Istituto di Analisi dei Sistemi ed Informatica
"Antonio Ruberti"
CONSIGLIO NAZIONALE DELLE RICERCHE

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

ON SOME FORBIDDEN CONFIGURATIONS FOR SELF-COMPLEMENTARY TRINUCLEOTIDE CIRCULAR CODES

R. 28, 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

A self-complementary trinucleotide circular code has two permuted sets which are either both circular codes or both not circular codes.

Key words: trinucleotide circular codes; self-complementary property; permuted circular sets.

1. Introduction

We continue our study of the combinatorial properties of maximum trinucleotide circular codes, i.e. circular codes with 20 trinucleotides. A trinucleotide is a word of three letters on the 4-letter alphabet $\{A, C, G, T\}$. The set of $4^3 = 64$ trinucleotides is a code in the sense of language theory, more precisely a uniform code, but not a circular code. 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.

In the past 50 years, circular codes have been studied in theoretical biology, mainly to understand the structure and the origin of the genetic code as well as the reading frame (construction) of genes, e.g. [6, 9, 10, 14]. In 1996, a circular code of 20 trinucleotides was identified statistically on two large and different gene populations, eukaryotes and prokaryotes [1]. Furthermore, this code has two properties: it is self-complementary and its two permuted sets are also circular codes. During the last years, circular codes are mathematical objects studied in discrete mathematics, theoretical computer science and theoretical biology, e.g. [2, 13, 4, 3, 25, 11, 7, 19, 16, 24, 8, 15, 20, 17, 18, 21, 12, 22, 23].

Among the $12,964,440$ trinucleotide circular codes, only 528 of them are self-complementary [1, 24, 18]. New propositions are identified here with these 528 self-complementary trinucleotide circular codes.

The 528 self-complementary trinucleotide circular codes are divided into two classes: a class of 216 circular codes where each code has two permuted sets X_1 and X_2 which are circular codes [1, 18] and a class containing the remaining 312 circular codes, denoted $\overline{C^3}$, for which the circularity of the permuted sets X_1 and of X_2 were not investigated so far.

For the $\overline{C^3}$ class, three cases are possible:

(i) X_1 is a circular code and X_2 is not a circular code;

(ii) X_1 is not a circular code and X_2 is a circular code;

(iii) X_1 and X_2 are not circular codes.

The main proposition of this paper will prove that only the case (iii) is verified. This result was obtained with a detailed identification of 51 "forbidden configurations" corresponding to 51 propositions which are collected in three groups (section Results). Thus, the 528 self-complementary trinucleotide circular codes are divided into two classes for which a certain symmetry holds. Indeed, even if these two classes have different cardinality (216 and 312), the first class contains 216 circular codes where X_1 and X_2 are both circular codes while the second class contains 312 circular codes where X_1 and X_2 are both non-circular codes.

2. Definitions

For the classical notions of *alphabet, empty word, length, factor, proper factor, prefix, proper prefix, suffix, proper suffix, alphabetical order*, we refer to [5]. 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 nonempty words) over $\mathcal{A}_4$;
- $\mathcal{A}_4^2$ is the set of the 16 words of length 2 (or *diletters or dinucleotides*) and
- $\mathcal{A}_4^3$ the set of the 64 words of length 3 (or *triletters or trinucleotides*).

We now recall two important genetic maps, the definitions of code and circular code [14, 5], and the C^3 self-complementary property of a circular code [1].

Definition 2.1. *The complementary map $\mathcal{C}$: $\mathcal{A}_4^+ \rightarrow \mathcal{A}_4^+$ 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. $\mathcal{C}(AAC) = GTT$. This map on words is naturally extended to word sets: a complementary trinucleotide set is obtained by applying the complementary map $\mathcal{C}$ to all its trinucleotides.*

Definition 2.2. *The circular permutation map $\mathcal{P}$: $\mathcal{A}_4^3 \rightarrow \mathcal{A}_4^3$ permutes circularly each trinucleotide $l_1l_2l_3$ as follows $\mathcal{P}(l_1l_2l_3) = l_2l_3l_1$. The k-th iterate of $\mathcal{P}$ is denoted $\mathcal{P}^k$. This map on words is also naturally extended to word sets: a permuted trinucleotide set is obtained by applying the circular permutation map $\mathcal{P}$ (or the k-th iterate of $\mathcal{P}$) to all its trinucleotides.*

Definition 2.3. *Code: 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 $i = 1, \ldots, n$.*

We consider in this paper only codes consisting of trinucleotides.

Definition 2.4. *Trinucleotide circular code: A set X_0 of trinucleotides is a trinucleotide circular code if, for each $x_1, \ldots, x_n, x'_1, \ldots, x'_m \in X_0$, $n, m \geq 1$, $p \in \mathcal{A}_4^*$, $s \in \mathcal{A}_4^+$, 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 circular code X_0 is self-complementary if, for each $y \in X_0$, $\mathcal{C}(y) \in X_0$.*

Definition 2.6. *If X_0 is a trinucleotide circular code, we denote by X_1 the permuted trinucleotide set $\mathcal{P}(X_0)$ and by X_2 the permuted trinucleotide set $\mathcal{P}^2(X_0)$.*

Definition 2.7. *A trinucleotide circular code X_0 is C^3 self-complementary if X_0, $X_1 = \mathcal{P}(X_0)$ and $X_2 = \mathcal{P}^2(X_0)$ are circular codes satisfying the following properties: $X_0 = \mathcal{C}(X_0)$ (self-complementary), $\mathcal{C}(X_1) = X_2$ and $\mathcal{C}(X_2) = X_1$.*

The concept of necklace was introduced by Pirillo [20] for circular codes in order to have an algorithmic characterization of circular codes. Let $l_1, l_2, \ldots, l_{n-1}, l_n, \ldots$ be letters in $\mathcal{A}_4$, $d_1, d_2, \ldots, d_{n-1}, d_n, \ldots$ be diletters in $\mathcal{A}_4^2$ and $n \geq 2$ be 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 $d_1l_2, d_2l_3, \ldots, d_{n-1}l_n, d_nl_{n+1} \in X$.*

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 $5LDCN$.

There are 528 self-complementary circular codes (Table 3 in [24] and Table 2(d) in [1]). There are 216 C^3 self-complementary circular codes (Tables 4a, 5a and 6a in [18] and Table 2(d) in [1]). Thus, there are $528 - 216 = 312$ circular codes which are self-complementary but not C^3 self-complementary. We denote them by $\overline{C^3}$. Using the three following propositions, we will prove that, for each $\overline{C^3}$ self-complementary circular code X_0, the sets X_1 and X_2 have both a necklace $5LDCN$, so nor X_1 neither X_2 are circular codes.

There are 28 self-complementary pairs of trinucleotides which are codified according to Table 1 (as in [24, 18]), giving to each pair the name of a letter of the english alphabet augmented by the two supplementary letters z' and z''.

$a = \{AAC, GTT\}$	$b = \{AAG, CTT\}$	$c = \{AAT, ATT\}$	$d = \{ACA, TGT\}$
$e = \{ACC, GGT\}$	$f = \{ACG, CGT\}$	$g = \{ACT, AGT\}$	$h = \{AGA, TCT\}$
$i = \{AGC, GCT\}$	$j = \{AGG, CCT\}$	$k = \{ATC, GAT\}$	$l = \{ATG, CAT\}$
$m = \{CAA, TTG\}$	$n = \{CAC, GTG\}$	$o = \{CAG, CTG\}$	$p = \{CCA, TGG\}$
$q = \{CCG, CGG\}$	$r = \{CGA, TCG\}$	$s = \{CTA, TAG\}$	$t = \{CTC, GAG\}$
$u = \{GAA, TTC\}$	$v = \{GAC, GTC\}$	$w = \{GCA, TGC\}$	$x = \{GCC, GGC\}$
$y = \{GGA, TCC\}$	$z = \{GTA, TAC\}$	$z' = \{TAA, TTA\}$	$z'' = \{TCA, TGA\}$

Table 1. The 28 self-complementary pairs of trinucleotides.

3. Results

Proposition 3.1. *If a circular code X_0 contains one of the four doublets*

$$\alpha_1 = \{a, p\}, \alpha_2 = \{b, y\}, \alpha_3 = \{e, m\}, \alpha_4 = \{j, u\}, \tag{3.1}$$

then neither its permuted set $X_1 = \mathcal{P}(X_0)$ nor its permuted set $X_2 = \mathcal{P}^2(X_0)$ are circular codes.

Proof. We prove the first case, the other cases being similar. If $\alpha_1 = \{a, p\} = \{AAC, GTT, CCA, TGG\} \subset X_0$ then $\mathcal{P}(\alpha_1) = \{ACA, TTG, CAC, GGT\} \subset X_1$. So, $A, CA, C, AC, A, CA, C, AC, A$ is a $5LDCN$ for $\mathcal{P}(\alpha_1)$, hence also for X_1 which, consequently, is not a circular code. Furthermore, $\mathcal{P}^2(\alpha_1) = \{CAA, TGT, ACC, GTG\} \subset X_2$ and $T, GT, G, TG, T, GT, G, TG, T$ is a $5LDCN$ for $\mathcal{P}^2(\alpha_1)$, hence also for X_2 which, consequently, is not a circular code. ∎

Proposition 3.2. *If a circular code X_0 contains one of the 24 triplets*

$$\beta_1 = \{a, r, w\}, \beta_2 = \{a, s, z''\}, \beta_3 = \{b, r, w\}, \beta_4 = \{b, z, z''\}, \beta_5 = \{c, s, z''\}, \beta_6 = \{c, z, z''\}, \tag{3.2}$$

$$\beta_7 = \{e, l, s\}, \beta_8 = \{e, o, r\}, \beta_9 = \{f, i, m\}, \beta_{10} = \{f, i, u\}, \beta_{11} = \{f, o, x\}, \beta_{12} = \{f, o, y\},$$
$$\beta_{13} = \{g, k, m\}, \beta_{14} = \{g, k, z'\}, \beta_{15} = \{g, l, u\}, \beta_{16} = \{g, l, z'\}, \beta_{17} = \{i, p, v\}, \beta_{18} = \{i, q, v\},$$
$$\beta_{19} = \{j, k, z\}, \beta_{20} = \{j, v, w\}, \beta_{21} = \{k, p, z\}, \beta_{22} = \{l, s, y\}, \beta_{23} = \{o, r, x\}, \beta_{24} = \{q, v, w\},$$

then neither its permuted set $X_1 = \mathcal{P}(X_0)$ nor its permuted set $X_2 = \mathcal{P}^2(X_0)$ are circular codes.

Proof. We prove the first case, the other cases being similar. If $\beta_1 = \{a, r, w\} = \{AAC, GTT, CGA, TCG, GCA, TGC\} \subset X_0$ then $\mathcal{P}(\beta_1) = \{ACA, TTG, GAC, CGT, CAG, GCT\} \subset X_1$. So $A, CA, G, AC, A, CA, G, AC, A$ is a $5LDCN$ for $\mathcal{P}(\beta_1)$, hence also for X_1 which, consequently, is not a circular code. Furthermore, $\mathcal{P}^2(\beta_1) = \{CAA, TGT, ACG, GTC, AGC, CTG\} \subset X_2$ and $T, GT, C, TG, T, GT, C, TG, T$ is a $5LDCN$ for $\mathcal{P}^2(\beta_1)$, hence also for X_2 which, consequently, is not a circular code. ∎

Proposition 3.3. *If a circular code X_0 contains one of the 23 quadruplets*

$$\gamma_1 = \{a, h, j, r\}, \gamma_2 = \{a, h, n, w\}, \gamma_3 = \{a, h, n, z'\}, \gamma_4 = \{a, l, n, y\}, \tag{3.3}$$
$$\gamma_5 = \{b, d, e, w\}, \gamma_6 = \{b, d, r, t\}, \gamma_7 = \{b, d, t, z'\}, \gamma_8 = \{b, k, p, t\},$$
$$\gamma_9 = \{c, d, t, u\}, \gamma_{10} = \{c, h, m, n\}, \gamma_{11} = \{d, e, l, t\}, \gamma_{12} = \{d, e, q, t\},$$
$$\gamma_{13} = \{d, f, p, u\}, \gamma_{14} = \{d, i, t, u\}, \gamma_{15} = \{d, p, t, x\}, \gamma_{16} = \{d, p, t, z\},$$
$$\gamma_{17} = \{e, s, t, u\}, \gamma_{18} = \{f, h, m, n\}, \gamma_{19} = \{h, j, k, n\}, \gamma_{20} = \{h, j, n, x\},$$
$$\gamma_{21} = \{h, i, m, y\}, \gamma_{22} = \{h, n, q, y\}, \gamma_{23} = \{h, n, s, y\},$$

then neither its permuted set $X_1 = \mathcal{P}(X_0)$ nor its permuted set $X_2 = \mathcal{P}^2(X_0)$ are circular codes.

Proof. We prove the first case, the other cases being similar. If $\gamma_1 = \{a, h, j, r\} = \{AAC, GTT, AGA, TCT, AGG, CCT, CGA, TCG\} \subset X_0$ then $\mathcal{P}(\gamma_1) = \{ACA, TTG, GAA, CTT, GGA, CTC, GAC, CGT\} \subset X_1$. So, $C, TT, G, GA, C, TT, G, GA, C$ is a $5LDCN$ for $\mathcal{P}(\gamma_1)$, hence also for X_1 which, consequently, is not a circular code. Furthermore, $\mathcal{P}^2(\gamma_1) = \{CAA, TGT, AAG, TTC, GAG, TCC, ACG, GTC\} \subset X_2$. Then $G, TC, C, AA, G, TC, C, AA, G$ is a $5LDCN$ for $\mathcal{P}^2(\gamma_1)$, hence also for X_2 which, consequently, is not a circular code. ∎

Table 2 reports the complete combinatorial study of the 312 $\overline{C^3}$ circular codes (self-complementary circular codes but not C^3). They are listed using the same order of Tables 4b, 5b and 6b in [18].

abcegikvxy	α_2	abcegivxyz"	α_2	acfghjopqz"	α_1	bcdegivxyz"	α_2	cdfglopqtu	β_{15}	dgkpuvwxyz'	β_{14}
abcfgjlopq	α_1	abcegkovxy	α_2	acfghlnoqy	β_{12}	bcdfgikptx	γ_8	cdflopqstu	γ_9	dhjpvwxzz'z"	β_{20}
abcflopqst	α_1	abceikvxyz	α_2	acghijknvx	γ_{19}	bcdfopqstz"	β_5	cdghijkpvx	β_{17}	djkpuvwxzz'	α_4
abciknvxyz	α_2	abceivxyzz"	α_2	acghijkpvx	α_1	bcdgikptvx	β_{17}	cdghijpvxz"	β_{17}	djopqrsuz'z"	α_4
abcknovxyz	α_2	abcekovxyz	α_2	acghijlnqv	β_{18}	bcefgijlmq	α_3	cdgijkpuvx	α_4	djpqsuvwz'z"	α_4
abclopqstv	α_1	abcfgijlpq	α_1	acghijlpqv	α_1	bcefglmoqt	α_3	cdgikptuvx	β_{17}	djpuvwxzz'z"	α_4
acegijkuvx	α_4	abcfgijpqz"	α_1	acghijpvxz"	α_1	bcefglmoqy	α_2	cefghijkmx	α_3	dkpuvwxyzz'	β_{21}
acfghijlpq	α_1	abcfgjopqz"	α_1	acgijknuvx	α_4	bcefjlmoqs	α_3	cefghikmxy	α_3	dpqstuvwz'z"	β_{24}
acfghijpqz"	α_1	abcfglnoqy	α_2	acgijkpuvx	α_1	bceflmoqsy	α_2	cefghilmqy	α_3	dptuvwxzz'z"	γ_{15}
achijknvxz	β_{19}	abcfglopqt	α_1	acgikptuvx	α_1	acikptuvxz	α_1	cefghjlmoq	α_3	eghlmoqryz'	α_3
acijknuvxz	α_4	abcfjlopqs	α_1	achijkpvxz	α_1	bcfglmnoqy	α_2	cefghlmoqy	α_3	eghmoqryz'z"	α_3
ahnrwxyzz'z"	β_1	abcfjopqsz"	α_1	achijnvxzz"	β_6	bcflmnoqsy	α_2	cefgijkmux	α_3	eghmvwxyz'z"	α_3
ahprwxyzz'z"	α_1	abcflnoqsy	α_2	achijpvxzz"	α_1	bcgijlmnqv	β_{18}	cefgijlmqu	α_3	egkmtuvwxz'	α_3
akptuvwxzz'	α_1	abcfopqstz"	α_1	achinvxyzz"	β_6	bdektvwxzz'	γ_5	cefgikmtux	α_3	egkmuvwxyz'	α_3
apuvwxyzz'z"	α_1	abcgijlpqv	α_1	acijkpuvxz	α_1	bdeoqrstz'z"	β_8	cefgjlmoqu	α_3	egmruwxyz'z"	α_3
bcdefgikxy	α_2	abcgiknvxy	α_2	bcfgijlmnq	β_9	bdeoqrsyz'z"	α_2	cefgkmotux	α_3	egmuvwxyz'z"	α_3
bcdefgixyz"	α_2	abcgikptvx	α_1	aehrwxyzz'z"	β_1	bdeqrstwz'z"	β_3	cefglmoqtu	α_3	ehlmoqrsyz'	α_3
bcdefloqst	β_7	abcgjlopqv	α_1	ahjlnoqrsz'	γ_1	bdeqrswyz'z"	α_2	ceflmoqstu	α_3	ehmoqrsyz'z"	α_3
bcefgjlmoq	α_3	abcgknovxy	α_2	ahjnoqrsz'z"	β_2	bderwxyzz'z"	α_2	ceghijkmvx	α_3	ehmorxyzz'z"	α_3
bceflmoqst	α_3	abcglopqtv	α_1	ahjnrwxzz'z"	β_1	ahlnoqrsyz'	β_{22}	ceghikmvxy	α_3	ehmvwxyzz'z"	α_3
bdpqrstwz'z"	β_3	abcijknvxz	β_{19}	ahjnvwxzz'z"	β_{20}	bdevwxyzz'z"	α_2	cegijkmuvx	α_3	ekmouvxyzz'	α_3
bdpqrswyz'z"	α_2	abcikptvxz	α_1	bdetvwxzz'z"	β_4	bdjpqrswz'z"	β_3	cegikmtuvx	α_3	ekmuvwxyzz'	α_3
blmnoqrsyz'	α_2	abcinvxyzz"	α_2	ahjpqrswz'z"	α_1	bdkptvwxzz'	β_{21}	cegkmotuvx	α_3	elmoqrstuz'	α_3
bmopqrsyz'z"	α_2	abcjlopqsv	α_1	ahjprwxzz'z"	α_1	bdopqrstz'z"	γ_6	cfghijlmpq	β_9	emoqrstuz'z"	α_3
cdefgijkux	α_4	abkptvwxzz'	α_1	ahjpvwxzz'z"	α_1	bdopqrsyz'z"	α_2	cfghilmnqy	β_9	emoruxyzz'z"	α_3
cdefgiktux	β_{10}	ablnoqrsyz'	α_2	ahjopqrsz'z"	α_1	bdpqstvwz'z"	β_{24}	cfghjlmnoq	γ_{10}	emotuvxzz'z"	α_3
cefghijlmq	α_3	abnoqrstz'z"	β_2	ahnoqrsyz'z"	β_2	bdprwxyzz'z"	α_2	cfghlmnoqy	β_{12}	emouvxyzz'z"	α_3
cfghijlmnq	β_9	abnoqrsyz'z"	α_2	ahnorxyzz'z"	β_{23}	bdptvwxzz'z"	β_4	cfgijlmnqu	α_4	emruwxyzz'z"	α_3
dgjpqruwz'z"	α_4	abnoqstvz'z"	β_2	ahnvwxyzz'z"	γ_2	bdpvwxyzz'z"	α_2	cfgjlmnoqu	α_4	emtuvwxzz'z"	α_3
djpqrsuwz'z"	α_4	abnorxyzz'z"	α_2	ahopqrsyz'z"	α_1	belmoqrstz'	α_3	cfglmnoqtu	β_{15}	ghjlmnoqrz'	β_{16}
dkptuvwxzz'	β_{21}	abnotvxzz'z"	β_4	ahpqrswyz'z"	α_1	belmoqrsyz'	α_2	cghijkmnvx	β_{13}	ghlmnoqryz'	β_{16}
eghmrwxyz'z"	α_3	abnovxyzz'z"	α_2	ahpvwxyzz'z"	α_1	bemoqrstzz"	α_3	cghijlmnqv	β_{18}	ghlmopqryz'	β_{16}
ehmrwxyzz'z"	α_3	abnrwxyzz'z"	α_2	ajknuvwxzz'	α_4	bemoqrsyz'z"	α_2	cghikmnvxy	β_{13}	gjlmnoqruz'	α_4
ekmotuvxzz"	α_3	abntvwxzz'z"	β_4	ajkpuvwxzz'	α_1	blmopqrsyz'	α_2	cgijkmnuvx	α_4	gjlmopqruz'	α_4
ekmtuvwxzz'	α_3	abnvwxyzz'z"	α_2	ajnuvwxzz'z"	α_4	bmnoqrsyz'z"	α_2	cgijlmnquv	α_4	gjmopqruz'z"	α_4
emuvwxyzz'z"	α_3	abopqrstz'z"	α_1	ajpuvwxzz'z"	α_1	bmpqrstwz'z"	β_3	cgikmntuvx	β_{13}	gjmpqruwz'z"	α_4
hlmnoqrsyz'	β_{22}	abopqrsyz'z"	α_1	akpuvwxyzz'	α_1	bmpqrswyz'z"	α_2	cgjlmnoquv	α_4	hjmnrwxzz'z"	γ_{20}
jlmnoqrsuz'	α_4	abopqstvz'z"	α_1	anoruxyzz'z"	β_{23}	bmpqstvwz'z"	β_{24}	cgkmnotuvx	β_{13}	hjmnvwxzz'z"	β_{20}
jlmnoqsuvz'	α_4	abpqrstwz'z"	α_1	anruwxyzz'z"	β_1	cdefghloqy	β_{12}	cglmnoqtuv	β_{15}	hlmopqrsyz'	β_{22}
jmopqrsuz'z"	α_4	abpqrswyz'z"	α_1	apruwxyzz'z"	α_1	cdefghoqyz"	β_{12}	chijkmnvxz	β_{19}	hmnoqrsyz'z"	γ_{22}
abcefgkotx	β_{11}	abpqstvwz'z"	α_1	aptuvwxzz'z"	α_1	cdefgijlqu	α_4	chikmnvxyz	γ_{10}	hmnorxyzz'z"	β_{23}
abcgijlnqv	β_{18}	abprwxyzz'z"	α_1	bcdefgkotx	β_{11}	cdefgikuxy	β_{10}	cijkmnuvxz	α_4	jkmnuvwxzz'	α_4
acehivxyzz"	β_6	abptvwxzz'z"	α_1	bcdefgkoxy	α_2	cdefgjloqu	α_4	deghoqryz'z"	β_8	jlmopqrsuz'	α_4
bcdfjopqsz"	β_5	abpvwxyzz'z"	α_1	bcdefgloqt	γ_{11}	cdefgkotux	β_{11}	degktuvwxz'	β_{14}	jlmopqsuvz'	α_4
degkuvwxyz'	β_{14}	acefgijkux	α_4	bcdefgloqy	α_2	cdefgloqtu	β_{15}	dehoqrsyz'z"	β_8	jmnoqrsuz'z"	α_4
ghjlmopqrz'	β_{16}	acefgiktux	β_{10}	bcdefgoqyz"	β_{12}	cdefloqstu	β_7	deoqrstuz'z"	β_8	jmnoqsuvz'z"	α_4
mnoruxyzz'z"	β_{23}	acefgkotux	β_{11}	bcdefjloqs	β_7	cdegijkuvx	α_4	deqrstuwz'z"	γ_{12}	jmnuvwxzz'z"	α_4
mpqstuvwz'z"	β_{24}	acehijkvxz	β_{19}	bcdefjoqsz"	β_5	cdegiktuvx	γ_9	dghjpvwxz'z"	β_{20}	jmopqsuvz'z"	α_4
abcefgikxy	α_2	acehijvxzz"	β_6	bcdefloqsy	α_2	cdfgijkpux	α_4	dgjkpuvwxz'	α_4	jmpqrsuwz'z"	α_4
abcefgixyz"	α_2	aceijkuvxz	α_4	bcdefoqstz"	β_5	cdfgijlpqu	α_4	dgjopqruz'z"	α_4	jmpqsuvwz'z"	α_4
abcefgkoxy	α_2	acfghilnqy	γ_4	bcdefoqsyz"	α_2	cdfgikptux	β_{10}	dgjpuvwxz'z"	α_4	kmptuvwxzz'	β_{21}
abcefloqst	β_7	acfghjlopq	α_1	bcdegikvxy	α_2	cdfgjlopqu	α_4	dgkptuvwxz'	β_{14}	lmnoqrsuyz'	β_{22}

Table 2. Complete combinatorial study of the 312 $\overline{C^3}$ circular codes. Only one "forbidden configuration"with one doublet (3.1), triplet (3.2) or quadruplet (3.3) is given for each $\overline{C^3}$ circular code.

Using the results of [1, 18] and Table 2, the following proposition can be deduced.

Proposition 3.4. *If a set of 20 trinucleotides is a self-complementary circular code then either its two permuted sets are both circular codes or its two permuted sets are both non-circular codes.*

Proof. Let X_0 be a self-complementary circular code among the 528 ones. If X_0 is one of the 216 C^3 self-complementary circular codes then both its permuted sets X_1 and X_2 are circular codes [1, 18]. If X_0 is one of the 312 $\overline{C^3}$ self-complementary circular codes then a "forbidden configuration"(a $5LDCN$ necklace) is identified (Table 2). If this "forbidden configuration"is a doublet (triplet and quadruplet, respectively) then Proposition 2 (3 and 4, respectively) applies. All the 312 $\overline{C^3}$ circular codes X_0 have a "forbidden configuration"proving that their permuted sets X_1 and X_2 are all non-circular codes. ∎

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] D.G. Arquès and C.J. Michel. A circular code in the protein coding genes of mitochondria. *J. Theor. Biol.* **189**, 273-290 (1997).

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

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

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

[6] F.H.C. Crick, J.S. Griffith, L.E. Orgel. Codes without commas. *Proc. Natl. Acad. Sci. USA* **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. *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. *Biologiske Meddel Danske Vidensk Selsk* **23**, 1-34 (1958).

[11] R. Jolivet, F. Rothen. Peculiar symmetry of DNA sequences and evidence suggesting its evolutionary origin in a primeval genetic code. In First European Workshop in Exo-/Astro-Biology. Eds. P. Ehrenfreund, O. Angerer, B. Battrick. pp. 173-176, Noordwijk, The Netheralnds, (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**, 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. Inf. 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. Proceedings of the 21st International Conference on Advanced Information Networking and Applications Workshops/Symposia (AINAW '07), vol. **2**, pp. 745-751, (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**, 17-26 (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.* Eds. C. Pellegrini, P. Cerrai, P. Freguglia, V. Benci, G. Israel, Kluwer, Boston, Mass, USA, 2003.

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

[22] G. Pirillo. Some remarks on prefix and suffix codes. *Pure Math. Appl.* **19**, 53-59 (2008).

[23] G. Pirillo. Non sharing border codes. *Adv. Appl. Math.* **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).