Research Article

Transition and Transversion on the Common Trinucleotide Circular Code

Emmanuel Benard and Christian J. Michel

Equipe de Bioinformatique Théorique, ICube, Université de Strasbourg, CNRS, 300 boulevard Sébastien Brant, 67400 Illkirch, France

Correspondence should be addressed to Christian J. Michel; c.michel@unistra.fr

Received 15 February 2013; Accepted 22 April 2013

Academic Editor: Alessandra Luminì

Copyright © 2013 E. Benard and C. J. Michel. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

In 1996, a trinucleotide circular code which is maximum, self-complementary, and $C^3$, called $X_0$, was identified statistically on a large gene population of eukaryotes and prokaryotes (Arquès and Michel (1996)). Transition and transversions I and II are classical molecular evolution processes. A comprehensive computer analysis of these three evolution processes in the code $X_0$ shows some new results; in particular (i) transversion I on the 2nd position of any subset of trinucleotides of $X_0$ generates trinucleotide circular codes which are always $C^3$ and (ii) transversion II on the three positions of any subset of trinucleotides of $X_0$ yields no trinucleotide circular codes. These new results extend our theory of circular code in genes to its evolution under transition and transversion.

1. Introduction

We continue our study of properties of trinucleotide circular codes [1–5], trinucleotide comma-free codes [1, 6], strong trinucleotide circular codes [7], and the common trinucleotide circular code $X_0$ identified in genes [8] (see also the recent statistical analysis by [9]) which could be a translation code [10]. 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 (called genetic code), more precisely a uniform code but not a circular code (see Remark 2). In the past 50 years, codes, comma-free codes, and circular codes have been mathematical objects 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, for example, [11–13]. In order to have an intuitive meaning of these notions, codes are written on a straight line while comma-free codes and circular codes are written on a circle, but in both cases, unique decipherability is required. Circular codes only belong to some subsets of the 64 trinucleotide set while comma-free codes are even more constrained subsets of circular codes [1].

Before the discovery of the genetic code, Crick et al. [11] proposed a maximum comma-free code of 20 trinucleotides for coding the 20 amino acids. This comma-free code turned out to be invalid (see, e.g., [14]). In 1996, a maximum circular code $X_0$ of 20 trinucleotides was identified statistically on a large gene population of eukaryotes and also on a large gene population of prokaryotes [8]

$$X_0 = \{ AAC, AAT, ACC, ATC, ATT, CAG, CTG, CTA, GAA, GAC, GAG, GAT, GCC, GGC, GTG, GTA, GTC, GTT, TAC, TTC \} .$$

This code $X_0$ has remarkable mathematical properties as it is a $C^3$ self-complementary maximum circular code (see the following). Since 1996, its properties have been studied in detail by different authors, for example, [9, 15–21]. Transition and transversions I and II are classical molecular evolution processes, for example, [22]. By using an algorithm based on the necklace, we perform here a comprehensive computer analysis of these three evolution processes in the code $X_0$. Some new results are identified with the code $X_0$ by computer analysis; in particular (i) transversion I on the 2nd position of any subset of trinucleotides of $X_0$ generates trinucleotide circular codes which are always $C^3$ and (ii) transversion II on the three positions of any subset of trinucleotides of $X_0$ yields no trinucleotide circular codes.
2. Preliminaries

The classical notions of language theory and codes can be found in [23, 24]. Let $A_4 = \{A, C, G, T\}$ denote the genetic alphabet, lexicographically ordered by $A < C < G < T$. The set of words (nonempty words, resp.) on $A_4$ is denoted by $A_4^*$ ($A_4^+$, resp.). The set of the 16 words of length 2 (dinucleotides or letter pairs) on $A_4$ is denoted by $A_4^2 = \{AA, AC, \ldots, TT\}$. The set of the 64 words of length 3 (trinucleotides or letter triplets) on $A_4$ is denoted by $A_4^3 = \{AAA, AAC, \ldots, TTT\}$.

**Definition 1.** A subset $X \subset A_4^*$ is a code on $A_4$ 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 $i = 1, \ldots, n$.

**Remark 2.** $A_4^3$ is a code.

Any nonempty subset of $A_4^3$ is a code called here trinucleotide code.

**Definition 3.** A trinucleotide code $X \subset A_4^3$ is circular if, for each $x_1, \ldots, x_n, x'_1, \ldots, x'_n \in X$, $n, m \geq 1$, $p \in A_4^3$, $s \in A_4^3$, the conditions $sx_1 \cdots x_n p = x'_1 \cdots x'_n s$ and $x_1 = ps$ imply $n = m$, $p = e$ (empty word) and $x_i = x'_i$ for $i = 1, \ldots, n$.

**Notation 1.** A trinucleotide circular code is noted $C$.

**Remark 4.** $A_4^3$ is not a trinucleotide circular code.

Let $l_1, l_2, \ldots, l_n, l_1$ be letters in $A_4$, $d_1, d_2, \ldots, d_{n-1}, d_n$ diletters in $A_4$, and $n$ an integer satisfying $n \geq 2$.

**Definition 5.** 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 (Letter Diletter Continued Necklace) for a subset $X \subset A_4^3$ if

\[
\begin{align*}
&l_1 d_1, l_2 d_2, \ldots, l_n d_n \in X, \\
&d_1 l_2, d_2 l_3, \ldots, d_{n-1} l_n, d_n l_{n+1} \in X.
\end{align*}
\]  

(2)

Only a few trinucleotide codes are circular. Two propositions based on the necklace concept allow to determine if a trinucleotide code is circular or not [2, 18].

**Proposition 6** (see [18]). Let $X$ be a trinucleotide code. The following conditions are equivalent:

(i) $X$ is a trinucleotide circular code;

(ii) $X$ has no nLDCN .

**Definition 7.** 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 (Letter Diletter Continued Closed Necklace) for a subset $X \subset A_4^3$ if

\[
\begin{align*}
&l_1 d_1, l_2 d_2, \ldots, l_n d_n \in X, \\
&d_1 l_2, d_2 l_3, \ldots, d_{n-1} l_n, d_n l_1 \in X.
\end{align*}
\]  

(3)

**Proposition 8** (see [2]). Let $X$ be a trinucleotide code. The following conditions are equivalent:

(i) $X$ is a trinucleotide circular code;

(ii) $X$ has no nLDCN .
Notation 3. The kth iterate of $\mathcal{P}$ is denoted by $\mathcal{P}^k$.

Remark 19. The trinucleotide codes $\mathcal{P}(X)$ and $\mathcal{P}^2(X)$ are the conjugated classes of the trinucleotide code $X$.

Definition 20. A trinucleotide circular code $X$ is self-complementary if, for each $x \in X$, $\mathcal{E}(x) \in X$.

Notation 4. A self-complementary trinucleotide circular code is noted SC.

Remark 21. A k-trinucleotide circular code for $k$ odd cannot be self-complementary.

Definition 22. A trinucleotide circular code $X$ is $C^3$ if $X$, $\mathcal{P}(X)$, and $\mathcal{P}^2(X)$ are trinucleotide circular codes.

Notation 5. A $C^3$ trinucleotide circular code $X$ is noted $C^3$.

Definition 23. A trinucleotide circular code $X$ is $C^3$ self-complementary maximum if $X$ is maximum, $X = \mathcal{E}(X)$ (self-complementary), and $\mathcal{P}(X)$ and $\mathcal{P}^2(X)$ are trinucleotide circular codes satisfying $\mathcal{E}(\mathcal{P}(X)) = \mathcal{P}(X)$.

Notation 6. A $C^3$ self-complementary maximum circular code is noted MSC$^3$.

The set $X_0$ of 20 trinucleotides identified in the gene populations of both eukaryotes and prokaryotes is a $C^3$ self-complementary maximum circular code MSC$^3$ [8]; that is, $X_0$ is maximum, $X_0 = \mathcal{E}(X_0)$, $\mathcal{P}(X_0) = X_1$, and $\mathcal{P}^2(X_0) = X_2$ are trinucleotide circular codes, and $\mathcal{E}(X_1) = X_2$.

We recall three classical evolution genetic maps: transition and transversions I and II, for example, [22] and extend their definitions to the positions of a trinucleotide.

Definition 24. The transition evolution genetic map $\mathcal{T}$: $\mathcal{A}_4^1 \rightarrow \mathcal{A}_4^1$ is defined by

$$\mathcal{T}(A) = G, \quad \mathcal{T}(C) = T, \quad \mathcal{T}(G) = A, \quad \mathcal{T}(T) = C.$$  \hspace{1cm} (9)

Definition 25. The transition map $\mathcal{T}$ on a letter $l$ can be applied in different positions of a trinucleotide $x = l_1l_2l_3$: $\mathcal{T}^i$, $i \in \{1, 2, 3\}$, is the transition on the position $i$ of $x$, $\mathcal{T}^i_j$, $i, j \in \{1, 2, 3\}$ with $i < j$, is the transition on the two positions $i$ and $j$ of $x$, and $\mathcal{T}^{1,2,3}$ is the transition on the three positions of $x$.

Example 26. $\mathcal{T}^1(AGC) = GCG$, $\mathcal{T}^2(ACG) = ATG$, $\mathcal{T}^3(ACG) = ACA$, $\mathcal{T}^{1,2}(ACG) = GTG$, $\mathcal{T}^{1,3}(ACG) = GCA$, $\mathcal{T}^{2,3}(ACG) = ATG$, and $\mathcal{T}^{1,2,3}(ACG) = GTA$.

Definition 27. The transition maps $\mathcal{T}^1$, $\mathcal{T}^i_j$, $\mathcal{T}^{1,2,3}$ on a trinucleotide $x$ are also extended to a trinucleotide code $X$, in a similar way to the genetic maps $\mathcal{E}$ and $\mathcal{P}$.

Definition 28. The transversion I evolution genetic map $\mathcal{V}^I_1$: $\mathcal{A}_4^1 \rightarrow \mathcal{A}_4^1$ is defined by

$$\mathcal{V}^I_1(A) = T, \quad \mathcal{V}^I_1(C) = G, \quad \mathcal{V}^I_1(G) = C, \quad \mathcal{V}^I_1(T) = A.$$  \hspace{1cm} (10)

Definition 29. The transversion I map $\mathcal{V}^I_1$ on a letter $l$ can also be applied in different positions of a trinucleotide $x = l_1l_2l_3$: $\mathcal{V}^i_1$, $i \in \{1, 2, 3\}$, is the transversion I on the position $i$ of $x$, $\mathcal{V}^i_j$, $i, j \in \{1, 2, 3\}$ with $i < j$, is the transversion I on the two positions $i$ and $j$ of $x$, and $\mathcal{V}^{1,2,3}_I$ is the transversion I on the three positions of $x$.

Example 30. $\mathcal{V}^I_1(ACG) = TCG$, $\mathcal{V}^I_1(ACG) = AGG$, $\mathcal{V}^I_1(ACG) = ACC$, $\mathcal{V}^{1,2}_I(ACG) = TGG$, $\mathcal{V}^{1,3}_I(ACG) = TCC$, $\mathcal{V}^{2,3}_I(ACG) = AGC$, and $\mathcal{V}^{1,2,3}_I(ACG) = TGC$.

Definition 31. The transversion I maps $\mathcal{V}^i_1$, $\mathcal{V}^i_j$, $\mathcal{V}^{1,2,3}_I$ on a trinucleotide $x$ are also extended to a trinucleotide code $X$, in a similar way to the genetic maps $\mathcal{E}$ and $\mathcal{P}$.

Definition 32. The transversion II evolution genetic map $\mathcal{V}^{II}_1$: $\mathcal{A}_4^1 \rightarrow \mathcal{A}_4^1$ is defined by

$$\mathcal{V}^{II}_1(A) = C, \quad \mathcal{V}^{II}_1(C) = A, \quad \mathcal{V}^{II}_1(G) = T, \quad \mathcal{V}^{II}_1(T) = G.$$  \hspace{1cm} (11)

Definition 33. The transversion II map $\mathcal{V}^{II}_1$ on a letter $l$ can also be applied in different positions of a trinucleotide $x = l_1l_2l_3$: $\mathcal{V}^{ii}_1$, $i \in \{1, 2, 3\}$, is the transversion II on the position $i$ of $x$, $\mathcal{V}^{ij}_1$, $i, j \in \{1, 2, 3\}$ with $i < j$, is the transversion II on the two positions $i$ and $j$ of $x$, and $\mathcal{V}^{1,2,3}_II$ is the transversion II on the three positions of $x$.

Example 34. $\mathcal{V}^{II}_1(ACG) = CCG$, $\mathcal{V}^{II}_1(ACG) = AAG$, $\mathcal{V}^{II}_1(ACG) = ACT$, $\mathcal{V}^{1,2}_II(ACG) = CAG$, $\mathcal{V}^{1,3}_II(ACG) = CCT$, $\mathcal{V}^{2,3}_II(ACG) = AAT$, and $\mathcal{V}^{1,2,3}_II(ACG) = CAT$.

Definition 35. The transversion II maps $\mathcal{V}^{ii}_1$, $\mathcal{V}^{ij}_1$, $\mathcal{V}^{1,2,3}_II$ on a trinucleotide $x$ are also extended to a trinucleotide code $X$, in a similar way to the genetic maps $\mathcal{E}$ and $\mathcal{P}$.

Definition 36. The evolution genetic maps in $l$ trinucleotides of a trinucleotide circular code are defined by $\mathcal{F}(l)$ for transition, $\mathcal{F}^i(l)$ for transversion I, and $\mathcal{F}^{ii}(l)$ for transversion II.

3. Results

An evolution genetic map, that is, $\mathcal{F}(l)$, $\mathcal{F}^i(l)$, and $\mathcal{F}^{ii}(l)$, in $l$ trinucleotides of the common trinucleotide circular code $X_0$ leads to $S(l) = \binom{l}{2}$ trinucleotide codes which are potentially circular. Table 1 gives these numbers $S(l)$.

Based on Proposition 6 allowing to test if a trinucleotide code is circular or not (algorithm not detailed, see, e.g., [2]), computer analyses of a great number of trinucleotide codes allow to identify here new properties with the common trinucleotide circular code $X_0$ observed in genes under evolution by transition and transversion.
The transition $\mathcal{T}^i$ generates a maximum number of trinucleotide circular codes $C$ for

$$\max \{ c(\mathcal{T}^i(l)), i = 1, 2, 3, l = 1, \ldots, 20 \} = c(\mathcal{T}^1(7)) = c(\mathcal{T}^3(7)) = 1436$$

and a maximum number of $C^3$ self-complementary maximum circular codes MSC$^3$ for

$$\max \{ \text{msc}^3(\mathcal{T}^i(l)), i = 1, 2, 3, l = 1, \ldots, 20 \} = \text{msc}^3(\mathcal{T}^1(6)) = \text{msc}^3(\mathcal{T}^3(6)) = 20.$$
Table 3: Transition map $\mathcal{S}^{i,j}(l)$ in $l$ trinucleotides of the common trinucleotide circular code $X_c$. Number $c(\mathcal{S}^{i,j}(l))$ of circular codes $C$, number $mc(\mathcal{S}^{i,j}(l))$ of maximum circular codes MC, number $sc(\mathcal{S}^{i,j}(l))$ of self-complementary circular codes SC, number $c^3(\mathcal{S}^{i,j}(l))$ of circular codes $C^3$, and number $msc^3(\mathcal{S}^{i,j}(l))$ of $C^3$' self-complementary maximum circular codes MSC$^3$.

<table>
<thead>
<tr>
<th>$l$</th>
<th>$\mathcal{S}^{1,1}(l)$ or $\mathcal{S}^{3,1}(l)$</th>
<th>$\mathcal{S}^{1,1}(l)$</th>
<th>$\mathcal{S}^{3,1}(l)$</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$c$</td>
<td>$mc$</td>
<td>$sc$</td>
</tr>
<tr>
<td>1</td>
<td>9</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>2</td>
<td>37</td>
<td>4</td>
<td>6</td>
</tr>
<tr>
<td>3</td>
<td>91</td>
<td>4</td>
<td>0</td>
</tr>
<tr>
<td>4</td>
<td>148</td>
<td>6</td>
<td>11</td>
</tr>
<tr>
<td>5</td>
<td>166</td>
<td>6</td>
<td>0</td>
</tr>
<tr>
<td>6</td>
<td>129</td>
<td>4</td>
<td>6</td>
</tr>
<tr>
<td>7</td>
<td>67</td>
<td>4</td>
<td>0</td>
</tr>
<tr>
<td>8</td>
<td>21</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>9</td>
<td>4</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>10</td>
<td>3</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>11</td>
<td>4</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>12</td>
<td>3</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>13</td>
<td>4</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>14</td>
<td>1</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>15, ..., 20</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
</tbody>
</table>

Table 4: Transition map $\mathcal{S}^{1,2,3}(l)$ in $l$ trinucleotides of the common trinucleotide circular code $X_c$. Number $c(\mathcal{S}^{1,2,3}(l))$ of circular codes $C$, number $mc(\mathcal{S}^{1,2,3}(l))$ of maximum circular codes MC, number $sc(\mathcal{S}^{1,2,3}(l))$ of self-complementary circular codes SC, number $c^3(\mathcal{S}^{1,2,3}(l))$ of circular codes $C^3$, and number $msc^3(\mathcal{S}^{1,2,3}(l))$ of $C^3$' self-complementary maximum circular codes MSC$^3$.

<table>
<thead>
<tr>
<th>$l$</th>
<th>$\mathcal{S}^{1,2,3}(l)$</th>
<th>$\mathcal{S}^{1,2,3}(l)$</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$c$</td>
<td>$mc$</td>
</tr>
<tr>
<td>1</td>
<td>8</td>
<td>0</td>
</tr>
<tr>
<td>2</td>
<td>28</td>
<td>4</td>
</tr>
<tr>
<td>3</td>
<td>56</td>
<td>0</td>
</tr>
<tr>
<td>4</td>
<td>71</td>
<td>7</td>
</tr>
<tr>
<td>5</td>
<td>64</td>
<td>0</td>
</tr>
<tr>
<td>6</td>
<td>56</td>
<td>8</td>
</tr>
<tr>
<td>7</td>
<td>64</td>
<td>0</td>
</tr>
<tr>
<td>8</td>
<td>72</td>
<td>8</td>
</tr>
<tr>
<td>9</td>
<td>64</td>
<td>0</td>
</tr>
<tr>
<td>10</td>
<td>56</td>
<td>8</td>
</tr>
<tr>
<td>20</td>
<td>1</td>
<td>1</td>
</tr>
</tbody>
</table>

$\mathcal{S}^{1,2,3}(l)$ always generates trinucleotide circular codes. Indeed, for $l = 1, \ldots, 20$

$$c\left(\mathcal{S}^{1,2,3}(l)\right) > 0.$$  \hspace{1cm} (23)

The lists of trinucleotide circular codes $C$ associated with $c(\mathcal{S}^{1,2,3}(l))$ and $c(\mathcal{S}^{1,2,3}(20-l))$ are different for $l = 1, \ldots, 9$ (not shown). The transition $\mathcal{S}^{1,2,3}$ generates a maximum number of trinucleotide circular codes $C$ for

$$\max \left\{ c\left(\mathcal{S}^{1,2,3}(l)\right) \right\}, \ l = 1, \ldots, 20 \} = c\left(\mathcal{S}^{1,2,3}(8)\right) = c\left(\mathcal{S}^{1,2,3}(12)\right) = 72$$

and a maximum number of $C^3$ self-complementary maximum circular codes MSC$^3$ for

$$\max \left\{ msc^3\left(\mathcal{S}^{1,2,3}(l)\right) \right\}, \ l = 1, \ldots, 20 \}$$

The numbers $c^3(\mathcal{S}^{1,2}(l)) = c^3(\mathcal{S}^{2,3}(l))$ of circular codes $C^3$ have a particular growth function

$$c^3\left(\mathcal{S}^{1,2}(14)\right) = c^3\left(\mathcal{S}^{2,3}(14)\right) = 1,$$

$$c^3\left(\mathcal{S}^{1,2}(l)\right) = c^3\left(\mathcal{S}^{2,3}(l)\right) = 0 \quad \text{for } l = 10, \ldots, 13.$$  \hspace{1cm} (22)

3.1.3. Transition Map $\mathcal{S}^{1,2,3}$

Result 3 (Table 4). The transition $\mathcal{S}^{1,2,3}$ always generates trinucleotide circular codes. Indeed, for $l = 1, \ldots, 20$

$$c\left(\mathcal{S}^{1,2,3}(l)\right) > 0.$$  \hspace{1cm} (23)

3.2. Transversion I Map

3.2.1. Transversion I Map $\mathcal{V}_i^1$

Result 4 (Table 5). For $l = 1, \ldots, 20$

$$c\left(\mathcal{V}_i^1(l)\right) = c\left(\mathcal{V}_i^3(l)\right),$$

$$mc\left(\mathcal{V}_i^1(l)\right) = mc\left(\mathcal{V}_i^3(l)\right),$$

$$sc\left(\mathcal{V}_i^1(l)\right) = sc\left(\mathcal{V}_i^3(l)\right).$$
\[ c^3(\varphi_1^1(l)) = c^3(\varphi_1^3(l)), \]
\[ \text{msc}^3(\varphi_1^1(l)) = \text{msc}^3(\varphi_1^3(l)). \]

(26)

The lists of trinucleotide circular codes \( C \) associated with \( c(\varphi_1^1(l)) \) and \( c(\varphi_1^3(l)) \) are different for \( l = 1, \ldots, 9 \) (not shown). No trinucleotide code is circular after a certain number of transversions \( \varphi_1^i \) in the trinucleotides of the common trinucleotide circular code \( X_0 \). Precisely, for \( l = 10, \ldots, 20 \)
\[ c(\varphi_1^1(l)) = c(\varphi_1^3(l)) = 0 \] (27)

and for \( l = 18, 19, 20 \)
\[ c(\varphi_1^2(l)) = 0. \] (28)

The transversion \( \varphi_1^i \) generates a maximum number of trinucleotide circular codes \( C \) for
\[ \max\{c(\varphi_1^i(l)), i = 1, 2, 3, l = 1, \ldots, 20\} \]
\[ = c(\varphi_1^2(9)) = 24310 \] (29)

and a maximum number of \( c^3 \) self-complementary maximum circular codes \( \text{MSC}^3 \) for
\[ \max\{\text{msc}^3(\varphi_1^i(l)), i = 1, 2, 3, l = 1, \ldots, 20\} \]
\[ = \text{msc}^3(\varphi_1^2(8)) = 70. \] (30)

A remarkable code property only found with transversion \( \varphi_1^2 \) is, for \( l = 1, \ldots, 20 \),
\[ c(\varphi_1^2(l)) = c^3(\varphi_1^3(l)), \] (31)

and furthermore, after a detailed computer analysis, the lists of trinucleotide circular codes \( C \) and \( C^3 \) associated with \( c(\varphi_1^i(l)) \) and \( c^3(\varphi_1^i(l)) \), respectively, are identical for \( l = 1, \ldots, 17 \).

3.2.2. Transversion \( \varphi_1^i \) Map \( \varphi_1^{i,j} \)

Result 5 (Table 6). For \( l = 1, \ldots, 20 \)
\[ c(\varphi_1^{1,2}(l)) = c(\varphi_1^{2,3}(l)), \]
\[ \text{mc}(\varphi_1^{1,2}(l)) = \text{mc}(\varphi_1^{2,3}(l)), \]
\[ \text{sc}(\varphi_1^{1,2}(l)) = \text{sc}(\varphi_1^{2,3}(l)), \]
\[ c^3(\varphi_1^{1,2}(l)) = c^3(\varphi_1^{2,3}(l)), \]
\[ \text{msc}^3(\varphi_1^{1,2}(l)) = \text{msc}^3(\varphi_1^{2,3}(l)). \]

(32)

The lists of trinucleotide circular codes \( C \) associated with \( c(\varphi_1^{1,2}(l)) \) and \( c(\varphi_1^{2,3}(l)) \) are different for \( l = 1, \ldots, 12 \) (not shown). No trinucleotide code is circular after a certain number of transversions \( \varphi_1^{i,j} \) in the trinucleotides of the common trinucleotide circular code \( X_0 \). Precisely, for \( l = 13, \ldots, 20 \)
\[ c(\varphi_1^{1,2}(l)) = c(\varphi_1^{2,3}(l)) = 0 \] (33)

and for \( l = 19, 20 \)
\[ c(\varphi_1^{1,3}(l)) = 0. \] (34)

The transversion \( \varphi_1^{i,j} \) generates a maximum number of trinucleotide circular codes \( C \) for
\[ \max\{c(\varphi_1^{i,j}(l)), i, j = 1, 2, 3, i < j, l = 1, \ldots, 20\} \]
\[ = c(\varphi_1^{1,2}(6)) = c(\varphi_1^{2,3}(6)) = 630 \] (35)

and a maximum number of \( c^3 \) self-complementary maximum circular codes \( \text{MSC}^3 \) for
\[ \max\{\text{msc}^3(\varphi_1^{i,j}(l)), i, j = 1, 2, 3, i < j, l = 1, \ldots, 20\} \]
\[ = \text{msc}^3(\varphi_1^{1,2}(4)) = \text{msc}^3(\varphi_1^{2,3}(4)) = 6. \] (36)

The numbers \( \text{sc}(\varphi_1^{i,3}(l)) \) of self-complementary circular codes \( SC \) have a particular growth function
\[ \text{sc}(\varphi_1^{1,3}(l)) = 1 \quad \text{for} \quad l = 12, 14, 16, 18, \]
\[ \text{sc}(\varphi_1^{1,3}(l)) = 0 \quad \text{for} \quad l = 8, 10. \] (37)

The numbers \( c^3(\varphi_1^{i,3}(l)) \) of circular codes \( C^3 \) have a particular growth function
\[ c^3(\varphi_1^{1,3}(l)) = 1 \quad \text{for} \quad l = 16, 18, \]
\[ c^3(\varphi_1^{1,3}(17)) = 2, \] (38)
\[ c^3(\varphi_1^{1,3}(l)) = 0 \quad \text{for} \quad l = 7, \ldots, 15. \]

3.2.3. Transversion \( \varphi_1^i \) Map \( \varphi_1^{1,2,3} \)

Result 6 (Table 7). The transversion \( \varphi_1^{1,2,3} \) always generates trinucleotide circular codes. Indeed, for \( l = 1, \ldots, 20 \)
\[ c(\varphi_1^{1,2,3}(l)) > 0. \] (39)

The lists of trinucleotide circular codes \( C \) associated with \( c(\varphi_1^{1,2,3}(l)) \) and \( c(\varphi_1^{1,2,3}(20-l)) \) are different for \( l = 1, \ldots, 9 \) (not shown). The transversion \( \varphi_1^{1,2,3} \) generates a maximum number of trinucleotide circular codes \( C \) for
\[ \max\{c(\varphi_1^{1,2,3}(l)), l = 1, \ldots, 20\} = c(\varphi_1^{1,2,3}(10)) = 66 \] (40)

and a maximum number of \( c^3 \) self-complementary maximum circular codes \( \text{MSC}^3 \) for
\[ \max\{\text{msc}^3(\varphi_1^{1,2,3}(l)), l = 1, \ldots, 20\} \]
\[ = \text{msc}^3(\varphi_1^{1,2,3}(4)) = \text{msc}^3(\varphi_1^{1,2,3}(16)) = 9. \] (41)
Table 5: Transversion I map $Y^{(l)}_{V,1}$ in $l$ trinucleotides of the common trinucleotide circular code $X$. Number $c(Y^{(l)}_{V,1})$ of circular codes $C$, number $mc(Y^{(l)}_{V,1})$ of maximum circular codes $MC$, number $sc(Y^{(l)}_{V,1})$ of self-complementary circular codes $SC$, number $c^3(Y^{(l)}_{V,1})$ of circular codes $C^3$, and number $msc^3(Y^{(l)}_{V,1})$ of $C^3$ self-complementary maximum circular codes $MSC^3$.

<table>
<thead>
<tr>
<th>$l$</th>
<th>$c$</th>
<th>$mc$</th>
<th>$sc$</th>
<th>$c^3$</th>
<th>$msc^3$</th>
<th>$c$</th>
<th>$mc$</th>
<th>$sc$</th>
<th>$c^3$</th>
<th>$msc^3$</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>8</td>
<td>0</td>
<td>0</td>
<td>8</td>
<td>0</td>
<td>16</td>
<td>0</td>
<td>0</td>
<td>16</td>
<td>0</td>
</tr>
<tr>
<td>2</td>
<td>31</td>
<td>4</td>
<td>5</td>
<td>30</td>
<td>4</td>
<td>122</td>
<td>8</td>
<td>12</td>
<td>122</td>
<td>8</td>
</tr>
<tr>
<td>3</td>
<td>76</td>
<td>0</td>
<td>0</td>
<td>71</td>
<td>0</td>
<td>590</td>
<td>0</td>
<td>0</td>
<td>590</td>
<td>0</td>
</tr>
<tr>
<td>4</td>
<td>130</td>
<td>6</td>
<td>8</td>
<td>114</td>
<td>6</td>
<td>2030</td>
<td>28</td>
<td>56</td>
<td>2030</td>
<td>28</td>
</tr>
<tr>
<td>5</td>
<td>160</td>
<td>0</td>
<td>0</td>
<td>125</td>
<td>0</td>
<td>5278</td>
<td>0</td>
<td>0</td>
<td>5278</td>
<td>0</td>
</tr>
<tr>
<td>6</td>
<td>142</td>
<td>4</td>
<td>5</td>
<td>92</td>
<td>4</td>
<td>10738</td>
<td>56</td>
<td>132</td>
<td>10738</td>
<td>56</td>
</tr>
<tr>
<td>7</td>
<td>87</td>
<td>0</td>
<td>0</td>
<td>42</td>
<td>0</td>
<td>17446</td>
<td>0</td>
<td>0</td>
<td>17446</td>
<td>0</td>
</tr>
<tr>
<td>8</td>
<td>32</td>
<td>1</td>
<td>1</td>
<td>10</td>
<td>1</td>
<td>22880</td>
<td>70</td>
<td>174</td>
<td>22880</td>
<td>70</td>
</tr>
<tr>
<td>9</td>
<td>5</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>24310</td>
<td>0</td>
<td>0</td>
<td>24310</td>
<td>0</td>
</tr>
<tr>
<td>10</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>20878</td>
<td>56</td>
<td>132</td>
<td>20878</td>
<td>56</td>
</tr>
<tr>
<td>11</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>14378</td>
<td>0</td>
<td>0</td>
<td>14378</td>
<td>0</td>
</tr>
<tr>
<td>12</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>7826</td>
<td>28</td>
<td>56</td>
<td>7826</td>
<td>28</td>
</tr>
<tr>
<td>13</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>3290</td>
<td>0</td>
<td>0</td>
<td>3290</td>
<td>0</td>
</tr>
<tr>
<td>14</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1030</td>
<td>8</td>
<td>12</td>
<td>1030</td>
<td>8</td>
</tr>
<tr>
<td>15</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>226</td>
<td>0</td>
<td>0</td>
<td>226</td>
<td>0</td>
</tr>
<tr>
<td>16</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>31</td>
<td>1</td>
<td>1</td>
<td>31</td>
<td>1</td>
</tr>
<tr>
<td>17</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>2</td>
<td>0</td>
<td>0</td>
<td>2</td>
<td>0</td>
</tr>
<tr>
<td>18, ..., 20</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
</tbody>
</table>

Table 6: Transversion I map $Z^{(l)}_{V,1}$ in $l$ trinucleotides of the common trinucleotide circular code $X$. Number $c(Z^{(l)}_{V,1})$ of circular codes $C$, number $mc(Z^{(l)}_{V,1})$ of maximum circular codes $MC$, number $sc(Z^{(l)}_{V,1})$ of self-complementary circular codes $SC$, number $c^3(Z^{(l)}_{V,1})$ of circular codes $C^3$, and number $msc^3(Z^{(l)}_{V,1})$ of $C^3$ self-complementary maximum circular codes $MSC^3$.

<table>
<thead>
<tr>
<th>$l$</th>
<th>$c$</th>
<th>$mc$</th>
<th>$sc$</th>
<th>$c^3$</th>
<th>$msc^3$</th>
<th>$c$</th>
<th>$mc$</th>
<th>$sc$</th>
<th>$c^3$</th>
<th>$msc^3$</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>10</td>
<td>2</td>
<td>0</td>
<td>9</td>
<td>0</td>
<td>6</td>
<td>2</td>
<td>0</td>
<td>6</td>
<td>0</td>
</tr>
<tr>
<td>2</td>
<td>48</td>
<td>5</td>
<td>5</td>
<td>38</td>
<td>4</td>
<td>15</td>
<td>5</td>
<td>3</td>
<td>15</td>
<td>5</td>
</tr>
<tr>
<td>3</td>
<td>147</td>
<td>8</td>
<td>0</td>
<td>102</td>
<td>0</td>
<td>20</td>
<td>4</td>
<td>0</td>
<td>20</td>
<td>0</td>
</tr>
<tr>
<td>4</td>
<td>319</td>
<td>10</td>
<td>8</td>
<td>191</td>
<td>6</td>
<td>15</td>
<td>3</td>
<td>5</td>
<td>15</td>
<td>3</td>
</tr>
<tr>
<td>5</td>
<td>514</td>
<td>12</td>
<td>0</td>
<td>265</td>
<td>0</td>
<td>8</td>
<td>4</td>
<td>0</td>
<td>8</td>
<td>4</td>
</tr>
<tr>
<td>6</td>
<td>630</td>
<td>10</td>
<td>5</td>
<td>282</td>
<td>4</td>
<td>11</td>
<td>3</td>
<td>1</td>
<td>11</td>
<td>1</td>
</tr>
<tr>
<td>7</td>
<td>595</td>
<td>8</td>
<td>0</td>
<td>240</td>
<td>0</td>
<td>22</td>
<td>4</td>
<td>0</td>
<td>22</td>
<td>0</td>
</tr>
<tr>
<td>8</td>
<td>435</td>
<td>5</td>
<td>1</td>
<td>162</td>
<td>1</td>
<td>32</td>
<td>4</td>
<td>0</td>
<td>32</td>
<td>0</td>
</tr>
<tr>
<td>9</td>
<td>245</td>
<td>2</td>
<td>0</td>
<td>88</td>
<td>0</td>
<td>38</td>
<td>2</td>
<td>0</td>
<td>38</td>
<td>0</td>
</tr>
<tr>
<td>10</td>
<td>103</td>
<td>1</td>
<td>0</td>
<td>34</td>
<td>0</td>
<td>38</td>
<td>2</td>
<td>0</td>
<td>38</td>
<td>0</td>
</tr>
<tr>
<td>11</td>
<td>29</td>
<td>0</td>
<td>0</td>
<td>8</td>
<td>0</td>
<td>34</td>
<td>0</td>
<td>0</td>
<td>34</td>
<td>0</td>
</tr>
<tr>
<td>12</td>
<td>4</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>39</td>
<td>0</td>
<td>1</td>
<td>39</td>
<td>0</td>
</tr>
<tr>
<td>13</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>56</td>
<td>0</td>
<td>0</td>
<td>56</td>
<td>0</td>
</tr>
<tr>
<td>14</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>65</td>
<td>0</td>
<td>1</td>
<td>65</td>
<td>0</td>
</tr>
<tr>
<td>15</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>52</td>
<td>0</td>
<td>0</td>
<td>52</td>
<td>0</td>
</tr>
<tr>
<td>16</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>27</td>
<td>0</td>
<td>1</td>
<td>27</td>
<td>0</td>
</tr>
<tr>
<td>17</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>8</td>
<td>0</td>
<td>2</td>
<td>8</td>
<td>0</td>
</tr>
<tr>
<td>18</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>19, 20</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
</tbody>
</table>

{18, ..., 20}
The lists of trinucleotide circular codes \( C \) associated with \( c(\mathcal{V}^3_{\text{II}}(l)) \) and \( c(\mathcal{V}^3_{\text{II}}(l)) \) are different for \( l = 1, \ldots, 8 \) (not shown). No trinucleotide code is circular after a certain number of transversions \( \mathcal{V}^3_{\text{II}} \) in the trinucleotides of the common trinucleotide circular code \( X_0 \). Precisely, for \( l = 9, \ldots, 20 \)

\[
c(\mathcal{V}^3_{\text{II}}(l)) = c(\mathcal{V}^3_{\text{II}}(l)) = 0 \tag{43}
\]
and for \( l = 12, \ldots, 20 \)

\[
c(\mathcal{V}^2_{\text{II}}(l)) = 0. \tag{44}
\]

The transversion \( \mathcal{V}^2_{\text{II}} \) generates a maximum number of trinucleotide circular codes \( C \) for

\[
\text{max} \{c(\mathcal{V}^3_{\text{II}}(l)), i = 1, 2, 3, l = 1, \ldots, 20\}
\]

\[
= c(\mathcal{V}^3_{\text{II}}(5)) = 176 \tag{45}
\]
and a maximum number of \( C_3 \) self-complementary maximum circular codes \( \text{MSC}_C \) for

\[
\text{max} \{\text{mc}^3(\mathcal{V}^3_{\text{II}}(l)), i = 1, 2, 3, l = 1, \ldots, 20\}
\]

\[
= \text{mc}^3(\mathcal{V}^3_{\text{II}}(4)) = 6. \tag{46}
\]
The transversion II \( Y_{II}^{1,1} \) generates a maximum number of trinucleotide circular codes \( C \) for
\[
\max \{ c \left( Y_{II}^{1,1} (l) \right), i, j = 1, 2, 3, i < j , l = 1, \ldots , 20 \} = c \left( Y_{II}^{1,1} (6) \right) = 662
\]
and a maximum number of \( C^3 \) self-complementary maximum circular codes \( MSC^3 \) for
\[
\max \{ msc^3 \left( Y_{II}^{1,1} (l) \right), i, j = 1, 2, 3, i < j , l = 1, \ldots , 20 \} = msc^3 \left( Y_{II}^{1,1} (4) \right) = 6.
\]

The numbers \( c(Y_{II}^{1,2,3}(l)) \) or \( c(Y_{II}^{2,3,1}(l)) \) of circular codes \( C \) have a particular growth function
\[
c \left( Y_{II}^{1,2} (l) \right) = c \left( Y_{II}^{2,3} (l) \right) = 1 \quad \text{for} \quad l = 13, 14,
c \left( Y_{II}^{1,3} (l) \right) = c \left( Y_{II}^{2,2} (l) \right) = 0 \quad \text{for} \quad l = 6, \ldots , 12.
\]

3.3.3. Transversion II Map \( Y_{II}^{1,2,3} \)

Proposition 37. For \( l = 1, \ldots , 19 \)
\[
c \left( Y_{II}^{1,2,3} (l) \right) = 0
\]
and obviously, by letter invariance, \( c(Y_{II}^{1,2,3}(20)) = 1 \) as in Tables 4 and 7.

Proof. The common trinucleotide circular code \( X_0 \) can be partitioned according to the maps \( Y_{II}^{1,2,3}, \mathcal{P}, \) and \( \mathcal{Q}^2 \) as shown in Table 10.

Let a partition \( P_i = \{ x, x' \}, i\in\{1, \ldots , 10 \}, \) composed of two trinucleotides \( x, x' \in X_0. \) For \( l = 1, \) any transversion II of a trinucleotide \( x \in P_i \) generates a trinucleotide \( y \) which is a permuted trinucleotide of the other trinucleotide \( x' \in P_i. \) So, any transversion II of a trinucleotide \( x \in X_0 \) leads to a trinucleotide which is not circular. For \( 2 \leq l \leq 19, \) the proof needs a computer analysis of the necklace for the nontrivial cases when two transversions II occur with two trinucleotides in the same partitions.

Remark 38. Very surprisingly, for the three maps of transition, transversions I and II, \( Y_{I}^{1}(l) \), \( Y_{II}^{1}(l) \), and \( Y_{II}^{1}(l), \) \( i \in \{1, 2, 3 \}, \) \( Y_{II}^{2,3}(l), \) \( Y_{II}^{1,2}(l), \) \( Y_{II}^{1,3}(l), \) \( Y_{II}^{2,1}(l), \) \( Y_{II}^{2,3}(l), \) \( Y_{II}^{2,1}(l), \) \( Y_{II}^{3,1}(l), \) \( Y_{II}^{2,3}(l), \) \( Y_{II}^{2,1}(l), \) \( i, j \in \{1, 2, 3 \} \) with \( i < j \) and \( Y_{II}^{1,2,3}(l) \) (not for \( Y_{II}^{1,2,2}(l) \) and \( Y_{II}^{1,2,3}(l) \)), the numbers \( mc^3 \) of self-complementary maximum circular codes \( MSC^3 \) for the first even values of \( l \) follow a series of binomial coefficients. For \( Y_{II}^{1}(l), Y_{II}^{2}(l), \) and \( Y_{II}^{3}(l), \) \( i \in \{1, 2, 3 \}, Y_{II}^{1,2}(l), \) \( Y_{II}^{3,1}(l), \) \( Y_{II}^{2,1}(l), \) \( Y_{II}^{2,3}(l), \) and \( Y_{II}^{1,3}(l), \) \( i, j \in \{1, 2, 3 \} \) with \( i < j \), the numbers \( mc \) of maximum circular codes MC for the first even values of \( l \) follow a series of binomial coefficients. For \( Y_{II}^{1,2,3}(l), \) the numbers \( c^3 \) of circular codes \( C^3 \) for the values \( l \) and \( 20 - l \) with \( l = 1, \ldots , 8 \) follow a series of binomial coefficients. These binomial properties with some numbers of circular codes for the three maps of transition, transversions I and II have no combinatorial explanation so far.

4. Conclusion

A comprehensive computer analysis of transition and transversions I and II in the \( C^3 \) self-complementary maximum circular code \( X_0 \) shows some new results; in particular (i) transversion I \( Y_{II}^{2}(l) \) on the 2nd position of any subset of trinucleotides of \( X_0 \) generates trinucleotide circular codes.
which are always $C^3$ and (ii) transversion II $\gamma^{1,2,3}_{II}$ on the three positions of any subset of trinucleotides of $X_0$ yields no trinucleotide circular codes. In addition to the classical self-complementary (Definition 20) partition of $X_0$ known since 1996, a new partition of $X_0$ based on the transversion II map $\gamma^{1,2,3}_{II}$ (Definition 33) and the circular permutation maps $\sigma$ and $\sigma^2$ (Definition 18) is also identified here. These results here extend our theory of circular code in genes to its evolution under transition and transversion.

References


