (2005).

[25] N. Štambuk. On circular coding properties of gene and protein sequences. *Croatica Chemica Acta* **72**, 999-1008 (1999).