Trinucleotides

Giuseppe Pirillo

Dipartimento di matematica “Ulisse Dini”

26 03 2013

UUU	Phe
UUC	Phe
UUA	Leu
UUG	Leu
CUU	Leu
CUC	Leu
CUA	Leu
CUG	Leu
AUU	Ile
AUC	Ile
AUA	Ile
AUG	Met
GUU	Val
GUC	Val
GUA	Val
GUG	Val

UCU	Ser
UCC	Ser
UCA	Ser
UCG	Ser
CCU	Pro
CCC	Pro
CCA	Pro
CCG	Pro
ACU	Thr
ACC	Thr
ACA	Thr
ACG	Thr
GCU	Ala
GCC	Ala
GCA	Ala
GCG	Ala

UAU	Tyr
UAC	Tyr
UAA	STOP
UAG	STOP
CAU	His
CAC	His
CAA	Gln
CAG	Gln
AAU	Asn
AAC	Asn
AAA	Lys
AAG	Lys
GAU	Asp
GAC	Asp
GAA	Glu
GAG	Glu

UGU	Cys
UGC	Cys
UGA	STOP
UGG	Trp
CGU	Arg
CGC	Arg
CGA	Arg
CGG	Arg
AGU	Ser
AGC	Ser
AGA	Arg
AGG	Arg
GGU	Gly
GGC	Gly
GGA	Gly
GGG	Gly

Example

Ibis redibis non morieris in bello

Ibis, redibis, non morieris in bello

Ibis, redibis, non morieris in bello

Ibis, redibis non, morieris in bello

Example

ourteawashot

Example

ourteawashot

We have the following possibilities:

Example

ourteawashot

We have the following possibilities:

- our tea was hot

Example

ourteawashot

We have the following possibilities:

- our tea was hot
- o urt eaw ash ot

Example

ourteawashot

We have the following possibilities:

- our tea was hot
- o urt eaw ash ot
- ou rte awa sho t

Example

ourteawashot

We have the following possibilities:

- our tea was hot
- o urt eaw ash ot
- ou rte awa sho t

Example

ourteawashot

We have the following possibilities:

- our tea was hot
- o urt eaw ash ot
- ou rte awa sho t

The first one is an English statement that is meaningful and, in appropriate context, it is even true!

Example

ourteawashot

We have the following possibilities:

- our tea was hot
- o urt eaw ash ot
- ou rte awa sho t

The first one is an English statement that is meaningful and, in appropriate context, it is even true!
The others two seems not meaningful and the problem if they are true or not ... is not meaningful!

Example

Similarly for the italian and mathematical sentence:

Example

Similarly for the italian and mathematical sentence:

- DUE PIU DUE FAN TRE PIU UNO

Example

Similarly for the italian and mathematical sentence:

- DUE PIU DUE FAN TRE PIU UNO
- D UEP IUD UEF ANT REP IUU NO

Example

Similarly for the italian and mathematical sentence:

- DUE PIU DUE FAN TRE PIU UNO
- D UEP IUD UEF ANT REP IUU NO
- DU EPI UDU EFA NTR EPI UUN O

Roughly speaking, one can consider coding regions of DNA as “digital computer messages”. where the coding words are strings of three symbols, i.e. trinucleotides. The meanings of the trinucleotides in coding regions of DNA are the coded amino acids which are sequentially assembled in order to form the proteins. It must be point out that the decoding frequency of some of these processes can attain more than 20 amino acids per second. So, as in the transmission of any digital message, a faithful and efficient protein synthesis needs a very good synchronization of the decoding process with the correct reading frame of codons. This ability is the *reading frame maintenance*.

Definition

Let A be a set which we call an *alphabet*; a *word* w on the alphabet A is a finite sequence of elements of A

$$w = (a_1, a_2, \ldots, a_n), \quad a_i \in A.$$

Definition

Let A be a set which we call an *alphabet*; a *word* w on the alphabet A is a finite sequence of elements of A

$$w = (a_1, a_2, \ldots, a_n), \quad a_i \in A.$$

The genetic alphabet

$$\mathcal{A}_4 = \{A, C, G, T\}$$

- $\mathcal{A}_4^*$ (respectively $\mathcal{A}_4^+$) is the set of words (respectively non empty words) over $\mathcal{A}_4$,

Definition

Let A be a set which we call an *alphabet*; a *word* w on the alphabet A is a finite sequence of elements of A

$$w = (a_1, a_2, \ldots, a_n), \quad a_i \in A.$$

The genetic alphabet

$$\mathcal{A}_4 = \{A, C, G, T\}$$

- $\mathcal{A}_4^*$ (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 (or *diletters*)

Definition

Let A be a set which we call an *alphabet*; a *word* w on the alphabet A is a finite sequence of elements of A

$$w = (a_1, a_2, \ldots, a_n), \quad a_i \in A.$$

The genetic alphabet

$$\mathcal{A}_4 = \{A, C, G, T\}$$

- $\mathcal{A}_4^*$ (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 (or *diletters*) and
- $\mathcal{A}_4^3$ is the set of the 64 words of length three (or *trinucleotides*).

Definition

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., $\mathcal{C}(AAC) = GTT$.

Definition

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., $\mathcal{C}(AAC) = GTT$.

Definition

The (left) *circular permutation map* $\mathcal{P}$: $\mathcal{A}_4^3 \to \mathcal{A}_4^3$ permutes circularly each trinucleotide $l_1 l_2 l_3$ as follows $\mathcal{P}\ (l_1 l_2 l_3) = l_2 l_3 l_1$.

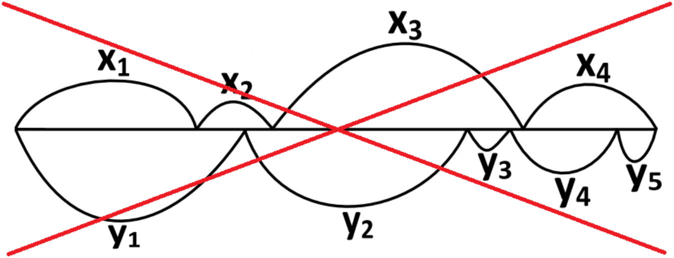
Definition

Code: A set X 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

Code: A set X 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 $i = 1, \ldots, n$.

AAA
ATC, TCA
ATA, TAT
ACA, CAC


x1
x2
x3
x4
y1
y2
y3
y4
y5

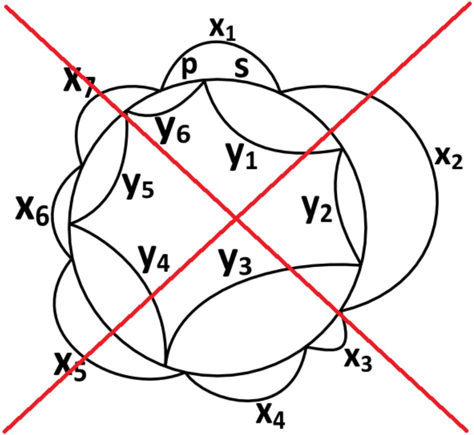
Definition

A code X is $\mathrm{circular}$ 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

A code X is $\mathrm{circular}$ 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 $i = 1, \ldots, n$.

AAA
ATC, TCA
ATA, TAT
ACA, CAC


x1
p
s
x7
y6
y1
x2
y5
x6
y2
y4
y3
x5
x3
x4

Definition

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

Definition

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

Definition

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

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

Definition

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

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 $\mathcal{C}(X_2) = X_1$).

Combinatorial algorithms

Definition

Letter Diletter Necklaces (*LDN*): We say that the ordered sequence l_1, d_1, l_2, d_2, $\ldots$, d_{n-1}, l_n, d_n is an *nLDN* for a subset $X \subset \mathcal{A}_4^3$ if $l_1 d_1$, $l_2 d_2$, $\ldots$, $l_n d_n \in X$ and $d_1 l_2$, $d_2 l_3$, $\ldots$, $d_{n-1} l_n \in X$.

Combinatorial algorithms

Definition

Letter Diletter Necklaces (LDN): We say that the ordered sequence l_1, d_1, l_2, d_2, $\ldots$, d_{n-1}, l_n, d_n is an $nLDN$ for a subset $X \subset \mathcal{A}_4^3$ if $l_1 d_1$, $l_2 d_2$, $\ldots$, $l_n d_n \in X$ and $d_1 l_2$, $d_2 l_3$, $\ldots$, $d_{n-1} l_n \in X$.

Definition

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 $l_1 d_1$, $l_2 d_2$, $\ldots$, $l_n d_n \in X$ and $d_1 l_2$, $d_2 l_3$, $\ldots$, $d_{n-1} l_n$, $d_n l_{n+1} \in X$.

$$\boxed{l_1}\ \boxed{\quad d_1 \quad}\ \boxed{l_2}\ \boxed{\quad d_2 \quad}\ \cdots\ \boxed{l_n}\ \boxed{\quad d_n \quad}\ \boxed{l_{n+1}}$$

$$\boxed{l_1}\ \boxed{\quad d_1 \quad}\ \boxed{l_2}\ \boxed{\quad d_2 \quad}\ \cdots\ \boxed{l_n}\ \boxed{\quad d_n \quad}\ \boxed{l_{n+1}}$$

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

Definition

Diletter Letter Necklaces (DLN): We say that the ordered sequence d_1, l_1, d_2, l_2, $\ldots$ l_{n-1}, d_n, l_n is an $nDLN$ for a subset $X \subset \mathcal{A}_4^3$ if $d_1 l_1$, $d_2 l_2$, $\ldots$, $d_n l_n \in X$ and $l_1 d_2$, $l_2 d_3$, $\ldots$, $l_{n-1} d_n \in X$.

Definition

Diletter Letter Necklaces (DLN): We say that the ordered sequence d_1, l_1, d_2, l_2, ... l_{n-1}, d_n, l_n is an $nDLN$ for a subset $X \subset \mathcal{A}_4^3$ if $d_1 l_1$, $d_2 l_2$, ..., $d_n l_n \in X$ and $l_1 d_2$, $l_2 d_3$, ..., $l_{n-1} d_n \in X$.

Definition

Diletter Letter Continued Necklaces ($DLCN$): We say that the ordered sequence d_1, l_1, d_2, l_2, ... l_{n-1}, d_n, l_n, d_{n+1} is an $(n+1)DLCN$ for a subset $X \subset \mathcal{A}_4^3$ if $d_1 l_1$, $d_2 l_2$, ..., $d_n l_n \in X$ and $l_1 d_2$, $l_2 d_3$, ..., $l_{n-1} d_n$, $l_n d_{n+1} \in X$.

In 1995 D. G. Arquès and C. J. Michel presented this set of trinucleotides:

$$X_0 = \{AAC, AAT, ACC, ATC, ATT, CAG, CTC, CTG, GAA, GAC,$$

$$GAG, GAT, GCC, GGC, GGT, GTA, GTC, GTT, TAC, TTC\}.$$

X_0 is a circular code with remarkable properties.
Now we show with some examples how it is important in reading frame maintenance. **We consider only words of X_0^* and factors of words of X_0^*.**

Consider the following word of length 12 $AGGTAATTACCA$

Consider the following word of length 12 $AGGTAATTACCA$ and its three possible factorization f_0, f_1 and f_2 (denoted by dots):
$f_0 = AGG.TAA.TTA.CCA$;

Consider the following word of length 12 $AGGTAATTACCA$ and its three possible factorization f_0, f_1 and f_2 (denoted by dots):

$f_0 = AGG.TAA.TTA.CCA$;

$f_1 = A.GGT.AAT.TAC.CA$;

Consider the following word of length 12 $AGGTAATTACCA$ and its three possible factorization f_0, f_1 and f_2 (denoted by dots):

$f_0 = AGG.TAA.TTA.CCA$;

$f_1 = A.GGT.AAT.TAC.CA$;

$f_2 = AG.GTA.ATT.ACC.A$.

Consider the following word of length 12 $AGGTAATTACCA$ and its three possible factorization f_0, f_1 and f_2 (denoted by dots):

$f_0 = AGG.TAA.TTA.CCA$;

$f_1 = A.GGT.AAT.TAC.CA$;

$f_2 = AG.GTA.ATT.ACC.A$.

Even if one factorization, namely f_0, is incompatible with Arquès and Michel circular code, there remain two others, namely f_1 and f_2, that are compatible with Arquès and Michel circular code. So a window of length 12 is not enough to retrieve of the good reading frame.

Now, consider the following sequence of length 13

$$AGGTAATTACCAG$$

Now, consider the following sequence of length 13

$$AGGTAATTACCAG$$

and the following three factorizations of it:

$g_0 = AGG.TAA.TTA.CCA.G;$

Now, consider the following sequence of length 13

$$AGGTAATTACCAG$$

and the following three factorizations of it:

$g_0 = AGG.TAA.TTA.CCA.G$;

$g_1 = A.GGT.AAT.TAC.CAG$;

Now, consider the following sequence of length 13

$$AGGTAATTACCAG$$

and the following three factorizations of it:

$g_0 = AGG.TAA.TTA.CCA.G$;

$g_1 = A.GGT.AAT.TAC.CAG$;

$g_2 = AG.GTA.ATT.ACC.AG$.

Now, consider the following sequence of length 13

$$AGGTAATTACCAG$$

and the following three factorizations of it:

$g_0 = AGG.TAA.TTA.CCA.G$;

$g_1 = A.GGT.AAT.TAC.CAG$;

$g_2 = AG.GTA.ATT.ACC.AG$.

Only one of them, namely g_1, is compatible with Arquès and Michel circular code. So a window of length 13 is enough to retrieve of the good reading frame.

This is just an example. But in their papers Arquès and Michel proved formally this statement (namely, *a window of length 13 is enough to retrieve of the good reading frame*) and a proposition on necklaces confirms it.

Definition

Trinucleotide comma-free code: A trinucleotide code X is comma-free if, for each $y \in X$ and $u, v \in \mathcal{A}_4^*$ such that $uyv = x_1 \cdots x_n$ with $x_1, \ldots, x_n \in X$, $n \geq 1$, it holds that $u, v \in X^*$.

Definition

Trinucleotide comma-free code: A trinucleotide code X is comma-free if, for each $y \in X$ and $u, v \in \mathcal{A}_4^*$ such that $uyv = x_1 \cdots x_n$ with $x_1, \ldots, x_n \in X$, $n \geq 1$, it holds that $u, v \in X^*$.

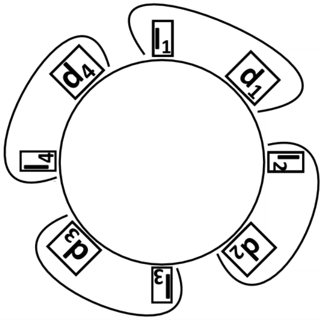
Proposition

Let X be a subset of $\mathcal{A}_4^3$. The following conditions are equivalent:
a) X is a comma-free code;
b) X has no $2LDN$ and no $2DLN$.

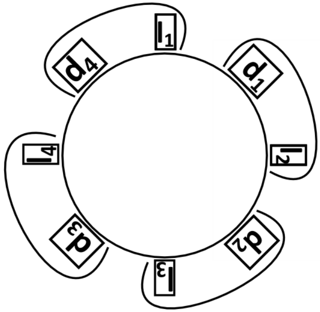
Proposition

Let X be a trinucleotide code. The following conditions are equivalent:
(i) X is a circular code.
(ii) X has no 5 Letter Diletter Continued Necklaces.

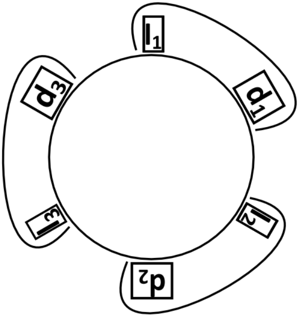
We give here a short idea of the proof.



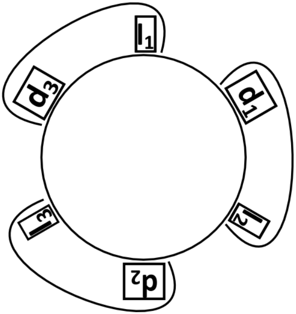
(a) First decomposition



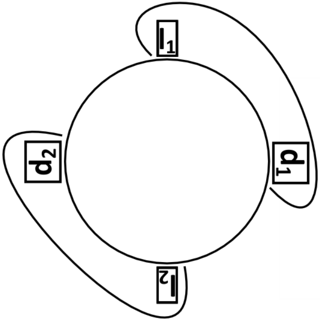
(b) Second decomposition



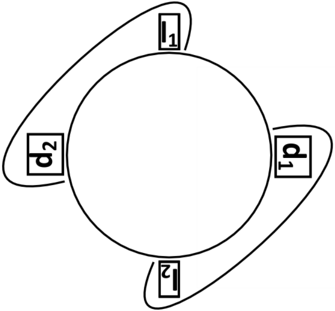
(c) First decomposition



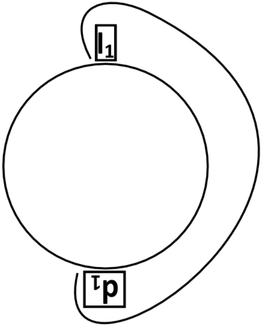
(d) Second decomposition



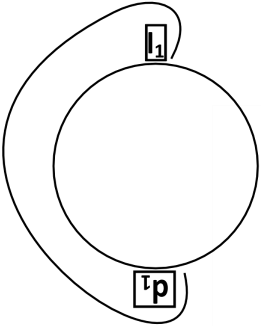
(e) First decomposition



(f) Second decomposition



(g) First decomposition



(h) Second decomposition

In 2005 the self-complementary trinucleotide circular codes were studied and, coding with a suitable alphabet the self-complementary trinucleotide pairs, were presented in the following tables the growth function of the self-complementary trinucleotide circular codes and the complete list of all the self-complementary trinucleotide circular codes.

2	4	6	8	10	12	14	16	18	20
28	334	2176	8294	19100	27264	24324	13240	4032	528

Definition

Let X be a trinucleotide code. For $k \in \{2, 3, 4, 5\}$, we say that X belongs to the class C^{kLDN} if X has no $kLDN$ and that X belongs to the class C^{kDLN} if X has no $kDLN$. Similarly, for $k \in \{3, 4, 5\}$, we say that X belongs to the class C^{kLDCN} if X has no $kLDCN$ and that X belongs to the class C^{kDLCN} if X has no $kDLCN$.

Definition

Let X be a trinucleotide code. For $k \in \{2,3,4,5\}$, we say that X belongs to the class C^{kLDN} if X has no $kLDN$ and that X belongs to the class C^{kDLN} if X has no $kDLN$. Similarly, for $k \in \{3,4,5\}$, we say that X belongs to the class C^{kLDCN} if X has no $kLDCN$ and that X belongs to the class C^{kDLCN} if X has no $kDLCN$.

Notation

$I^n = C^{nLDN} \cap C^{nDLN}$, $I^nC = C^{nLDCN} \cap C^{nDLCN}$, $U^n = C^{nLDN} \cup C^{nDLN}$, $U^nC = C^{nLDCN} \cup C^{nDLCN}$.

C. Michel, G. Pirillo and M. Pirillo A relation between trinucleotide comma-free codes and trinucleotide circular codes Theoret. Comput. Sci., 401, 2008, 17–26.

C^{2LDN} 0	C^{3LDCN} 96	C^{3LDN} 96	C^{4LDCN} 96	C^{4LDN} $64+96$	C^{5LDCN} $368+64+96$	C^{5LDN} $368+64+96$
C^{2DLN} 0	C^{3DLCN} 0	C^{3DLN} 96	C^{4DLCN} $64+96$	C^{4DLN} $64+96$	C^{5DLCN} $64+96$	C^{5DLN} $368+64+96$
I^2 0	I^3C 0	I^3 96	I^4C 96	I^4 $64+96$	I^5C $64+96$	I^5 $368+64+96$
U^2 0	U^3C 96	U^3 96	U^4C $64+96$	U^4 $64+96$	U^5C $368+64+96$	U^5 $368+64+96$

C^{2LDN}	C^{3LDCN}	C^{3LDN}	C^{4LDCN}	C^{4LDN}	C^{5LDCN}	C^{5LDN}
0	56	56	56	$56+56$	$104+56+56$	$104+56+56$
C^{2DLN}	C^{3DLCN}	C^{3DLN}	C^{4DLCN}	C^{4DLN}	C^{5DLCN}	C^{5DLN}
0	0	56	$56+56$	$56+56$	$56+56$	$104+56+56$
I^2	I^3C	I^3	I^4C	I^4	I^5C	I^5
0	0	56	56	$56+56$	$56+56$	$104+56+56$
U^2	U^3C	U^3	U^4C	U^4	U^5C	U^5
0	56	56	$56+56$	$56+56$	$104+56+56$	$104+56+56$

C.J. Michel, G. Pirillo, M.A. Pirillo. A classification of 20-trinucleotide circular codes, Information and Computation, 212, 2012, 55-63.

C^{2LDN} α_1	C^{3LDCN} α_4	C^{3LDN} α_7	C^{4LDCN} α_9	C^{4LDN} α_{12}	C^{5LDCN} α_{14}	C^{5LDN} α_{14}
C^{2DLN} α_1	C^{3DLCN} α_5	C^{3DLN} α_7	C^{4DLCN} α_{10}	C^{4DLN} α_{12}	C^{5DLCN} α_{13}	C^{5DLN} α_{14}
I^2 α_2	I^3C α_3	I^3 α_6	I^4C α_8	I^4 α_{11}	I^5C α_{13}	I^5 α_{14}
U^2 α_3	U^3C α_6	U^3 α_8	U^4C α_{11}	U^4 α_{13}	U^5C α_{14}	U^5 α_{14}

C^{2LDN} 1,584	C^{3LDCN} 294,912	C^{3LDN} 423,552	C^{4LDCN} 5,088,264	C^{4LDN} 5,528,688	C^{5LDCN} 12,964,440	C^{5LDN} 12,964,440
C^{2DLN} 1,584	C^{3DLCN} 4,920	C^{3DLN} 423,552	C^{4DLCN} 578,496	C^{4DLN} 5,528,688	C^{5DLCN} 5,940,648	C^{5DLN} 12,964,440
I^2 408	I^3C 2,760	I^3 297,072	I^4C 550,032	I^4 5,116,728	I^5C 5,940,648	I^5 12,964,440
U^2 2,760	U^3C 297,072	U^3 550,032	U^4C 5,116,728	U^4 5,940,648	U^5C 12,964,440	U^5 12,964,440

In 2011 was extended a hierarchy relation already presented in 2008, the chain of inclusion that starts from trinucleotide comma-free codes and ends by trinucleotide circular codes, adding on the left side (at the beginning of the chain) two classes of codes, called *DLD* and *LDL* codes, which are stronger than the comma-free codes.

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

.

Proposition

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.

Proposition

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.

Proposition

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).

Proposition

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

A computer calculus

Forbidden configurations.

2	44	64	153	0	0	0	0

4	66	24	846	936	7236	...	...	0

C. Michel, G. Pirillo, A permuted set of a trinucleotide circular code coding the 20 amino acids in variant nuclear codes, Journal of Theoretical Biology, 319, 2013, 116121.

Theorem

The following set Y of 20 trinucleotides

$$Y = \{ACG, ACT, AGA, AGG, AGT, ATA, ATC, CAA, CAC, CAG, CCT, GCC, GCG, GCT, GGT, TCG, TCT, TGA, TGT, TTA\}$$

is a circular code (maximal). More precisely, Y is the $11,056,585th$ among $12,964,440$ maximal circular codes (in the lexicographical order) of 20 trinucleotides and belongs to the classes $C^{5LDN} = C^{5LDCN} = C^{5DLN}$.

Theorem

The trinucleotide circular code Y has a permuted set $\mathcal{P}^2(Y)$ of 20 trinucleotides

$$\mathcal{P}^2(Y) = \{AAG, AAT, ACA, ATG, ATT, CAT, CCA, CGC, GAC, GAG, GCA, GGC, GTC, TAC, TAG, TCC, TGC, TGG, TTC, TTG\}$$

which codes the 20 amino acids in the variant nuclear codes 6 and 15.

The **ennone**