BioSystems 118 (2014) 39–50

Contents lists available at ScienceDirect

BioSystems

journal homepage: www.elsevier.com/locate/biosystems

Bijective transformation circular codes and nucleotide exchanging RNA transcription

Christian J. Michel[a,*], Hervé Seligmann[b,c,1]

[a] Equipe de Bioinformatique Théorique, ICube, Université de Strasbourg, CNRS, 300 Boulevard Sébastien Brant, 67400 Illkirch, France
[b] National Natural History Museum Collections, The Hebrew University of Jerusalem, 91904 Jerusalem, Israel
[c] Department of Life Sciences, Ben Gurion University, 84105 Beer Sheva, Israel

ARTICLE INFO

Article history:
Received 13 September 2013
Received in revised form 6 February 2014
Accepted 12 February 2014
Available online 21 February 2014

Keywords:
Circular code
Bijective transformation
Reading frame retrieval
RNA transcription
Nucleotide exchanging

ABSTRACT

The C^3 self-complementary circular code X identified in genes of prokaryotes and eukaryotes is a set of 20 trinucleotides enabling reading frame retrieval and maintenance, i.e. a framing code (Arquès and Michel, 1996; Michel, 2012, 2013). Some mitochondrial RNAs correspond to DNA sequences when RNA transcription systematically exchanges between nucleotides (Seligmann, 2013a,b). We study here the 23 bijective transformation codes $\Pi(X)$ of X which may code nucleotide exchanging RNA transcription as suggested by this mitochondrial observation. The 23 bijective transformation codes $\Pi(X)$ are C^3 trinucleotide circular codes, seven of them are also self-complementary. Furthermore, several correlations are observed between the Reading Frame Retrieval (RFR) probability of bijective transformation codes $\Pi(X)$ and the different biological properties of $\Pi(X)$ related to their numbers of RNAs in GenBank's EST database, their polymerization rate, their number of amino acids and the chirality of amino acids they code. Results suggest that the circular code X with the functions of reading frame retrieval and maintenance in regular RNA transcription, may also have, through its bijective transformation codes $\Pi(X)$, the same functions in nucleotide exchanging RNA transcription. Associations with properties such as amino acid chirality suggest that the RFR of X and its bijective transformations molded the origins of the genetic code's machinery.

1. Introduction

The genetic code is a set of 64 trinucleotides {*AAA*,. . .,*TTT*} (called codons) coding the 20 amino acids constituting the proteins. Three trinucleotides {*ATG*,*GTG*,*TTG*} (called start codons) have two functions: they code an amino acid and they are also initiation signals for protein synthesis. The classical start codon is *ATG* coding *Met*. Three trinucleotides {*TAA*,*TAG*,*TGA*} (called stop codons) which do not code for an amino acid, are termination signals for protein synthesis. As the trinucleotides are the DNA words coding the amino acids, genes are DNA sequences which are read modulo 3 letters among the three possible frames. Classically, only one frame, called reading frame, which begins with a start codon and ends with a stop codon, codes the corresponding protein sequence according to the genetic code. However, ribosomal slippage uses a non-reading frame (frameshift translation) which is terminated early by stop codons (Seligmann and Pollock, 2004; Itzkovitz and Alon, 2007; Seligmann, 2007, 2010). Start and stop codons are not the only punctuation signals in genes. Indeed, a set X of 20 trinucleotides identified in prokaryotic and eukaryotic genes is a circular code which both codes amino acids and maintains the reading frame (Arquès and Michel, 1996; Michel, 2012, 2013):

$$X = \{AAC, AAT, ACC, ATC, ATT, CAG, CTC, CTG, GAA, GAC, GAG, GAT, GCC, GGC, GGT, GTA, GTC, GTT, TAC, TTC\}. \quad (1)$$

These 20 trinucleotides X are overrepresented in reading frames of prokaryotic and eukaryotic genes, as compared to their frequencies in the two non-coding shifted frames (Arquès and Michel, 1996). Furthermore, the different trinucleotides of X vary in efficiency for frame maintenance (termed stability in Table 7 in Ahmed et al., 2010). We briefly recall a few definitions and properties of the common trinucleotide circular code X (1) which are involved here.

* Corresponding author. Tel.: +33 368854462.
E-mail addresses: c.michel@unistra.fr (C.J. Michel), varanuseremius@gmail.com (H. Seligmann).
[1] Current address: Unité de Recherche sur les Maladies Infectieuses et Tropicales Emergentes, Faculté de Médecine, URMITE CNRS-IRD 198 UMR 6236, Université de la Méditerranée, 13385 Marseille, France.

http://dx.doi.org/10.1016/j.biosystems.2014.02.002
0303-2647/

1.1. Definitions and few properties of the common trinucleotide circular code X

Notation 1. The letters (or nucleotides or bases) define the genetic alphabet $A_4 = \{A,C,G,T\}$. The set of non-empty words (words resp.) over A_4 is denoted by A_4^+ (A_4^* resp.). The set of the 16 words of length 2 (dinucleotides or diletters) on A_4 is denoted by $A_4^2 = \{AA, \ldots, TT\}$. The set of the 64 words of length 3 (trinucleotides or triletters) on A_4 is denoted by $A_4^3 = \{AAA, \ldots, TTT\}$. The set of 20 amino acids is denoted by $A_{20} = \{Ala, Arg, Asn, Asp, Cys, Gln, Glu, Gly, His, Ile, Leu, Lys, Met, Phe, Pro, Ser, Thr, Trp, Tyr, Val\}$ using the 3-letter convention. The (universal) genetic code is a surjective map $\mathcal{G} : A_4^3 \backslash \{TAA, TAG, TGA\} \to A_{20}$ giving the amino acid $aa = \mathcal{G}(l_1 l_2 l_3)$ coded by the codon $l_1 l_2 l_3$. The three stop codons $\{TAA, TAG, TGA\}$ do not code for an amino acid. Let $x_1 \cdots x_n$ be the concatenation of the words x_i for $i = 1,\ldots,n$, where the symbol "$\cdot$" is the concatenation of words. We use here the definitions in coding theory.

Definition 1. Code: A set Y of words in A_4^3 is a code if, for each $x_1, \ldots, x_n, y_1, \ldots, y_m \in Y$, $n, m \geq 1$, the condition $x_1 \cdots x_n = y_1 \cdots y_m$ implies $n = m$ and $x_i = y_i$ for $i = 1, \ldots, n$.

The set A_4^3 itself is a code. Consequently, its non-empty subsets are codes. Here, we call them trinucleotide codes.

Definition 2. Trinucleotide circular code: A trinucleotide code $Y \subset A_4^3$ is circular if, for each $x_1, \ldots, x_n, y_1, \ldots, y_m \in Y$, $n, m \geq 1$, $r \in A_4^*$, $s \in A_4^+$, the conditions $sx_2 \cdots x_n r = y_1 \cdots y_m$ and $x_1 = rs$ imply $n = m$, $r = \varepsilon$ (empty word) and $x_i = y_i$ for $i = 1, \ldots, n$.

Definition 3. Complementarity map $\mathcal{C}$: The complementarity map $\mathcal{C} : A_4^+ \to A_4^+$ is defined by $\mathcal{C}(A) = T, \mathcal{C}(C) = G, \mathcal{C}(G) = C, \mathcal{C}(T) = A$ and, according to the property of the complementary and antiparallel double helix (for DNA and RNA), by $\mathcal{C}(u \cdot v) = \mathcal{C}(v) \cdot \mathcal{C}(u)$ for all $u, v \in A_4^+$, e.g. $\mathcal{C}(ACG) = CGT$. The complementarity map $\mathcal{C}$ on a trinucleotide x is naturally extended to a trinucleotide code Y by applying the complementarity map $\mathcal{C}$ to all its trinucleotides: $\mathcal{C}(Y) = \{y | \forall y' \in Y, y = \mathcal{C}(y')\}$.

Definition 4. Circular permutation map $\mathcal{P}$: The circular permutation map $\mathcal{P} : A_4^3 \to A_4^3$ permutes circularly each trinucleotide $l_0 l_1 l_2$ as follows $\mathcal{P}(w_0) = l_1 l_2 l_0$, e.g. $\mathcal{P}(ACG) = CGA$. The k th iterate of $\mathcal{P}$ is denoted $\mathcal{P}^k$, e.g. $\mathcal{P}^2(ACG) = GAC$. The circular permutation map $\mathcal{P}$ on a trinucleotide x is naturally and similarly extended to a trinucleotide code Y by applying the circular permutation map $\mathcal{P}$ (or the k th iterate of $\mathcal{P}$) to all its trinucleotides: $\mathcal{P}(Y) = \{y | \forall y' \in Y, y = \mathcal{P}(y')\}$.

Definition 5. Self-complementary trinucleotide circular code: A trinucleotide circular code Y is self-complementary if, for each $y \in Y$, $\mathcal{C}(y) \in Y$.

Definition 6. Permuted trinucleotide code: A trinucleotide code $Y_1 = \mathcal{P}(Y)$ of a trinucleotide code Y is permuted if, for each $y \in Y$, $\mathcal{P}(y) \in \mathcal{P}(Y)$. The permuted trinucleotide code $Y_2 = \mathcal{P}^2(Y)$ is defined similarly.

Definition 7. C^3 trinucleotide circular code: A trinucleotide circular code Y is C^3 if the permuted trinucleotide codes $Y_1 = \mathcal{P}(Y)$ and $Y_2 = \mathcal{P}^2(Y)$ are circular codes.

Definition 8. C^3 self-complementary trinucleotide circular code: A trinucleotide circular code Y is C^3 self-complementary if Y, $Y_1 = \mathcal{P}(Y)$ and $Y_2 = \mathcal{P}^2(Y)$ are trinucleotide circular codes satisfying the following properties $Y = \mathcal{C}(Y)$ (self-complementary), $\mathcal{C}(Y_1) = Y_2$ and $\mathcal{C}(Y_2) = Y_1$ (Y_1 and Y_2 are complementary).

Result 1. (Arquès and Michel, 1996). The common trinucleotide set $X = X_0$ (1) coding the reading frames (frames 0) in eukaryotic and prokaryotic genes is a C^3 self-complementary trinucleotide circular code. The circular code $X_1 = \mathcal{P}(X)$ contains the 20 following trinucleotides

$$X_1 = \{AAG, ACA, ACG, ACT, AGC, AGG, ATA, ATG, CCA, CCG, GCG, GTG, TAG, TCA, TCC, TCG, TCT, TGC, TTA, TTG\} \quad (2)$$

and the circular code $X_2 = \mathcal{P}^2(X)$, the 20 following trinucleotides

$$X_2 = \{AGA, AGT, CAA, CAC, CAT, CCT, CGA, CGC, CGG, CGT, CTA, CCT, GCA, GCT, GGA, TAA, TAT, TGA, TGG, TGT\}. \quad (3)$$

Definition 9. Trinucleotide comma-free code: A trinucleotide code $Y \subset A_4^3$ is comma-free if, for each $y \in Y$ and $u, v \in A_4^*$ such that $uyv = x_1 \cdots x_n$ with $x_1, \ldots, x_n \in Y$, $n \geq 1$, it holds that $u, v \in Y^*$.

Result 2. (Michel, 2012). The subset $\tilde{X} = \{CAG, CTC, CTG, GAG\}$ of X (1) is a trinucleotide comma-free code and furthermore, C^3 self-complementary.

1.2. Circular code involved in frameshift and overlapping genes

Frameshift genes use two frames to code for a protein. Indeed, a programmed ribosomal frameshift occurs so that the second part of the protein coded downstream the frameshift site is translated in the +1 or +2 frame compared to the first part of the protein upstream the frameshift site which is translated in reading frame 0. There is a loss of the X circular code signal at the frameshift site (Ahmed et al., 2007). Furthermore, the X circular code signal downstream the frameshift site is shifted from the circular code signal upstream the frameshift site according to the type +1 or +2 of frameshift (Ahmed and Michel, 2011). The transition phase of the X circular code signal at the frameshift site is compatible with a frameshift solely induced by a frameshifting mutation, i.e. a nucleotide insertion or deletion.

Overlapping genes use more than one frame to code for more than one protein, typically two proteins. In mitochondria, overlapping genes can be expressed when RNA transcription systematically exchanges between nucleotides (Seligmann, 2012, 2013a,b,c). Precisely, in addition to the regular transcription of DNA to RNA, there are 23 different potential types of nucleotide exchanges which can be divided into two classes: nine symmetric exchanges of the type $a \leftrightarrow b$ (Seligmann, 2013a) and 14 asymmetric exchanges of the type $a \to b \to c \to d \to a$ (Seligmann, 2013b). About 100 RNA transcripts corresponding to regions of the human mitochondrial genome after systematic nucleotide exchange have been detected for seven symmetric nucleotide exchanges and four asymmetric nucleotide exchanges in GenBank's ESTs database (Seligmann, 2013a,b). As GenBank's EST database includes about 10,000 human transcripts from mitochondria, this observation suggests that the nucleotide exchanging transcripts represent about 1% of RNAs. Otherwise, the trinucleotides of the circular code X are underrepresented in the regular reading frames involved in predicted overlapping genes, as detected by comparing their frequencies in adjacent (upstream and downstream) regions of the main frame of the regular gene coded without nucleotide exchange (Seligmann, 2012, 2013a,b,c).

As the circular code X is associated to the reading frames in eukaryotic and prokaryotic genes, i.e. to the regular RNA transcription, the working hypothesis studied here is to relate the 23 bijective transformation codes of X with the 23 nucleotide exchanges which generate non-regular RNA transcription observed, for example, in mitochondrial overlapping genes. We will show that the 23 bijective transformation codes $\Pi(X)$ of X are C^3 trinucleotide circular codes, i.e. with the property of reading frame retrieval and maintenance, and furthermore, seven of them are in addition self-complementary. The trinucleotides of X which occur in the bijective transformation codes $\Pi(X)$ are analyzed for their probability to retrieve the reading frame. These "conserved"

trinucleotides of X in $\Pi(X)$ may also explain the paucity of X in mitochondrial overlapping genes (Seligmann, 2012, 2013a,b,c). Furthermore, the 23 bijective transformation codes $\Pi(X)$ are also analyzed according to their polymerization rate, their number of amino acids and the chirality of amino acids they code.

2. Method

2.1. The 23 bijective transformation codes of the C^3 self-complementary trinucleotide circular code X

We described the 23 bijective transformation codes $\Pi(X) = \{\pi_1(X), \ldots, \pi_{23}(X)\}$ of the C^3 self-complementary trinucleotide circular code $X = \pi_0(X)$ (Table 1). The notation of bijective transformations used here is based on (i) the transcript data identified from the human mitochondrial genome by Seligmann (2013a,b) in previous papers; and (ii) the biological function of the polymerase. These biological observations suggest that bijective transformations of RNA transcripts using only two bases are simpler than bijective transformations of three bases which are also simpler than bijective transformations of four bases. Another notation of bijective transformations of circular codes has recently been proposed by Fimmel et al. (2013, pp. 225–226) in a combinatorial work based on group theory.

2.1.1. Partition into symmetric and asymmetric bijective transformation codes

The 23 bijective transformation codes $\Pi(X)$ of X can be partitioned into nine symmetric bijective transformation codes $\Pi_S(X) = \{\pi_1(X), \ldots, \pi_9(X)\}$ and 14 asymmetric bijective transformation codes $\Pi_A(X) = \{\pi_{10}(X), \ldots, \pi_{23}(X)\}$ (Table 1). The number $N(n,p)$ of bijective transformation codes at p letters among n letters is (obviously) equal to

$$N(n,p) = \frac{n!}{(n-p)!p}.$$

Note: If $p = n$ then $N(n,n) = (n-1)!$.

The nine symmetric bijective transformation codes $\Pi_S(X)$ can again be partitioned into

(i) N(4,2)=6 symmetric bijective transformation codes $\Pi_{S,2}(X)$ at 2 letters

$$\Pi_{S,2}(X) = \{\pi_1(X) : (AC), \pi_2(X) : (AG), \pi_3(X) : (AT), \pi_4(X) : (CG), \pi_5(X) : (CT), \pi_6(X) : (GT)\}$$

where $\pi_i(X) : (l_1 l_2)$ is the ith bijective transformation in the lexicographical order of the letter $l_1 \in A_4$ into the letter $l_2 \in A_4$, $l_2 \neq l_1$, and reciprocally;

(ii) N(4,2)/2=3 symmetric bijective transformation codes $\Pi_{S,2,2}(X)$ of two disjoint bijective transformations at 2 letters

$$\Pi_{S,2,2}(X) = \{\pi_7(X) : (AC)(GT), \pi_8(X) : (AG)(CT), \pi_9(X) : (AT)(CG)\}$$

where $\pi_i(X) : (l_1 l_2)(l_3 l_4)$ is the ith bijective transformation in the lexicographical order of the letter $l_1 \in A_4$ into the letter $l_2 \in A_4$, $l_2 \neq l_1$, and reciprocally, and of the letter $l_3 \in A_4$, $l_3 \neq l_2 \neq l_1$, into the letter $l_4 \in A_4$, $l_4 \neq l_3 \neq l_2 \neq l_1$, and reciprocally.

The 14 asymmetric bijective transformation codes $\Pi_A(X)$ can also be partitioned into

(i) N(4,3)=8 asymmetric bijective transformation codes $\Pi_{A,3}(X)$ at 3 letters

$$\Pi_{A,3}(X) = \{\pi_{10}(X) : (ACG), \pi_{11}(X) : (ACT), \pi_{12}(X) : (AGC), \pi_{13}(X) : (AGT), \pi_{14}(X) : (ATC), \pi_{15}(X) : (ATG), \pi_{16}(X) : (CGT), \pi_{17}(X) : (CTG)\}$$

where $\pi_i(X) : (l_1 l_2 l_3)$ is the ith bijective transformation in the lexicographical order of the letter $l_1 \in A_4$ into the letter $l_2 \in A_4$, $l_2 \neq l_1$, the letter l_2 into the letter $l_3 \in A_4$, $l_3 \neq l_2 \neq l_1$, and the letter l_3 into the letter l_1;

(ii) N(4,4)=6 asymmetric bijective transformation codes $\Pi_{A,4}(X)$ at 4 letters

$$\Pi_{A,4}(X) = \{\pi_{18}(X) : (ACGT), \pi_{19}(X) : (ACTG), \pi_{20}(X) : (AGCT), \pi_{21}(X) : (AGTC), \pi_{22}(X) : (ATCG), \pi_{23}(X) : (ATGC)\}$$

where $\pi_i(X) : (l_1 l_2 l_3 l_4)$ is the ith bijective transformation in the lexicographical order of the letter $l_1 \in A_4$ into the letter $l_2 \in A_4$, $l_2 \neq l_1$, the letter l_2 into the letter $l_3 \in A_4$, $l_3 \neq l_2 \neq l_1$, the letter l_3 into the letter $l_4 \in A_4$, $l_4 \neq l_3 \neq l_2 \neq l_1$, and the letter l_4 into the letter l_1.

Note that the transformations at 1 ($X = \pi_0(X)$), 2, 3 and 4 letters are the transformations of order 1, 2, 3 and 4, respectively, according to the notation in Fimmel et al. (2013, pp. 225–226).

2.1.2. Partition into complementary and non-complementary bijective transformation codes

The 23 bijective transformation codes $\Pi(X)$ of X can also be partitioned into seven self-complementary bijective transformation codes $\Pi_C(X) = \{\pi_3(X), \pi_4(X), \pi_7(X), \pi_8(X), \pi_9(X), \pi_{19}(X), \pi_{21}(X)\}$ and 16 non self-complementary bijective transformation codes $\Pi_{\overline{C}}(X) = \Pi(X) \backslash \Pi_C(X)$ of X (Table 1).

2.2. Reading frame retrieval (RFR) probability of a bijective transformation code π(X)

A method was developed for measuring the Reading Frame Retrieval (RFR) probability (originally called stability) for sets of 20 trinucleotides, in particular for the C^3 self-complementary trinucleotide circular code X (Ahmed et al., 2010). The RFR probability of X is based on the mean value of RFR probabilities of 20 trinucleotides of X. This RFR method is extended here to the RFR probability for a bijective transformation code $\pi(X)$ of X.

Let $x_0 = l_0 l_1 l_2$, $l_i \in A_4$ with $0 \leq i \leq 2$, be a trinucleotide of X (in frame 0). The cardinality of X is $\mathrm{Card}(X) = 20$. By convention, the reading frame established by a start codon $\{ATG, GTG, TTG\}$ is the frame 0, and the frames 1 and 2 are the reading frame 0 shifted by 1 and 2 nucleotides in the $5' - 3'$ direction, respectively. Let the di-trinucleotide w be a concatenation of the two trinucleotides x_0 and $x'_0 = l'_0 l'_1 l'_2$ of X, i.e. $w = x_0 x'_0 \in X^2$. By assuming that the trinucleotides of X are equiprobable, then there are 400 possible di-trinucleotides $w \in X^2$ leading to an occurrence probability $\Pr(w) = 1/\mathrm{Card}(X)^2$. We denote by $x_0(w)$, $x_1(w)$ and $x_2(w)$ the trinucleotides $l_0 l_1 l_2$ in frame 0, $l_1 l_2 l'_0$ in frame 1 and $l_2 l'_0 l'_1$ in frame 2 of a di-trinucleotide $w \in X^2$, respectively. The concatenation of the two trinucleotides x_0 and x'_0 of X may yield a trinucleotide $x_f(w) \in X$ but in a frame $f \neq 0$. For example, the concatenation of the trinucleotides $x_0 = TAC \in X$ and $x'_0 = CTC \in X$, i.e. $w = TACCTC$, leads to the trinucleotide $x_1(w) = ACC \in X$ which thus occurs in frame 0 but also in frame 1 (see the trinucleotides of X in (1)).

In order to measure the RFR probability of each trinucleotide x of X, the frequency of x of X to occur not only in the reading frame 0 but also in the 2 shifted frames 1 and 2, is determined. Let the

Table 1

Probability Pr(*x*,*X*) (%) (Eq. (4)) of Reading Frame Retrieval (RFR) of each trinucleotide *x* of the C^3 self-complementary trinucleotide circular code $X = \pi_0(X)$ in the 23 bijective transformation codes $\Pi(X) = \{\pi_1(X), \ldots, \pi_{23}(X)\}$. The lines 1–2 give the RFR probability Pr(*x*,*X*) (%) of *x* of *X*. The subset $\tilde{X} = \{CAG, CTC, CTG, GAG\}$ of the four trinucleotides $\tilde{x}$ is a C^3 self-complementary trinucleotide comma-free code with a RFR probability $\Pr(\tilde{x}, X) = 1$. The lines 3–14 present the RFR probability Pr(*x*,*X*) (%) of *x* of *X* in the six symmetric bijective transformation codes $\Pi_{S,2}(X) = \{\pi_1(X), \pi_2(X), \pi_3(X), \pi_4(X), \pi_5(X), \pi_6(X)\}$ at 2 letters. The lines 15–20 give the RFR probability Pr(*x*,*X*) (%) of *x* of *X* in the three symmetric bijective transformation codes $\Pi_{S,2,2}(X) = \{\pi_7(X), \pi_8(X), \pi_9(X)\}$ of two disjoint transformations at 2 letters. The lines 21–36 present the RFR probability Pr(*x*,*X*) (%) of *x* of *X* in the eight asymmetric bijective transformation codes $\Pi_{A,3}(X) = \{\pi_{10}(X), \pi_{11}(X), \pi_{12}(X), \pi_{13}(X), \pi_{14}(X), \pi_{15}(X), \pi_{16}(X), \pi_{17}(X)\}$ at 3 letters. The lines 37–48 present the RFR probability Pr(*x*,*X*) (%) of *x* of *X* in the six asymmetric bijective transformations $\Pi_{A,4}(X) = \{\pi_{18}(X), \pi_{19}(X), \pi_{20}(X), \pi_{21}(X), \pi_{22}(X), \pi_{23}(X)\}$ at 4 letters. Underlined trinucleotides *x* in the codes $\Pi(X)$ belong to *X*. Bijective transformations $\{\pi_3(X), \pi_4(X), \pi_7(X), \pi_8(X), \pi_9(X), \pi_{19}(X), \pi_{21}(X)\}$ in bold are C^3 self-complementary trinucleotide circular codes. The next to last column gives the RFR probability $\Pr(\pi(X))$ (%) (Eq. (5)) of *X* and its 23 codes $\Pi(X)$. The last column presents the relative RFR probability $\mathrm{PrRFR}(\pi(X))$ (%) (Eq. (6)) of *X* and its 23 codes $\Pi(X)$.

																					$\Pr(\pi(X))$	$\mathrm{PrRFR}(\pi(X))$
$X = \pi_0(X)$	*AAC*	*AAT*	*ACC*	*ATC*	*ATT*	*CAG*	*CTC*	*CTG*	*GAA*	*GAC*	*GAG*	*GAT*	*GCC*	*GGC*	*GGT*	*GTA*	*GTC*	*GTT*	*TAC*	*TTC*		
Pr(*x*,*X*) (%)	80	76.9	69	76.9	76.9	100	100	100	76.9	87	100	76.9	87	87	69	71.4	87	80	71.4	76.9	82.5	100
$\pi_1(X)$: (*AC*)	*CCA*	*CCT*	*CAA*	*CTA*	*CTT*	*ACG*	*ATA*	*ATG*	*GCC*	*GCA*	*GCG*	*GCT*	*GAA*	*GGA*	*GGT*	*GTC*	*GTA*	*GTT*	*TCA*	*TTA*		
Pr(x, $\pi_1(X)$) (%)	0	0	0	0	0	0	0	0	87	0	0	0	76.9	0	69	87	71.4	80	0	0	23.6	28.6
$\pi_2(X)$: (*AG*)	*GGC*	*GGT*	*GCC*	*GTC*	*GTT*	*CGA*	*CTC*	*CTA*	*AGG*	*AGC*	*AGA*	*AGT*	*ACC*	*AAC*	*AAT*	*ATG*	*ATC*	*ATT*	*TGC*	*TTC*		
Pr(x, $\pi_2(X)$) (%)	87	69	87	87	80	0	100	0	0	0	0	0	69	80	76.9	0	76.9	76.9	0	76.9	48.3	58.6
$\boldsymbol{\pi_3}(X)$: (AT)	***TTC***	***TTA***	***TCC***	***TAC***	***TAA***	***CTG***	***CAC***	***CAG***	***GTT***	***GTC***	***GTG***	***GTA***	***GCC***	***GGC***	***GGA***	***GAT***	***GAC***	***GAA***	***ATC***	***AAC***		
Pr(x, $\pi_3(X)$) (%)	76.9	0	0	71.4	0	100	0	100	80	87	0	71.4	87	87	0	76.9	87	76.9	76.9	80	57.9	70.2
$\boldsymbol{\pi_4}(X)$: (CG)	***AAG***	***AAT***	***AGG***	***ATG***	***ATT***	***GAC***	***GTG***	***GTC***	***CAA***	***CAG***	***CAC***	***CAT***	***CGG***	***CCG***	***CCT***	***CTA***	***CTG***	***CTT***	***TAG***	***TTG***		
Pr(x, $\pi_4(X)$) (%)	0	76.9	0	0	76.9	87	0	87	0	100	0	0	0	0	0	0	100	0	0	0	26.4	32.0
$\pi_5(X)$: (*CT*)	*AAT*	*AAC*	*ATT*	*ACT*	*ACC*	*TAG*	*TCT*	*TCG*	*GAA*	*GAT*	*GAG*	*GAC*	*GTT*	*GGT*	*GGC*	*GCA*	*GCT*	*GCC*	*CAT*	*CCT*		
Pr(x, $\pi_5(X)$) (%)	76.9	80	76.9	0	69	0	0	0	76.9	76.9	100	87	80	69	87	0	0	87	0	0	48.3	58.6
$\pi_6(X)$: (*GT*)	*AAC*	*AAG*	*ACC*	*AGC*	*AGG*	*CAT*	*CGC*	*CGT*	*TAA*	*TAC*	*TAT*	*TAG*	*TCC*	*TTC*	*TTG*	*TGA*	*TGC*	*TGG*	*GAC*	*GGC*		
Pr(x, $\pi_6(X)$) (%)	80	0	69	0	0	0	0	0	0	71.4	0	0	0	76.9	0	0	0	0	87	87	23.6	28.6
$\boldsymbol{\pi_7}(X)$: (AC)(GT)	***CCA***	***CCG***	***CAA***	***CGA***	***CGG***	***ACT***	***AGA***	***AGT***	***TCC***	***TCA***	***TCT***	***TCG***	***TAA***	***TTA***	***TTG***	***TGC***	***TGA***	***TGG***	***GCA***	***GGA***		
Pr(x, $\pi_7(X)$) (%)	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
$\boldsymbol{\pi_8}(X)$: (AG)(CT)	***GGT***	***GGC***	***GTT***	***GCT***	***GCC***	***TGA***	***TCT***	***TCA***	***AGG***	***AGT***	***AGA***	***AGC***	***ATT***	***AAT***	***AAC***	***ACG***	***ACT***	***ACC***	***CGT***	***CCT***		
Pr(x, $\pi_8(X)$) (%)	69	87	80	0	87	0	0	0	0	0	0	0	76.9	76.9	80	0	0	69	0	0	31.3	37.9
$\boldsymbol{\pi_9}(X)$: (AT)(CG)	***TTG***	***TTA***	***TGG***	***TAG***	***TAA***	***GTC***	***GAG***	***GAC***	***CTT***	***CTG***	***CTC***	***CTA***	***CGG***	***CCG***	***CCA***	***CAT***	***CAG***	***CAA***	***ATG***	***AAG***		
Pr(x, $\pi_9(X)$) (%)	0	0	0	0	0	87	100	87	0	100	100	0	0	0	0	0	100	0	0	0	28.7	34.8
$\pi_{10}(X)$: (*ACG*)	*CCG*	*CCT*	*CGG*	*CTG*	*CTT*	*GCA*	*GTG*	*GTA*	*ACC*	*ACG*	*ACA*	*ACT*	*AGG*	*AAG*	*AAT*	*ATC*	*ATG*	*ATT*	*TCG*	*TTG*		
Pr(x, $\pi_{10}(X)$) (%)	0	0	0	100	0	0	0	71.4	69	0	0	0	0	0	76.9	76.9	0	76.9	0	0	23.6	28.5
$\pi_{11}(X)$: (*ACT*)	*CCT*	*CCA*	*CTT*	*CAT*	*CAA*	*TCG*	*TAT*	*TAG*	*GCC*	*GCT*	*GCG*	*GCA*	*GTT*	*GGT*	*GGA*	*GAC*	*GAT*	*GAA*	*ACT*	*AAT*		
Pr(x, $\pi_{11}(X)$) (%)	0	0	0	0	0	0	0	0	87	0	0	0	80	69	0	87	76.9	76.9	0	76.9	27.7	33.6
$\pi_{12}(X)$: (AGC)	*GGA*	*GGT*	*GAA*	*GTA*	*GTT*	*AGC*	*ATA*	*ATC*	*CGG*	*CGA*	*CGC*	*CGT*	*CAA*	*CCA*	*CCT*	*CTG*	*CTA*	*CTT*	*TGA*	*TTA*		
Pr(x, $\pi_{12}(X)$) (%)	0	69	76.9	71.4	80	0	0	76.9	0	0	0	0	0	0	0	100	0	0	0	0	23.7	28.7
$\pi_{13}(X)$: (*AGT*)	*GGC*	*GGA*	*GCC*	*GAC*	*GAA*	*CGT*	*CAC*	*CAT*	*TGG*	*TGC*	*TGT*	*TGA*	*TCC*	*TTC*	*TTA*	*TAG*	*TAC*	*TAA*	*AGC*	*AAC*		
Pr(x, $\pi_{13}(X)$) (%)	87	0	87	87	76.9	0	0	0	0	0	0	0	0	76.9	0	0	71.4	0	0	80	28.3	34.3
$\pi_{14}(X)$: (*ATC*)	*TTA*	*TTC*	*TAA*	*TCA*	*TCC*	*ATG*	*ACA*	*ACG*	*GTT*	*GTA*	*GTG*	*GTC*	*GAA*	*GGA*	*GGC*	*GCT*	*GCA*	*GCC*	*CTA*	*CCA*		
Pr(x, $\pi_{14}(X)$) (%)	0	76.9	0	0	0	0	0	0	80	71.4	0	87	76.9	0	87	0	0	87	0	0	28.3	34.3
$\pi_{15}(X)$: (*ATG*)	*TTC*	*TTG*	*TCC*	*TGC*	*TGG*	*CTA*	*CGC*	*CGA*	*ATT*	*ATC*	*ATA*	*ATG*	*ACC*	*AAC*	*AAG*	*AGT*	*AGC*	*AGG*	*GTC*	*GGC*		
Pr(x, $\pi_{15}(X)$) (%)	76.9	0	0	0	0	0	0	0	76.9	76.9	0	0	69	80	0	0	0	0	87	87	27.7	33.6
$\pi_{16}(X)$: (*CGT*)	*AAG*	*AAC*	*AGG*	*ACG*	*ACC*	*GAT*	*GCG*	*GCT*	*TAA*	*TAG*	*TAT*	*TAC*	*TGG*	*TTG*	*TTC*	*TCA*	*TCG*	*TCC*	*CAG*	*CCG*		
Pr(x, $\pi_{16}(X)$) (%)	0	80	0	0	69	76.9	0	0	0	0	0	71.4	0	0	76.9	0	0	0	100	0	23.7	28.7
$\pi_{17}(X)$: (*CTG*)	*AAT*	*AAG*	*ATT*	*AGT*	*AGG*	*TAC*	*TGT*	*TGC*	*CAA*	*CAT*	*CAC*	*CAG*	*CTT*	*CCT*	*CCG*	*CGA*	*CGT*	*CGG*	*GAT*	*GGT*		
Pr(x, $\pi_{17}(X)$) (%)	76.9	0	76.9	0	0	71.4	0	0	0	0	0	100	0	0	0	0	0	0	76.9	69	23.6	28.5
$\pi_{18}(X)$: (*ACGT*)	*CCG*	*CCA*	*CGG*	*CAG*	*CAA*	*GCT*	*GAG*	*GAT*	*TCC*	*TCG*	*TCT*	*TCA*	*TGG*	*TTG*	*TTA*	*TAC*	*TAG*	*TAA*	*ACG*	*AAG*		
Pr(x, $\pi_{18}(X)$) (%)	0	0	0	100	0	0	100	76.9	0	0	0	0	0	0	0	71.4	0	0	0	0	17.4	21.1
$\boldsymbol{\pi_{19}}(X)$: (ACTG)	***CCT***	***CCG***	***CTT***	***CGT***	***CGG***	***TCA***	***TGT***	***TGA***	***ACC***	***ACT***	***ACA***	***ACG***	***ATT***	***AAT***	***AAG***	***AGC***	***AGT***	***AGG***	***GCT***	***GGT***		
Pr(x, $\pi_{19}(X)$) (%)	0	0	0	0	0	0	0	0	69	0	0	0	76.9	76.9	0	0	0	0	0	69	14.6	17.7
$\pi_{20}(X)$: (*AGCT*)	*GGT*	*GGA*	*GTT*	*GAT*	*GAA*	*TGC*	*TAT*	*TAC*	*CGG*	*CGT*	*CGC*	*CGA*	*CTT*	*CCT*	*CCA*	*CAG*	*CAT*	*CAA*	*AGT*	*AAT*		
Pr(x, $\pi_{20}(X)$) (%)	69	0	80	76.9	76.9	0	0	71.4	0	0	0	0	0	0	0	100	0	0	0	76.9	27.6	33.4
$\boldsymbol{\pi_{21}}(X)$: (AGTC)	***GGA***	***GGC***	***GAA***	***GCA***	***GCC***	***AGT***	***ACA***	***ACT***	***TGG***	***TGA***	***TGT***	***TGC***	***TAA***	***TTA***	***TTC***	***TCG***	***TCA***	***TCC***	***CGA***	***CCA***		
Pr(x, $\pi_{21}(X)$) (%)	0	87	76.9	0	87	0	0	0	0	0	0	0	0	76.9	0	0	0	0	0	0	16.4	19.9
$\pi_{22}(X)$: (*ATCG*)	*TTG*	*TTC*	*TGG*	*TCG*	*TCC*	*GTA*	*GCG*	*GCA*	*ATT*	*ATG*	*ATA*	*ATC*	*AGG*	*AAG*	*AAC*	*ACT*	*ACG*	*ACC*	*CTG*	*CCG*		
Pr(x, $\pi_{22}(X)$) (%)	0	76.9	0	0	0	71.4	0	0	76.9	0	0	76.9	0	0	80	0	0	69	100	0	27.6	33.4
$\pi_{23}(X)$: (*ATGC*)	*TTA*	*TTG*	*TAA*	*TGA*	*TGG*	*ATC*	*AGA*	*AGC*	*CTT*	*CTA*	*CTC*	*CTG*	*CAA*	*CCA*	*CCG*	*CGT*	*CGA*	*CGG*	*GTA*	*GGA*		
Pr(x, $\pi_{23}(X)$) (%)	0	0	0	0	0	76.9	0	0	0	0	100	100	0	0	0	0	0	0	71.4	0	17.4	21.1

indicator function $\delta(x, X)$ be equal to 1 if the trinucleotide x belongs to X, 0 otherwise

$$\delta(x, X) = \begin{cases} 1 & \text{if } x \in X, \\ 0 & \text{otherwise.} \end{cases}$$

Let the indicator function $\delta(x, x_f(w), X)$ be equal to 1 if the trinucleotide $x_f(w)$ in the shifted frame f, $1 \le f \le 2$, of w is equal to the trinucleotide $x \in X$

$$\delta(x, x_f(w), X) = \begin{cases} 1 & \text{if } x = x_f(w) \text{ with } x \in X \text{ and } w \in X^2, \\ 0 & \text{otherwise.} \end{cases}$$

Then, the RFR probability $\Pr(x,X)$ of a trinucleotide $x \in X$ is equal to

$$\Pr(x, X) = \frac{\delta(x, X)}{1 + \frac{1}{\mathrm{Card}(X)} \sum_{1 \le f \le 2} \sum_{w \in X^2} \delta(x, x_f(w), X)}. \quad (4)$$

When the trinucleotide x does not occur in the shifted frames then its RFR probability $\Pr(x,X) = 1$.

The RFR probability $\Pr(x,X)$ of each trinucleotide x of the C^3 self-complementary trinucleotide circular code X is given in Table 1 (first two lines). The subset $\tilde{X} = \{CAG, CTC, CTG, GAG\}$ of the four trinucleotides $\tilde{x}$ is a C^3 self-complementary trinucleotide comma-free code with a RFR probability $\Pr(\tilde{x}, X) = 1$ for each trinucleotide $\tilde{x}$, in accordance with Definition 9.

Then, the RFR probability $\Pr(X)$ of the C^3 self-complementary trinucleotide circular code X is equal to

$$\Pr(X) = \frac{1}{\mathrm{Card}(X)} \sum_{x \in X} \Pr(x, X). \quad (5)$$

The RFR probability $\Pr(X)$ of X is equal to 82.5% (Table 1).

The RFR probability $\Pr(\pi(X))$ of a bijective transformation code $\pi(X)$ of X is measured by the RFR probabilities of trinucleotides x of X which occur ("conserved") in $\pi(X)$, i.e.

$$\Pr(\pi(X)) = \frac{1}{\mathrm{Card}(X)} \sum_{x \in \pi(X)} \Pr(x, X). \quad (6)$$

Note obviously that $\Pr(\pi_0(X)) = \Pr(X)$ as $\pi_0(X) = X$ and $\mathrm{Card}(X) = \mathrm{Card}(\pi(X))$ for the 23 bijective transformation codes $\Pi(X)$.

The relative RFR probability $\mathrm{PrRFR}(\pi(X))$ of a bijective transformation code $\pi(X)$ of X is equal to

$$\mathrm{PrRFR}(\pi(X)) = \frac{\Pr(\pi(X))}{\Pr(X)}. \quad (7)$$

Note obviously that

$$\mathrm{PrRFR}(X) = 1 \quad (8)$$

as $\pi_0(X) = X$.

The relative RFR probabilities $\mathrm{PrRFR}(\pi(X))$ of the 23 bijective transformation codes $\Pi(X)$ range from 70.2% for $\pi_3(X)$ to 0 for $\pi_7(X)$ (Table 1).

2.3. Classical statistical tests used

We briefly recall basic information about the three classical statistical tests used. The reader can refer to basic statistical textbooks for details, e.g. Sokal and Rohlf (2012), Woolson (1987). We systematically report here bilateral tests (two-tailed p values).

2.3.1. Pearson product moment correlation coefficient r

Let two series of paired data $X(x_1, \ldots, x_n)$ and $Y(y_1, \ldots, y_n)$ of same size n. Then, the Pearson product moment correlation coefficient r is a parametric measure equal to

$$r = \frac{\sum_{i=1}^{n} (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^{n} (x_i - \bar{x})^2 \sum_{i=1}^{n} (y_i - \bar{y})^2}}$$

where $\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$ and $\bar{y} = \frac{1}{n}\sum_{i=1}^{n} y_i$.

If $r = 0$ then the two data series X and Y are not linearly correlated. If $r \pm 1$ then the two data series X and Y are perfectly linearly correlated, positively if $r = 1$ and negatively if $r = -1$. The two data series X and Y are much more correlated if r is far from 0.

2.3.2. Spearman rank correlation coefficient r_s

We have completed the coefficient r by a non-parametric measure based on the Spearman rank correlation coefficient r_s which is classically defined by the coefficient r between ranked series. This statistic test for correlations is not affected by extreme values, hence positive results are not due to outliers.

2.3.3. Fisher's exact test

The data for two independent random samples are y_{ij} where $y_{ij} = 0$ or $y_{ij} = 1$ for the ith, $i = 1, 2, \ldots, n_j$, observation in the group j, $j = 1,2$. The two independent random samples arise from populations with probability p_1 (p_2 respectively) of responses 1 in group 1 (2 respectively). A 2×2 table is constructed where a (c respectively) is the number of responses 1 in group 1 (2 respectively) and b (d respectively) is the number of responses 0 in group 1 (2 respectively). The null hypothesis is $H_0: p_1 = p_2$. Under H_0, the probability P_a of the observed table is determined from the hypergeometric distribution

$$P_a = \frac{(a+b)!(a+c)!(b+d)!(c+d)!}{(a+b+c+d)!a!b!c!d!}$$

and the probabilities of all tables more extreme than it, are computed.

2.4. Polymerization rate of a bijective transformation code $\pi(X)$

We define a polymerization rate for a bijective transformation code $\pi(X)$. During polymerization of nucleotide sequences, such as RNA transcription, bijective transformations correspond to misinsertions of a given nucleotide by another nucleotide. These nucleotide misinsertions are characterized by measured kinetic parameters corresponding to the affinity k (μM) and the maximal rate $V_{\max}$ (s^{-1}) according to classical Michaelis–Menten enzyme kinetic analyses. The affinity k reflects the reaction's rate when the enzyme is in excess as compared to the substrate while the maximal rate $V_{\max}$ corresponds to substrate excess conditions. We use Table 2 in Lee and Johnson (2006) which gives the values of $k(l_1l_2)$ and $V_{\max}(l_1l_2)$ for each of the 16 nucleotide insertions (l_1l_2), l_1, $l_2 \in A_4$: the 12 misinsertions (l_1l_2) with $l_2 \neq l_1$, and the four regular ('correct') insertions (l_1l_1), as measured for the human mitochondrial DNA polymerase gamma. Note that the bijective transformation (l_1l_2) is associated to $dl_2TP : dl_1$ in Table 2 in Lee and Johnson (2006).

Assuming similarity between DNA and RNA synthesis, we define the polymerization rate $pol(l_1l_2)$ of a nucleotide insertion (l_1l_2), l_1, $l_2 \in A_4$, by

$$pol(l_1l_2) = k(l_1l_2) \times V_{\max}(l_1l_2).$$

Indeed, the affinity k and the maximal rate $V_{\max}$ tend to be inversely proportional. Thus, the product $k \times V_{\max}$ characterizes

Table 2
Polymerization rate $pol(\pi(X))$ (Eqs. (9) and (10)) of the C^3 self-complementary trinucleotide circular code $X = \pi_0(X)$ and its 23 bijective transformation codes $\Pi(X)$.

	$pol(\pi(X))$
$X = \pi_0(X)$	11.2
$\pi_1(X)$	80.6
$\pi_2(X)$	111.7
$\pi_3(X)$	61.3
$\pi_4(X)$	69.5
$\pi_5(X)$	39.6
$\pi_6(X)$	113.7
$\pi_7(X)$	183.2
$\pi_8(X)$	126.3
$\pi_9(X)$	119.3
$\pi_{10}(X)$	94.8
$\pi_{11}(X)$	102.1
$\pi_{12}(X)$	155.4
$\pi_{13}(X)$	168.4
$\pi_{14}(X)$	54.6
$\pi_{15}(X)$	107.1
$\pi_{16}(X)$	75.4
$\pi_{17}(X)$	122.1
$\pi_{18}(X)$	151.6
$\pi_{19}(X)$	147.9
$\pi_{20}(X)$	176.9
$\pi_{21}(X)$	161.7
$\pi_{22}(X)$	68.8
$\pi_{23}(X)$	150.9

the rate across the whole range of conditions (from low to high nucleotide concentrations).

The definition of the polymerization rate $pol(l_1 l_2)$ of a nucleotide insertion $(l_1 l_2)$ is extended to the polymerization rate $pol(\pi(X))$ of a bijective transformation code $\pi(X)$ by summing over the four nucleotide insertions $(l_1 l_2)$ depending on $\pi(X)$. Precisely, for l_1, l_2, l_3, $l_4 \in A_4$, $l_1 \neq l_2 \neq l_3 \neq l_4$, the polymerization rate $pol(X)$ of the C^3 self-complementary trinucleotide circular code X is equal to

$$pol(X) = pol(l_1 l_1) + pol(l_2 l_2) + pol(l_3 l_3) + pol(l_4 l_4) \quad (9)$$

and the polymerization rate $pol(\pi(X))$ of a bijective transformation code $\pi(X)$ is equal to

$$pol(\pi(X)) = \begin{cases} pol(l_1 l_2) + pol(l_2 l_1) + pol(l_3 l_3) + pol(l_4 l_4) & \text{if } \pi(X) : (l_1 l_2) \\ pol(l_1 l_2) + pol(l_2 l_1) + pol(l_3 l_4) + pol(l_4 l_3) & \text{if } \pi(X) : (l_1 l_2)(l_3 l_4) \\ pol(l_1 l_2) + pol(l_2 l_3) + pol(l_3 l_1) + pol(l_4 l_4) & \text{if } \pi(X) : (l_1 l_2 l_3) \\ pol(l_1 l_2) + pol(l_2 l_3) + pol(l_3 l_4) + pol(l_4 l_1) & \text{if } \pi(X) : (l_1 l_2 l_3 l_4) \end{cases}. \quad (10)$$

Example 1. For the C^3 self-complementary trinucleotide circular code X, the polymerization rate $pol(X) = pol(AA) + pol(CC) + pol(GG) + pol(TT) = 25 \times 0.0036 + 140 \times 0.003 + 150 \times 0.066 + 57 \times 0.013 = 11.151$ μM/s. For the bijective transformation code $\pi_1(X) : (AC)$, the polymerization rate $pol(\pi_1(X)) = pol(AC) + pol(CA) + pol(GG) + pol(TT) = 540 \times 0.1 + 160 \times 0.1 + 150 \times 0.066 + 57 \times 0.013 = 80.641$ μM/s.

Table 2 gives the polymerization rates $pol(\pi(X))$ of $X = \pi_0(X)$ and its 23 bijective transformation codes $\Pi(X)$. The C^3 self-complementary trinucleotide circular code X has the lowest polymerization rate $pol(X) = 11.2$.

2.5. Amino acid chirality of a bijective transformation code $\pi(X)$

We define an amino acid chirality of a bijective transformation code $\pi(X)$. One of the strongest, earliest and most mysterious properties of the genetic code and translation process is that amino acids usually occurring in organisms are all L-rotated enantiomers. All amino acids in A_{20} have an L- and D-rotated form, at the exception of *Gly* which has no asymmetrical molecular structure. The biological preference for L amino acids putatively reflects small excess of L versus D enantiomers in spontaneous non-biogenic amino acids, such as those naturally occurring in meteorites (Engel and Nagy, 1982; Engel and Macko, 1997; Pizzarello and Cronin, 2000; Pizzarello et al., 2008).

The stereochemistry of D-nucleoside aminoacylation by the L enantiomer of an amino acid (Fig. 3 in Han et al., 2010) differs from that by the D enantiomer of the same amino acid (Fig. 4 in Han et al., 2010). The conformation energies E for L and D enantiomers (E_L and E_D) of the 19 amino acids in $A_{20} \setminus Gly$ were determined with D nucleosides for the four nucleotides *A*, *C*, *G* and *T* (*U*) for RNA molecules constrained on a surface (Bailey, 1998). Thus, $19 \times 4 = 76$ differences $E_L - E_D$ between L and D conformation energies are quantified for each amino acid-nucleotide interaction (Table 1 in Han et al., 2010). A difference indicates which amino acid enantiomer preferentially interacts with a nucleotide, a negative value indicating that an L enantiomer is preferred. 89.5% of amino acid-nucleotide interactions indicate preference for L-enantiomers (68 negative values among 76).

In order to get sums of positive values for the conformation energies of Table 1 in Han et al. (2010), we set the conformation energy $\Delta E_{aa}(l)$ for an amino acid aa in interaction with a nucleotide l as follows

$$\Delta E_{aa}(l) = -(E_L - E_D)$$

where $l \in A_4$ and $aa \in A_{20} \setminus Gly$.

Then, the aminoacylation conformation energy $AE(aa)$ measuring the interactions of the amino acid $aa \in A_{20} \setminus Gly$ assigned by the genetic code $\mathcal{G}$ (see Notation 1) with the nucleotides $\mathcal{C}(l_2)$ and $\mathcal{C}(l_1)$ at the second and third positions, respectively, of the anticodon $\mathcal{C}(l_1 l_2 l_3) = \mathcal{C}(l_3)\mathcal{C}(l_2)\mathcal{C}(l_1)$ complementary to the codon $l_1 l_2 l_3$, is defined by

$$AE(aa) = \frac{1}{2 \times \mathrm{Card}(\mathcal{G}^{-1}(aa))} \sum_{l_1 l_2 | l_1 l_2 l_3 = \mathcal{G}^{-1}(aa)} (\Delta E_{aa}(\mathcal{C}(l_1)) + \Delta E_{aa}(\mathcal{C}(l_2))). \quad (11)$$

The normalization $1/(2 \times \mathrm{Card}(\mathcal{G}^{-1}(aa)))$ represents the number of codons coded by an amino acid times 2 for the two nucleotide positions. The anticodon is used, rather than the codon, because the amino acid is supposed to be linked to its anticodon in the primitive pre-tRNA translation system (Seligmann and Amzallag, 2002).

Example 2. For the amino acid arginine $aa = Arg$ coded by the six codons $l_1 l_2 l_3$ in {*AGA*, *AGG*, *CGA*, *CGC*, *CGG*, *CGT*}, there are two nucleotides $\mathcal{C}(l_1 = A) = T$, four nucleotides $\mathcal{C}(l_1 = C) = G$ and six nucleotides $\mathcal{C}(l_2 = G) = C$. From Table 1 in Han et al. (2010), $\Delta E_{Arg}(C) = 2.89$, $\Delta E_{Arg}(G) = 0.04$ and $\Delta E_{Arg}(T) = 2.63$. Then, the aminoacylation conformation energy for *Arg* is equal to $AE(Arg) = (4 \times 0.04 + 2 \times 2.63 + 6 \times 2.89)/(2 \times 6) \approx 1.897$.

Table 3 gives the aminoacylation conformation energy AE_{aa} for the 19 amino acids $aa \in A_{20} \setminus Gly$. Seventeen amino acids have positive values AE_{aa} meaning preference for L-enantiomers while two amino acids *His* and *Lys* have negative values AE_{aa} meaning preference for D-enantiomers (note that naturally occurring enantiomers in organisms are L also for *His* and *Lys*).

Table 3
Aminoacylation conformation energy $AE(aa)$ (Eq. (11)) for the 20 amino acids except glycine. Seventeen amino acids have positive values $AE(aa)$ meaning preference for L-enantiomers while two amino acids *His* and *Lys* have negative values $AE(aa)$ meaning preference for D-enantiomers.

aa	$AE(aa)$
Ala	4.450
Arg	1.897
Asn	9.540
Asp	11.620
Cys	4.905
Gln	8.110
Glu	16.795
His	−0.075
Ile	2.530
Leu	2.997
Lys	−1.060
Met	3.995
Phe	8.610
Pro	1.470
Ser	5.113
Thr	3.800
Trp	1.140
Tyr	1.145
Val	3.200

Finally, the amino acid chirality $chiralAA(\pi(X))$ of a bijective transformation code $\pi(X)$ per trinucleotide is defined by

$$chiralAA(\pi(X)) = \frac{1}{20} \sum_{l_1l_2l_3 \in \pi(X)\setminus\{\mathcal{G}^{-1}(Gly),Stop\}} AE(\mathcal{G}(l_1l_2l_3)) \qquad (12)$$

where the genetic map $\mathcal{G}(l_1l_2l_3)$ gives the amino acid coded by the codon $l_1l_2l_3$ except for seven codons: the four codons $\mathcal{G}^{-1}(Gly) = \{GGA, GGC, GGG, GGT\}$ as the chirality of *Gly* is not defined and the three stop codons $Stop = \{TAA, TAG, TGA\}$ which do not code an amino acid. Table 4 gives the type and the occurrence number of amino acids and stop codons coded by the C^3 self-complementary trinucleotide circular code $X = \pi_0(X)$ and its 23 bijective transformation codes $\Pi(X)$ according to the (universal) genetic code $\mathcal{G}$. The normalization 1/20 allows to consider bijective transformation codes $\pi(X)$ containing several codons coding *Gly* and/or stop codons, i.e. bijective transformation codes $\pi(X)$ with no preferential chirality.

Example 3. For the C^3 self-complementary trinucleotide circular code X, the amino acid chirality is equal to $chiralAA(X) = (AE(Ala) + 2AE(Asn) + 2AE(Asp) + AE(Gln) + 2AE(Glu) + 2AE(Gly) + 2AE(Ile) + 2AE(Leu) + AE(Phe) + AE(Thr) + AE(Tyr) + 3AE(Val))/20 = (4.45 + 2 \times 9.54 + 2 \times 11.62 + 8.11 + 2 \times 16.795 + 2 \times 2.53 + 2 \times 2.997 + 8.61 + 3.8 + 1.145 + 3 \times 3.2)/20 \approx 6.134$.

Table 5 gives the amino acid chirality $chiralAA(\pi(X))$ of $X = \pi_0(X)$ and its 23 bijective transformation codes $\Pi(X)$. The C^3 self-complementary trinucleotide circular code X has the highest chirality rate $chiralAA(X) \approx 6.134$, i.e. X has the most L-oriented amino acids.

2.6. Transcript data

The complete sequence of the human mitochondrial genome is downloaded from GenBank (entry NC012920). Then, the 23 bijectively transformed genomes of the human mitochondrial genome are determined. These 23 bijectively transformed genomes are tested for nucleotide alignments in GenBank's ESTs database (January 2013) by using the Blast software (Altschul et al., 1997) with standard default alignment parameters. 101 RNAs with very high similarities with regions of the bijectively transformed human mitochondrial genome are recorded in tables of Seligmann (2013a,b). 82 RNAs with lengths varying from 32 to 451 nucleotides

Table 4
Type and occurrence number of amino acids and stop codons coded by the C^3 self-complementary trinucleotide circular code $X = \pi_0(X)$ and its 23 bijective transformation codes $\Pi(X)$ according to the (universal) genetic code $\mathcal{G}$. The next to last column gives the total number $NbAA(\pi(X))$ of amino acids coded by X and $\Pi(X)$. The last column presents the total number $NbAAStop(\pi(X))$ of amino acids plus the class of stop codons (*TAA,TAG,TGA*) coded by X and $\Pi(X)$.

	Ala	Arg	Asn	Asp	Cys	Gln	Glu	Gly	His	Ile	Leu	Lys	Met	Phe	Pro	Ser	Thr	Trp	Tyr	Val	Stop	$NbAA(\pi(X))$	$NbAAStop(\pi(X))$
$X = \pi_0(X)$	1	0	2	2	0	1	2	2	0	2	2	0	0	1	0	0	1	0	1	3	0	12	12
$\pi_1(X)$	4	0	0	0	0	1	1	2	0	1	3	0	1	0	2	1	1	0	0	3	0	11	11
$\pi_2(X)$	1	3	2	0	1	0	0	2	0	2	2	0	1	1	0	2	1	0	0	2	0	12	12
$\pi_3(X)$	1	0	1	2	0	1	1	2	1	1	2	0	0	1	0	1	0	0	1	4	1	13	14
$\pi_4(X)$	0	2	1	1	0	2	0	0	2	1	4	1	1	0	2	0	0	0	0	2	1	11	12
$\pi_5(X)$	3	0	2	2	0	0	2	2	1	1	0	0	0	0	1	2	2	0	0	1	1	11	12
$\pi_6(X)$	0	3	1	1	1	0	0	1	1	0	1	1	0	1	0	2	1	1	2	0	3	13	14
$\pi_7(X)$	1	3	0	0	1	1	0	1	0	0	2	0	0	0	2	5	1	1	0	0	2	10	11
$\pi_8(X)$	2	3	2	0	0	0	0	2	0	1	0	0	0	0	1	4	3	0	0	1	1	9	10
$\pi_9(X)$	0	1	0	1	0	2	1	0	1	0	6	1	1	0	2	0	0	1	0	1	2	11	12
$\pi_{10}(X)$	1	2	1	0	0	0	0	0	0	2	3	1	1	0	2	1	4	0	0	2	0	11	11
$\pi_{11}(X)$	4	0	1	2	0	1	1	2	1	0	1	0	0	0	2	1	1	0	1	1	1	13	14
$\pi_{12}(X)$	0	4	0	0	0	1	1	2	0	2	4	0	0	0	2	1	0	0	0	2	1	9	10
$\pi_{13}(X)$	1	1	1	1	2	0	1	2	2	0	1	0	0	1	0	2	0	1	1	0	3	13	14
$\pi_{14}(X)$	3	0	0	0	0	0	1	2	0	0	2	0	1	1	1	2	2	0	0	4	1	10	11
$\pi_{15}(X)$	0	3	1	0	1	0	0	1	0	3	2	1	1	1	0	3	1	1	0	1	0	13	13
$\pi_{16}(X)$	2	1	1	1	0	1	0	0	0	0	1	1	0	1	1	3	2	1	2	0	2	13	14
$\pi_{17}(X)$	0	4	1	1	2	2	0	1	2	1	1	1	0	0	2	1	0	0	1	0	0	13	13
$\pi_{18}(X)$	1	1	0	1	0	2	1	0	0	0	2	1	0	0	2	4	1	1	1	0	2	12	13
$\pi_{19}(X)$	1	3	1	0	1	0	0	1	0	1	1	1	0	0	2	3	4	0	0	0	1	11	12
$\pi_{20}(X)$	0	4	1	1	1	2	1	2	1	0	1	0	0	0	2	1	0	0	2	1	0	13	13
$\pi_{21}(X)$	2	1	0	0	2	0	1	2	0	0	1	0	0	1	1	4	2	1	0	0	2	11	12
$\pi_{22}(X)$	2	1	1	0	0	0	0	0	0	3	2	1	1	1	1	2	3	1	0	1	0	13	13
$\pi_{23}(X)$	0	4	0	0	0	1	0	1	0	1	6	0	0	0	2	1	0	1	0	1	2	9	10

Table 5
Amino acids chirality *chiralAA*($\pi(X)$) (Eq. (12)) of the C^3 self-complementary trinucleotide circular code $X = \pi_0(X)$ and its 23 bijective transformation codes $\Pi(X)$.

	chiralAA($\pi(X)$)
$X = \pi_0(X)$	6.134
$\pi_1(X)$	3.984
$\pi_2(X)$	3.911
$\pi_3(X)$	4.913
$\pi_4(X)$	3.391
$\pi_5(X)$	5.711
$\pi_6(X)$	2.984
$\pi_7(X)$	3.130
$\pi_8(X)$	3.636
$\pi_9(X)$	3.733
$\pi_{10}(X)$	3.221
$\pi_{11}(X)$	4.730
$\pi_{12}(X)$	3.200
$\pi_{13}(X)$	3.904
$\pi_{14}(X)$	4.042
$\pi_{15}(X)$	3.437
$\pi_{16}(X)$	3.923
$\pi_{17}(X)$	3.415
$\pi_{18}(X)$	4.270
$\pi_{19}(X)$	3.127
$\pi_{20}(X)$	4.157
$\pi_{21}(X)$	3.984
$\pi_{22}(X)$	3.645
$\pi_{23}(X)$	2.430

involve seven symmetric bijective transformation codes: 2 for π_2, 34 for π_3, 2 for π_4, 24 for π_5, 8 for π_6, 1 for π_7 and 11 for π_9. 19 RNAs with lengths varying from 68 to 513 nucleotides involve four asymmetric bijective transformation codes: 3 for π_{16}, 1 for π_{18}, 11 for π_{20} and 4 for π_{22}. For the regular transcription associated to the C^3 self-complementary trinucleotide circular code X, 10,904 RNAs are detected, not by alignment, but according to the number of RNAs from *Homo sapiens* in GenBank with a mitochondrial origin indicated in their annotation.

3. Results

3.1. Properties of the 23 bijective transformation codes $\Pi(X)$ of the C^3 self-complementary trinucleotide circular code X

Obvious from a combinatorial point of view but important from a biological point of view, and also not mentioned previously, we give two propositions which relate the 23 bijective transformation codes $\Pi(X)$ of X to the circular code property.

Proposition 1. *The 23 bijective transformation codes $\Pi(X)$ of X are C^3 trinucleotide circular codes.*

Proof. By letter invariance, $\Pi(X)$ belongs to the set of the 221,328 C^3 trinucleotide circular codes (Michel, unpublished) or by Proposition 3 in Michel and Pirillo (2010) or by Theorem 1 in Fimmel et al. (2014).

Proposition 2. *The seven bijective transformation codes $\Pi_{\mathcal{C}}(X) = \{\pi_3(X), \pi_4(X), \pi_7(X), \pi_8(X), \pi_9(X), \pi_{19}(X), \pi_{21}(X)\}$ are C^3 self-complementary trinucleotide circular codes.*

Proof. By letter invariance for the complementarity map $\mathcal{C}$, $\Pi_{\mathcal{C}}(X)$ belongs to the set of the 216 C^3 self-complementary trinucleotide circular codes identified in Arquès and Michel (1996) or by Proposition 3 in Michel and Pirillo (2010) or by Theorem 2 in Fimmel et al. (2014).

Proposition 3. *The bijective transformation code $\pi_7(X)$ has no trinucleotide $x \in X$ and a RFR probability* $\mathrm{PrRFR}(\pi_7(X)) = 0$.

Proof. By inspection of Table 1.

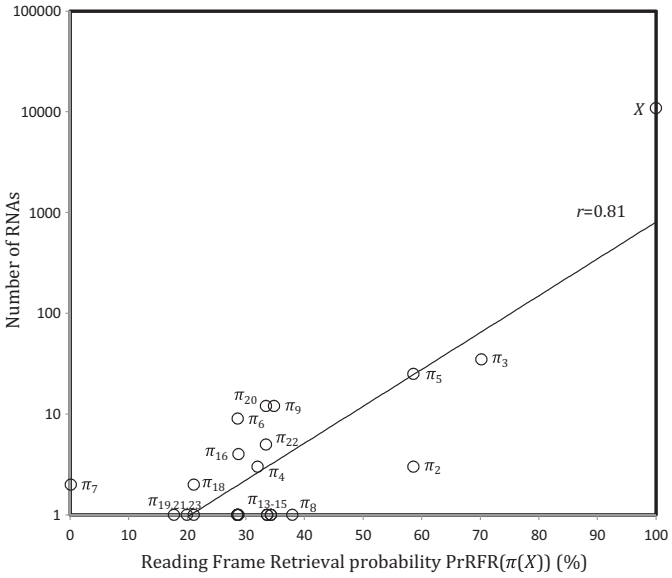

Fig. 1. Number of RNAs (logarithmic scale with number of RNAs plus one) in GenBank's EST database (January 2013) aligning with the bijectively transformed human mitochondrial genome as a function of the Reading Frame Retrieval (RFR) probability $\mathrm{PrRFR}(\pi(X))$ (%) (Eq. (7) and Table 1) of the C^3 self-complementary trinucleotide circular code $X = \pi_0(X)$ and its 23 bijective transformation codes $\Pi(X) = \{\pi_1(X), \ldots, \pi_{23}(X)\}$. The circular code X with a RFR probability $\mathrm{PrRFR}(X) = 1$ (Eq. (8)) has the highest number of RNAs (upper right corner). The juxtaposed datapoints of coordinates (29,1) correspond to π_1, π_{10}, π_{11}, π_{12} and π_{17}.

Proposition 4. *The bijective transformation code $\pi_3(X)$ has the maximum number (14) of trinucleotides $x \in X$ and the highest RFR probability* $\mathrm{PrRFR}(\pi_3(X)) = 0.702$.

Proof. By inspection of Table 1.

3.2. Number of RNAs aligning the bijectively transformed human mitochondrial genome and Reading Frame Retrieval (RFR) probability of X and $\Pi(X)$

The 23 bijective transformation codes $\Pi(X)$ of the C^3 self-complementary trinucleotide circular code X have 460 trinucleotides, 180 trinucleotides for the nine symmetric bijective transformation codes $\Pi_{\mathcal{S}}(X)$, and 280 trinucleotides for the 14 asymmetric bijective transformation codes $\Pi_{\mathcal{A}}(X)$. The nine symmetric codes $\Pi_{\mathcal{S}}(X)$ have 70 trinucleotides belonging to X, i.e. 38.9%, and the 14 asymmetric codes $\Pi_{\mathcal{A}}(X)$ have 82 trinucleotides belonging to X, i.e. 29.3% (Table 1). This percentage difference between symmetric and asymmetric bijective transformation codes is statistically significant ($p = 0.034$ with Fisher's exact test). In agreement with this theoretical observation, the number of RNAs (82, Section 2.6) from the GenBank's EST database involving symmetric bijective transformation codes is greater than the number of RNAs (19, Section 2.6) involving asymmetric bijective transformation codes.

Fig. 1 shows that the number of RNAs in GenBank's EST database (January 2013) aligning with the bijectively transformed human mitochondrial genome increases with the Reading Frame Retrieval RFR probability $\mathrm{PrRFR}(\pi(X))$ (Eq. (7) and Table 1) of the 23 bijective transformation codes $\Pi(X)$ of X: $r = 0.70$ with $p = 0.0002$ and $r_s = 0.42$ with $p = 0.048$ (bilateral tests).

If $X = \pi_0(X)$ is also considered with the 23 codes $\Pi(X)$, the statistical significance of the previous result obviously increases strongly: $r = 0.81$ with $p = 0.000002$ and $r_s = 0.49$ with $p = 0.019$ (bilateral tests).

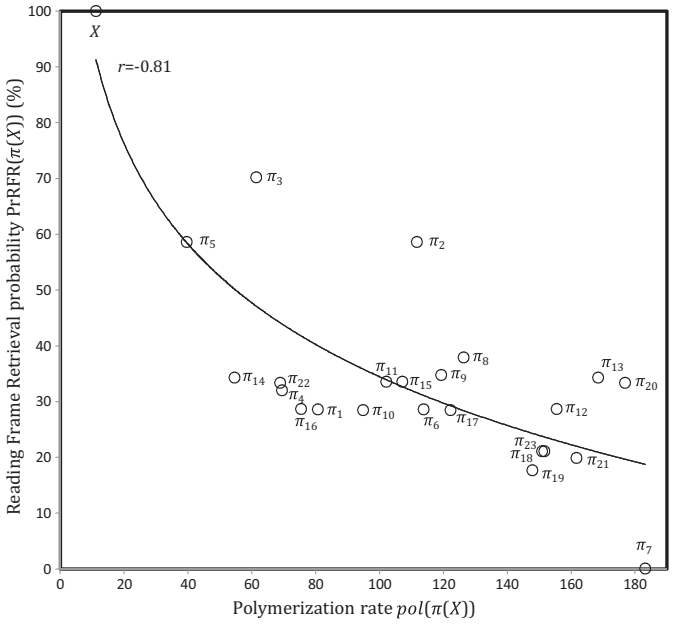

Fig. 2. Reading Frame Retrieval (RFR) probability PrRFR($\pi(X)$) (%) (Eq. (7) and Table 1) of the C^3 self-complementary trinucleotide circular code $X = \pi_0(X)$ and its 23 bijective transformation codes $\Pi(X) = \{\pi_1(X), \ldots, \pi_{23}(X)\}$ as a function of their polymerization rate $pol(\pi(X))$ (Eqs. (9) and (10), and Table 2). The circular code X with a RFR probability PrRFR(X) = 1 (Eq. (8)) has the lowest polymerization rate.

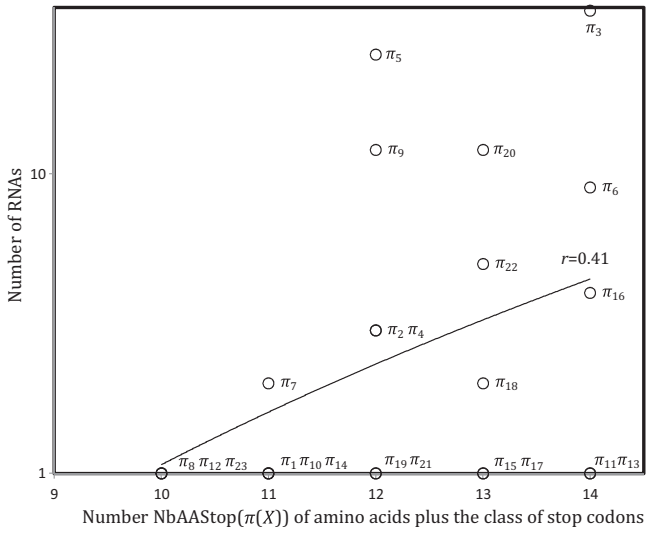

Fig. 3. Number of RNAs (logarithmic scale with number of RNAs plus one) in GenBank's EST database (January 2013) aligning with the bijectively transformed human mitochondrial genome as a function of the number NbAAStop($\pi(X)$) (Table 4) of amino acids plus stop codons which are coded by the 23 bijective transformation codes $\Pi(X) = \{\pi_1(X), \ldots, \pi_{23}(X)\}$ of X (X not included).

Furthermore, the previous correlation is also observed with some partitions of the 23 bijective transformation codes $\Pi(X)$:

(i) the nine symmetric codes $\Pi_S(X)$: $r = 0.69$ with $p = 0.041$ and $r_s = 0.58$ with $p = 0.103$ (bilateral tests; note that $\pi_7(X)$ has no statistical contribution according to Proposition 3);
(ii) the seven self-complementary codes $\Pi_C(X)$: $r = 0.83$ with $p = 0.020$ and $r_s = 0.50$ with $p = 0.220$ (bilateral tests; with the same remark for $\pi_7(X)$ as before);
(iii) the 16 non self-complementary codes $\Pi_{\bar{C}}(X)$: $r = 0.58$ with $p = 0.019$ and $r_s = 0.27$ with $p = 0.290$ (bilateral tests).

The previous result is not found with the 14 asymmetric codes $\Pi_A(X)$: $r = 0.29$ with $p = 0.310$ and $r_s = 0.23$ with $p = 0.400$ (bilateral tests).

Note that with these tests, the Spearman rank correlation coefficient r_s gives higher p values compared to the Pearson product moment correlation coefficient r.

3.3. Polymerization rate and reading frame retrieval (RFR) probability of X and Π(X)

Fig. 2 shows that the RFR probability PrRFR($\pi(X)$) (%) (Eq. (7) and Table 1) of the C^3 self-complementary trinucleotide circular code X and its 23 bijective transformation codes $\Pi(X)$ is inversely proportional to their polymerization rate $pol(\pi(X))$ (Eqs. (9) and (10), and Table 2): $r = -0.81$ with $p = 1.9 \times 10^{-6}$ and $r_s = -0.55$ with $p = 0.0088$ (bilateral tests). The circular code X with a RFR probability PrRFR(X) = 1 (Eq. (8)) has the lowest polymerization rate.

3.4. Amino acids coded by X and Π(X)

The C^3 self-complementary trinucleotide circular code X codes for NbAA(X) = 12 amino acids (Arquès and Michel (1996) and Table 4 therein). No stop codon {*TAA*, *TAG*, *TGA*} is coded by X. The 23 bijective transformation codes $\Pi(X)$ code from NbAA($\pi(X)$) = 9 amino acids ($\pi_8(X)$, $\pi_{12}(X)$, $\pi_{23}(X)$) to NbAA($\pi(X)$) = 13 amino acids ($\pi_3(X)$, $\pi_6(X)$, $\pi_{11}(X)$, $\pi_{13}(X)$, $\pi_{15}(X)$, $\pi_{16}(X)$, $\pi_{17}(X)$, $\pi_{20}(X)$, $\pi_{22}(X)$) (Table 4). If the class of stop codons is also considered as a coding signal (plus one if at least one stop codon is also coded by $\pi(X)$), then the 23 codes $\Pi(X)$ code from NbAAStop($\pi(X)$) = 10 amino acids plus stop codons ($\pi_8(X)$, $\pi_{12}(X)$, $\pi_{23}(X)$ to NbAAStop($\pi(X)$) = 14 amino acids plus stop codons ($\pi_3(X)$, $\pi_6(X)$, $\pi_{11}(X)$, $\pi_{13}(X)$, $\pi_{16}(X)$) (Table 4).

Fig. 3 suggests that the number of RNAs in GenBank's EST database (January 2013) aligning with the bijectively transformed human mitochondrial genome increases with the number NbAAStop($\pi(X)$) (Table 4) of amino acids plus stop codons which are coded by the 23 bijective transformation codes $\Pi(X)$ (X being not considered): $r = 0.41$ with $p = 0.052$ and $r_s = 0.46$ with $p = 0.032$ (bilateral tests). The statistical significances in this biological observation are not strong. In most statistical tests presented here, as with Fig. 3, a given direction is predictable according to biological rationales. For example, a positive correlation is expected between the numbers of RNAs following a bijective transformation of X and the numbers of genetic code signals coded by that bijective transformation of X. Hence, with unilateral tests, parametric and non-parametric statistical significances are $p = 0.026$ and $p = 0.016$, respectively, indicating a significant statistical association.

3.5. Amino acid chirality of X and Π(X)

Fig. 4 shows that the RFR probability PrRFR($\pi(X)$) (%) (Eq. (7) and Table 1) of the C^3 self-complementary trinucleotide circular code X and its 23 bijective transformation codes $\Pi(X)$ increases with their amino acid chirality $chiralAA(\pi(X))$ (Eq. (12) and Table 5): $r = 0.77$ with $p = 0.000009$ and $r_s = 0.58$ with $p = 0.0054$ (bilateral tests). The circular code X with a RFR probability PrRFR(X) = 1 (Eq. (8)) has the most L-oriented amino acids.

Fig. 5 indicates that the polymerization rate $pol(\pi(X))$ (Eqs. (9) and (10), and Table 2) of the C^3 self-complementary trinucleotide circular code X and its 23 bijective transformation codes $\Pi(X)$

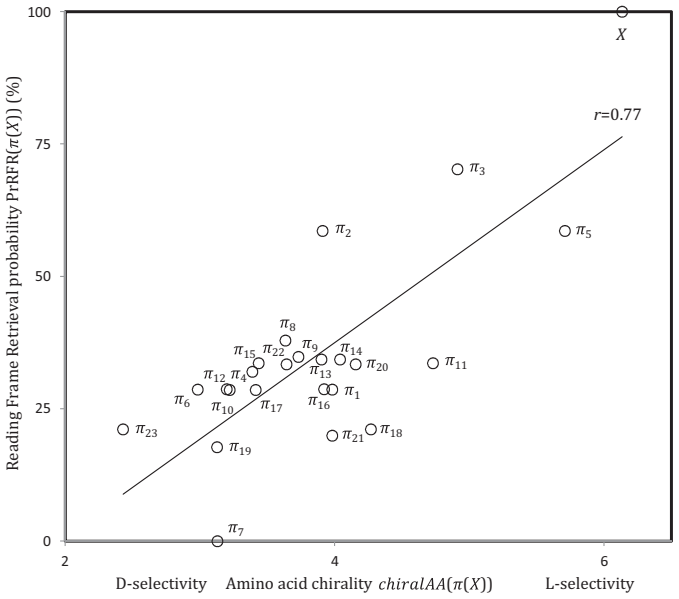

Fig. 4. Reading Frame Retrieval (RFR) probability PrRFR($\pi(X)$) (%) (Eq. (7) and Table 1) of the C^3 self-complementary trinucleotide circular code $X = \pi_0(X)$ and its 23 bijective transformation codes $\Pi(X) = \{\pi_1(X), \ldots, \pi_{23}(X)\}$ as a function of their amino acid chirality $chiralAA(\pi(X))$ (Eq. (12) and Table 5). The circular code X with a RFR probability PrRFR(X) = 1 (Eq. (8)) has the most L-oriented amino acids.

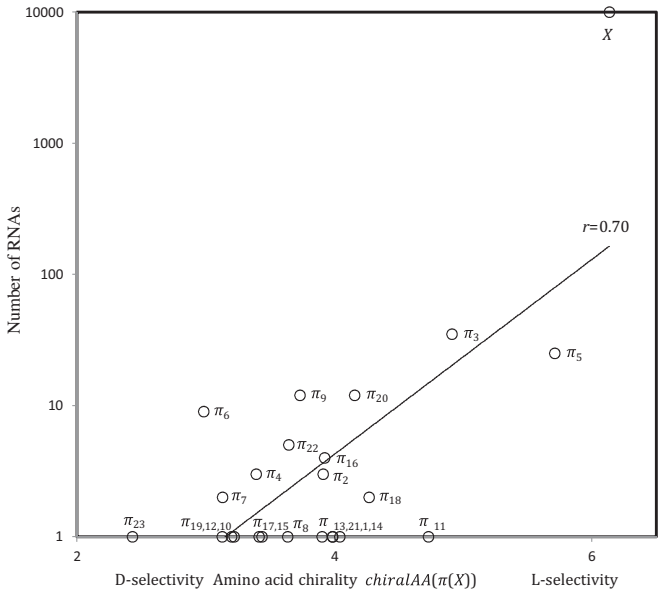

Fig. 6. Number of RNAs (logarithmic scale with number of RNAs plus one) in GenBank's EST database (January 2013) aligning with the bijectively transformed human mitochondrial genome as a function of the amino acid chirality $chiralAA(\pi(X))$ (Eq. (12) and Table 5) of the C^3 self-complementary trinucleotide circular code $X = \pi_0(X)$ and its 23 bijective transformation codes $\Pi(X) = \{\pi_1(X), \ldots, \pi_{23}(X)\}$. The circular code X with the highest number of RNAs has the most L-oriented amino acids.

decreases with their amino acid chirality $chiralAA(\pi(X))$ (Eq. (12) and Table 5): $r = -0.72$ with $p = 0.00008$ and $r_s = -0.42$ with $p = 0.040$ (bilateral tests). The circular code X with the lowest polymerization rate has the most L-oriented amino acids.

Fig. 6 suggests that the number of RNAs in GenBank's EST database aligning with the bijectively transformed human mitochondrial genome increases with the amino acid chirality $chiralAA(\pi(X))$ (Eq. (12) and Table 5) of the C^3 self-complementary trinucleotide circular code X and its 23 bijective transformation codes $\Pi(X)$: $r = 0.70$ with $p = 0.0001$ and $r_s = 0.45$ with $p = 0.032$ (bilateral tests). The circular code X with the highest number of RNAs has the most L-oriented amino acids.

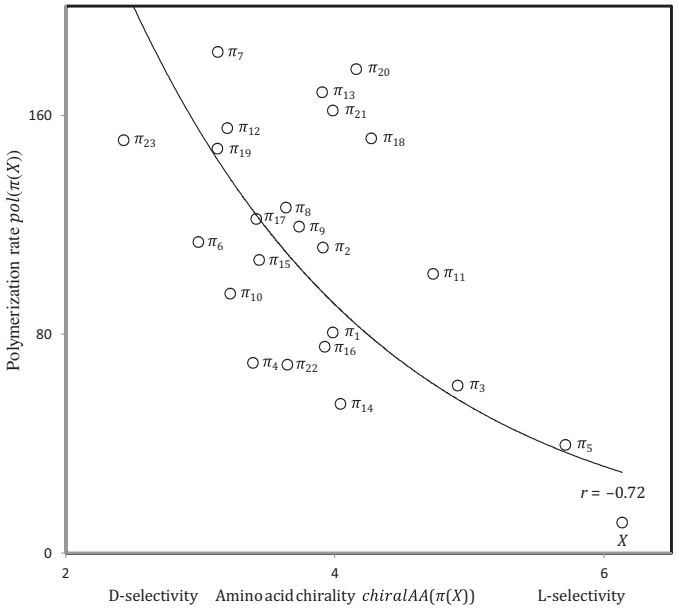

Fig. 5. Polymerization rate $pol(\pi(X))$ (Eqs. (9) and (10), and Table 2) of the C^3 self-complementary trinucleotide circular code $X = \pi_0(X)$ and its 23 bijective transformation codes $\Pi(X) = \{\pi_1(X), \ldots, \pi_{23}(X)\}$ as a function of their amino acid chirality $chiralAA(\pi(X))$ (Eq. (12) and Table 5). The circular code X with the lowest polymerization rate has the most L-oriented amino acids.

4. Discussion

The circular code X enables to retrieve and maintain the reading frames of genes (Arquès and Michel, 1996; Gonzalez et al., 2011). It is not known whether this mathematical property is actually used by the cell's translational apparatus. However, the identification of regions with X circular code motifs (stretches of trinucleotides of X) in 16S ribosomal and transfer RNAs which are in direct contact with messenger RNAs (reading frames of genes), suggests a possible translation (framing) code in genes based on circular code (Michel, 2012). Furthermore, the identification of X circular code motifs and a gene circular code property in transfer RNAs of prokaryotes and eukaryotes strengthens the previous concept (Michel, 2013).

On the other hand, in mitochondria, some RNAs correspond to DNA sequences when RNA transcription systematically exchanges between nucleotides (Seligmann, 2012, 2013a,b,c). Furthermore, these RNAs are produced by different types of nucleotide exchanges and at different frequencies (Seligmann, 2013a,b). This observation suggests that the C^3 self-complementary trinucleotide circular code X which is associated with the regular RNA transcription, may use its bijective transformation codes $\Pi(X)$ for coding nucleotide exchanging RNA transcription.

The 23 bijective transformation codes $\Pi(X)$ of X are C^3 trinucleotide circular codes, i.e. with the property of reading frame retrieval and maintenance, and furthermore, seven of them are in addition self-complementary (Propositions 1 and 2).

The Reading Frame Retrieval (RFR) probability (Eq. (4)) of bijective transformation codes $\Pi(X)$ is measured as function of

trinucleotides of X which occur ("conserved") in $\Pi(X)$. Indeed, we assume that the process of reading frame retrieval by the ribosome remains unchanged and follows the "mathematical property" of the circular code X, and not one of its bijective transformation codes $\Pi(X)$. This assumption also implies that the genetic code map remains unchanged, i.e. codons after bijective transformation code for amino acids according to the genetic code. An alternative hypothesis which we believe too complex to coexist with regular transcription and which was not considered here, would be that codons after bijective transformation code for the same amino acid as before bijective transformation. In other words, we believe that the bijective transformation codes $\Pi(X)$ do not change the genetic code map, but they only alter the RNA produced. According to this assumption, the bijective transformation codes $\Pi(X)$ have different probabilities of reading frame retrieval (RFR), ranging from 70.2% for $\pi_3(X)$ to 0 for $\pi_7(X)$ (Table 1). These RFR probabilities of $\Pi(X)$ depend on two factors: the number of trinucleotides of X in $\Pi(X)$ and their RFR probabilities (Table 1). In general, the less trinucleotides of X remain in $\Pi(X)$, the lower its RFR probability. From a biological point of view, according to the previous assumption that the genetic code map does not change the amino acid-codon assignments in the bijective transformation codes $\Pi(X)$, the ribosome will only slowly adapt to the alterations in the punctuation of the circular code X.

Several notable correlations exist between the RFR probability of bijective transformation codes $\Pi(X)$ and the different biological properties of $\Pi(X)$ related to their numbers of RNAs in GenBank's EST database (January 2013) aligning with the bijectively transformed human mitochondrial genome, their polymerization rate, their number of amino acids and their amino acid chirality. The RFR probability of bijective transformation codes $\Pi(X)$ correlates: (i) positively with the RNA numbers in GenBank's EST database (Fig. 1); (ii) negatively with the polymerization rates (Fig. 2); and (iii) positively with the amino acid chirality (Fig. 4). The RNA numbers in GenBank's EST database correlates: (i) positively with the numbers of amino acids (Fig. 3); and (ii) positively with the amino acid chirality (Fig. 6). The polymerization rate correlates negatively with the amino acid chirality (Fig. 5).

These results suggest that the polymerase system may preferentially use the bijective transformation codes $\Pi(X)$ coding for a large number of amino acids and with preferential L-selectivity. However, correlations are stronger with chirality (preferential L-selectivity with $r = 0.70$ in Fig. 6) than with coding diversity (number of amino acids with $r = 0.41$ in Fig. 3). Hence, the function of the circular code X and its bijective transformation codes $\Pi(X)$ with the structure of the genetic code and the transcription system that polymerizes RNA, may be more related to chirality of amino acids than to the coding diversity of X. Therefore, the circular code X and its bijective transformation codes $\Pi(X)$ probably occurred at the earliest stages of evolution of life's chemistry, when amino acids were selected according to the chiral stereophysical selectivity of their interactions with RNA.

The bijective transformation codes $\Pi(X)$ corresponding to frequent nucleotide exchanging polymerizations, have codons with a high probability of Reading Frame Retrieval (Fig. 1) and also codons coding for numerous translational signals (amino acids and stop codons) (Fig. 3).

The circular code X and its bijective transformation codes $\Pi(X)$ coevolved with the process that selected L-rotated amino acids. Chiral selectivity for L-enantiomers (Han et al., 2010) decreases with the order of inclusion of amino acids in the genetic code (Trifonov, 2000), suggesting that the codes X and $\Pi(X)$ evolved when the first amino acids were assigned to codons.

Symmetric bijective transformation codes $\Pi_S(X)$ are more preferentially used compared to asymmetric bijective transformation codes $\Pi_A(X)$, from a theoretical point of view with the RFR probability (see the results obtained with the different partitions of $\Pi(X)$ in Section 3.2) and from the observed numbers of RNAs in GenBank's EST database (82 RNAs for $\Pi_S(X)$ and 19 RNAs for $\Pi_A(X)$, Section 2.6).

In summary, a theoretical study of the C^3 self-complementary trinucleotide circular code X and its 23 bijective transformation codes $\Pi(X)$ based on the probability of Reading Frame Retrieval, the polymerization rate and the amino acid chirality, is developed here. It is applied to analyze the nucleotide exchanging RNA transcription in human mitochondria. The obtained results suggest that the circular code X with the functions of reading frame retrieval and maintenance in regular RNA transcription, may also have, through its bijective transformation codes $\Pi(X)$, the same functions in nucleotide exchanging RNA transcription.

Acknowledgments

We thank two anonymous reviewers for their advices.

References

Ahmed, A., Frey, G., Michel, C.J., 2007. Frameshift signals in genes associated with the circular code. In Silico Biol. 7, 155–168.
Ahmed, A., Frey, G., Michel, C.J., 2010. Essential molecular functions associated with circular code evolution. J. Theor. Biol. 264, 613–622.
Ahmed, A., Michel, C.J., 2011. Circular code signal in frameshift genes. J. Comp. Sci. Syst. Biol. 4, 7–15.
Altschul, S.F., Madden, T.L., Schaeffer, A.A., Zhang, J., Zhang, Z., Miller, W., Lipman, D.J., 1997. Gapped BLAST and PSI-BLAST: a new generation of protein database programs. Nucleic Acids Res. 25, 3389–3402.
Arquès, D.G., Michel, C.J., 1996. A complementary circular code in the protein coding genes. J. Theor. Biol. 182, 45–58.
Bailey, J.M., 1998. RNA-directed amino acid homochirality. FASEB J. 12, 503–507.
Engel, M.H., Macko, S.A., 1997. Isotopic evidence for extraterrestrial non-racemic amino acids in the Murchison meteorite. Nature 389, 265–268.
Engel, M.H., Nagy, B., 1982. Distribution and enantiomeric composition of amino acids in the Murchison meteorite. Nature 296, 837–840.
Fimmel, E., Danielli, A., Strüngmann, L., 2013. On dichotomic classes and bijections of the genetic code. J. Theor. Biol. 336, 221–230.
Fimmel, E., Giannerini, S., Gonzalez, D.L., Strüngmann, L., 2014. Circular codes, symmetries and transformations. J. Math. Biol. (in press).
Gonzalez, D.L., Giannerini, S., Rosa, R., 2011. Circular codes revisited: a statistical approach. J. Theor. Biol. 275, 21–28.
Han da, X., Wang, H.Y., Ji, Z.L., Hu, A.F., Zhao, Y.F., 2010. Amino acid homochirality may be linked to the origin of phosphate-based life. J. Mol. Evol. 70, 572–582.
Itzkovitz, S., Alon, U., 2007. The genetic code is nearly optimal for allowing additional information within protein-coding sequences. Genome Res. 17, 405–412.
Lee, H.R., Johnson, K.A., 2006. Fidelity of the human mitochondrial DNA polymerase. J. Biol. Chem. 281, 36236–36240.
Michel, C.J., 2012. Circular code motifs in transfer RNA and 16S ribosomal RNAs: a possible translation code in genes. Comput. Biol. Chem. 34, 24–37.
Michel, C.J., 2013. Circular code motifs in transfer RNAs. Comput. Biol. Chem. 45, 17–29.
Michel, C.J., Pirillo, G., 2010. Identification of all trinucleotide circular codes. Comput. Biol. Chem. 34, 122–125.
Pizzarello, S., Cronin, J.R., 2000. Non-racemic amino acids in the Murray and Murchison meteorites. Geochim. Cosmochim. Acta 64, 329–338.
Pizzarello, S., Huang, Y., Alexandre, M.R., 2008. Molecular asymmetry in extraterrestrial chemistry: insights from a pristine meteorite. Proc. Natl. Acad. Sci. U.S.A. 105, 3700–3704.
Seligmann, H., Amzallag, G.N., 2002. Chemical interactions between amino acid and RNA: multiplicity of the levels of specificity explains origin of the genetic code. Naturwissenschaften 89, 542–551.
Seligmann, H., Pollock, D.D., 2004. The ambush hypothesis: hidden stops prevent off-frame gene reading. DNA Cell Biol. 23, 701–705.
Seligmann, H., 2007. Cost minimization of ribosomal frameshifts. J. Theor. Biol. 249, 162–167.
Seligmann, H., 2010. The ambush hypothesis at the whole-organism level: off frame, 'hidden' stops in vertebrate mitochondrial genes increase developmental stability. Comput. Biol. Chem. 34, 80–85.
Seligmann, H., 2012. Overlapping genes coded in the 3′-to-5′-direction in mitochondrial genes and 3′-to-5′ polymerization of non-complementary RNA by an 'invertase'. J. Theor. Biol. 315, 38–52.
Seligmann, H., 2013a. Systematic asymmetric nucleotide exchanges produce human mitochondrial RNAs cryptically encoding for overlapping protein coding genes. J. Theor. Biol. 324, 1–20.

Seligmann, H., 2013b. Polymerization of non-complementary RNA: systematic symmetric nucleotide exchanges mainly involving uracil produce mitochondrial RNA transcripts coding for cryptic overlapping genes. BioSystems 111, 156–174.
Seligmann, H., 2013c. Triplex DNA:RNA, 3′-to-5′ inverted RNA and protein coding in mitochondrial genomes. J. Comput. Biol. 200, 660–671.
Sokal, R.R., Rohlf, F.J., 2012. Biometry: The Principles and Practice of Statistics in Biological Research, fourth ed. Freeman W.H. and Co, New York.
Trifonov, E.N., 2000. Consensus temporal order of amino acids and evolution of the triplet code. Gene 261, 139–151.
Woolson, R.F., 1987. Statistical Methods for the Analysis of Biomedical Data. Wiley Series in Probability and Mathematical Statistics. John Wiley & Sons, New York.