



HAL
open science

Palindromic Vectors, Symmetry and Symmetry as Symmetry Descriptors of Binary Data

Jean Marc Girault, Sébastien Ménigot

► **To cite this version:**

Jean Marc Girault, Sébastien Ménigot. Palindromic Vectors, Symmetry and Symmetry as Symmetry Descriptors of Binary Data. *Entropy*, 2022, 24 (1), 10.3390/e24010082 . hal-03508162

HAL Id: hal-03508162

<https://hal.science/hal-03508162v1>

Submitted on 3 Jan 2022

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons Attribution 4.0 International License

Palindromic Vectors, Symmetry and Symmetry as Symmetry Descriptors of Binary Data

Jean-Marc Girault ^{1,2,*} and Sébastien Ménigot ^{1,2}

¹ Groupe ESEO, 49000 Angers, France; sebastien.menigot@eseo.fr

² Laboratoire d'Acoustique de l'Université du Mans (LAUM), UMR 6613, Institut d'Acoustique-Graduate School (IA-GS), CNRS, Le Mans Université, 72085 Le Mans, France

* Correspondence: jean-marc.girault@eseo.fr

Version accepted in Entropy

Abstract: Today, the palindromic analysis of biological sequences, based exclusively on the study of “mirror” symmetry properties, is almost unavoidable. However, other types of symmetry, such as those present in friezes, could allow us to analyze binary sequences from another point of view. New tools, such as symmetry and symmetry, based on new types of palindromes allow us to discriminate binarized $1/f$ noise sequences better than Lempel–Ziv complexity. These new palindromes with new types of symmetry also allow for better discrimination of binarized DNA sequences. A relative error of 6% of symmetry is obtained from the HUMHBB and YEAST1 DNA sequences. A factor of 4 between the slopes obtained from the linear fits of the local symmetries for the two DNA sequences shows the discriminative capacity of the local symmetry. Moreover, it is highlighted that a certain number of these new palindromes of sizes greater than 30 bits are more discriminating than those of smaller sizes assimilated to those from an independent and identically distributed random variable.

Keywords: palindrome; palindromic vectors; symmetry; symmetry; symmetry; entropy; complexity; descriptor; binary sequence

Received: November 24, 2021. Accepted: December 31, 2021. Published: January 3, 2022.

Link: <https://doi.org/10.3390/e24010082>

1. Introduction

The palindromic analysis of discrete sequences has partly revolutionized molecular biology and is widely used as shown by the following work [1–8], to name a few. Very recently, the study of quantum behavior [9], encountered in palindromes within the DNA structure, revealed that the symmetry properties of the unitary structure, other than those present in classical palindromes, play an important role in the origin and cause of mutations.

In the continuity of the work carried out by Tibatan and Sarisaman [9], our article aims to highlight the symmetry links between the concept of frieze and the concept of palindrome, which have been insufficiently exploited until now in the analysis of binary data.

The “mirror” symmetry on which the concept of palindrome was based is certainly the basis of the oldest symmetry descriptors. Its greatest success is undoubtedly derived from the analysis of biological sequences (DNA, RNA and proteins), even if in this case the definition of DNA palindromes is slightly different from the classical definition. (Let us consider the sequence of characters ‘ATGGCCAT’. It is qualified as a 8-palindrome sequence. It is composed of a 4-pattern on the right (‘CCAT’) obtained by a mirror reflection of its 4-pattern complementary on the left (‘ATGG’): ‘ATGG’ \bullet ‘CCAT’ \bullet , where \bullet indicates the mirror reflection and \bullet indicates the complementary. Note that ‘T’ is the complementary of ‘A’, and ‘C’ is the complementary of ‘G’.)

To fix ideas, a palindrome of size m , called “ m -palindrome”, is a discrete sequence composed of two contiguous symmetrical (mirror) sub-sequences each composed of k -patterns with $k = \lfloor m/2 \rfloor$. For example, the alphabetic character sequence dddddddbbbbbbb is a 16–palindrome composed of two 8–patterns:

35 *ddddddd, bbbbbbb*. Even if the theoretical research around the palindrome is still going on, as shown by the
 36 recent article by Gabric and Shallit [10] to name but a few, it is the older work of Allouche et al. [11,12], which
 37 is used as a starting point in this work and in particular the notion of **palindromic complexity**.

38 Today, when studying a word or a discrete sequence, its analysis is still limited to only one type of symmetry:
 39 the “mirror” symmetry. Wanting to extract many more intrinsic features in the discrete sequences studied can
 40 consist of looking for other types of symmetries, as it is explicitly the case in friezes.

41 A frieze is a horizontal strip composed of an infinite number of symmetrical patterns, i.e., a periodic
 42 geometric object. As an illustration, five types of alphabetical sequences of 16 characters, having the same
 43 symmetries as friezes, are presented as follows: *bbbbbbbbbbbbbb, dbdbdbdbdbdbdb, bpbpbpbpbpbpbpb,*
 44 *bqbqbqbqbqbqbqb, bqpdbqpdbqpdbqpd.*

45 If the objective is indeed to extend the analysis of discrete periodic sequences to other types of sequences,
 46 then the search for all symmetric patterns is the next step. To reach this goal, the concept of palindrome and
 47 then that of frieze is presented in Section 2. Then, the concept of palindromes is extended and new tools such as
 48 symmetropy and symmentropy are proposed in Section 3. Finally, the set of symmetry descriptors are tested on
 49 binarized $1/f$ noises and binarized DNA sequences in Section 4; then, the results are discussed in Section 5.

50 2. Palindromes and Friezes

51 In this section, we recall the concept of palindromes [11–13] and the concept of friezes [14,15].

52 2.1. Palindromes

53 For a binary sequence, an m -palindrome is, by definition, a grouping of m bits that form an m -pattern
 54 of mirror symmetry. In other words, for a binary sequence $\mathbf{X} = \{x(1), x(2), \dots, x(M)\}$ composed of M
 55 bits, an m -palindrome can be defined as the concatenation of two k -patterns: $\mathbf{X}_m(i) = [\mathbf{X}_k(i) \Gamma_R[\mathbf{X}_k(i)]$,
 56 with $k = \lfloor m/2 \rfloor$ being the order of the palindrome. The first k -pattern $\mathbf{X}_k(i) = \{x(i), x(i+1), \dots, x(i+)$
 57 $k-1)\}$, $1 \leq i \leq M-k+1$ is the reference pattern, and the second k -pattern obtained by $\Gamma_R[\mathbf{X}_k(i)]$ is the
 58 symmetric pattern, where $\Gamma_R[\bullet]$ is the transformation corresponding to the mirror symmetry, a reflection.

59 For example, for the binary sequence $\mathbf{X} = \{01100110\}$ of 8 bits, the first 4-pattern of \mathbf{X} of order 2 is written
 60 as $\mathbf{X}_4(1) = [\mathbf{X}_2(1) \Gamma_R[\mathbf{X}_2(1)] = [\{01\}\{10\}] = \{0110\}$, with $\mathbf{X}_2(1) = \{01\}$ and $\Gamma_R[\mathbf{X}_2(1)] = \{10\}$. In the
 61 same way, the 8-palindrome of \mathbf{X} of order 4 is written as $\mathbf{X}_8(1) = [\mathbf{X}_4(1) \Gamma_R[\mathbf{X}_4(1)] = [\{0110\}\{0110\}] =$
 62 $\{01100110\}$, with $\mathbf{X}_4(1) = \{0110\}$ and $\Gamma_R[\mathbf{X}_4(1)] = \{0110\}$.

63 A palindrome of odd length can be seen as the concatenation of a pattern of size $(m-1)$ and its mirror, for
 64 which the rightmost bit of the $(m-1)$ -reference pattern (bit in bold in the following example) and the leftmost
 65 bit of the $(m-1)$ -mirror pattern (bit in bold in the following example) are merged to give only one. Example:
 66 $[\{01\}\{10\}] = \{01 \mathbf{10}\}$ becomes $\{010\}$.

67 Although there is a plethora of scalar descriptors such as those indicated in [11–13] to name but a few,
 68 here, we limit ourselves to the concept of palindromic complexity \tilde{c} computed from \mathbf{D} , which lists, from the
 69 palindromic dictionary, the cardinal of the **different** palindromic words of size m :

$$70 \quad \mathbf{D} = [d(0), d(1), \dots, d(m), \dots, d(M)]^t. \quad (1)$$

71 where $d(m)$ is the cardinal of “palindrome words” of size m [11] present in the binary sequence. The empty
 72 palindrome obtained for $m=0$ is e and $\{e, 0, 1\}$ are the trivial palindromes. The **palindromic complexity** \tilde{c} ,
 73 which corresponds to the cardinal of \mathbf{D} , is defined by the following:

$$74 \quad \tilde{c} = \text{card}(\mathbf{D}). \quad (2)$$

75 In order to measure the level of mirror symmetry present in a binary sequence, we propose to count the
 76 frequency of occurrence of m -palindromic patterns in the binary sequence studied by the following:

$$77 \quad \mathbf{V} = [v(0), v(1), \dots, v(m), \dots, v(M)]^t. \quad (3)$$

78 where $v(m)$ is the frequency of occurrence of a palindromic pattern of size m . The “mirror” symmetry level $\tilde{\sigma}$ is
 79 the sum of all occurrences for non-trivial palindromes:

$$80 \quad \tilde{\sigma} = \sum_{m=2}^M v(m). \quad (4)$$

81 An illustration given in Table 1 for the binary sequence $\mathbf{X} = \{01101001\}$, specifies the value of the
 82 palindromic complexity $\tilde{c} = 5$. There are, in all, five non-zero elements in $\mathbf{D} = \{1, 2, 2, 2, 2, 0, 0, 0, 0\}$ which is
 83 itself computed from the empirical palindromic dictionary $Dict = \{e, 0, 1, 00, 11, 010, 101, 0110, 1001\}$.

Table 1. $Dict$, $d(m)$ and $v(m)$ calculated from the binary sequence $\mathbf{X} = \{01101001\}$ composed of $M = 8$ bits. There are in total $\tilde{c} = 5$ sizes of palindromes (0, 1, 2, 3, 4) derived from the dictionary and used in the binary sequence \mathbf{X} . There are two palindromes of size 2, two palindromes of size 3, and two palindromes of size 4, so a total of $\tilde{\sigma} = 6 = 2 + 2 + 2$ palindromes composing the binary sequence.

m	0	1	2	3	4	5	6	7	8
$Dict$	e	0,1	00,11	101,010	0110,1001	-	-	-	-
$d(m)$	1	2	2	2	2	0	0	0	0
$v(m)$	8	8	2	2	2	0	0	0	0

84 2.2. Friezes

85 As stated in the Introduction, a frieze is a periodic horizontal band composed of a few basic symmetrical
 86 patterns repeated *ad infinitum*. There are only seven different types of friezes [14,15] (see Figure 1) obtained from
 87 five types of isometries (isometry is a geometrical transformation that leaves the objects invariant thus transformed
 88 while preserving the distances, which is the case for the five following operations: translation, vertical reflection,
 89 horizontal reflection, inversion, and glide reflection). (**TRIGH**: Translation, vertical **R**eflection, **I**nversion and
 90 **G**lide reflection, **H**orizontal reflection). There are only 5 possible types of periodic discrete sequences obtained
 91 from 4 types of isometries (**TRIG**: Translation, vertical **R**eflection, **I**nversion and **G**lide reflection), vertical
 92 reflection not allowing to obtain a 1D-sequences.

93 For example, from the friezes in Figure 1 and replacing Γ by $\{10\}$, we can construct five types of periodic
 94 discrete sequences, all having different types of symmetry:

- 95 • sequence $\mathbf{X} = \{10101010 \dots\}$ obtained with translations;
- 96 • sequence $\mathbf{X} = \{10011001 \dots\}$ obtained with vertical reflections (mirror);
- 97 • sequence $\mathbf{X} = \{10011001 \dots\}$ obtained with glide reflections and translations;
- 98 • sequence $\mathbf{X} = \{10101010 \dots\}$ obtained with inversions and translations;
- 99 • sequence $\mathbf{X} = \{10100101 \dots\}$ obtained with inversions and vertical reflections.

100 Among the five previous sequences, two are composed of mirror palindromes (the second and the last). By
 101 no longer limiting the search to mirror palindromes, it should be possible to describe binary sequences more
 102 precisely; this is the subject of the next section.

┌	┌	┌	┌	┌
┌-----┐	┌-----┐	┌-----┐	┌-----┐	┌-----┐
┌	┌	┌	┌	┌
┌-----┐	┌-----┐	┌-----┐	┌-----┐	┌-----┐
┌	┌	┌	┌	┌
┌-----┐	┌-----┐	┌-----┐	┌-----┐	┌-----┐
┌	┌	┌	┌	┌
┌-----┐	┌-----┐	┌-----┐	┌-----┐	┌-----┐
┌	┌	┌	┌	┌
┌-----┐	┌-----┐	┌-----┐	┌-----┐	┌-----┐
┌	┌	┌	┌	┌
┌-----┐	┌-----┐	┌-----┐	┌-----┐	┌-----┐
┌	┌	┌	┌	┌
┌-----┐	┌-----┐	┌-----┐	┌-----┐	┌-----┐
┌	┌	┌	┌	┌
┌-----┐	┌-----┐	┌-----┐	┌-----┐	┌-----┐
┌	┌	┌	┌	┌
┌-----┐	┌-----┐	┌-----┐	┌-----┐	┌-----┐
┌	┌	┌	┌	┌
┌-----┐	┌-----┐	┌-----┐	┌-----┐	┌-----┐

Figure 1. The seven types of friezes with a \lceil pattern. The friezes 1, 2, 4, 5 and 6 can constitute periodic discrete sequences because no pattern appears with the same abscissa. This is not the case for friezes 3 and 7, which cannot constitute a discrete sequence. Among the five periodic sequences, friezes 2 and 6 are composed of palindromes.

103 3. Methods

104 In this section, we propose to extend the different palindromic vector and scalar descriptors by integrating
 105 the different types of symmetry revealed in the friezes. Then, new palindromic descriptors such as the notions of
 106 symmetry and symmentropy are proposed.

107 As mentioned later, through the notion of friezes, several types of symmetries can be considered using the
 108 combination of only four isometries (TRIG). On this basis, we propose to generalize the notion of palindromes
 109 by taking into account all types of symmetries.

110 For a binary sequence $\mathbf{X} = \{x(1), x(2), \dots, x(M)\}$ composed of M bits, an m -palindrome of type
 111 $j \in \{T, R, I, G\}$ can be defined as the concatenation of two k -patterns: $\mathbf{X}_m(i) = [\mathbf{X}_k(i) \Gamma_j[\mathbf{X}_k(i)]]$ with
 112 $k = m/2$. The first k -pattern $\mathbf{X}_k(i) = \{x(i), x(i+1), \dots, x(i+k-1)\}$, $1 \leq i \leq M-k+1$ is the reference
 113 pattern, and the second k -pattern is the one obtained by one of the four isometries $\Gamma_j[\mathbf{X}_k(i)]$ with $j \in \{T, R, I, G\}$:

- 114 - $\Gamma_T[\mathbf{X}_k(i)] = \{x(i), x(i+1), \dots, x(i+k-1)\}$. A translation is simply a “copy and paste”;
- 115 - $\Gamma_R[\mathbf{X}_k(i)] = \{x(i+k-1), \dots, x(i+1), x(i)\}$. A vertical reflection is simply a “copy, return and paste”;
- 116 - $\Gamma_I[\mathbf{X}_k(i)] = \{x(i+k-1), \dots, x(i+1), x(i)\}$. An inversion is simply a
 117 “copy-complement-return-paste”;
- 118 - $\Gamma_G[\mathbf{X}_k(i)] = \{x(i), x(i+1), \dots, x(i+k-1)\}$. A glide reflection is simply a “copy-complement-paste”;

119 where \bullet is the logical function NOT, also called a complement. For example, with the binary sequence
 120 $\mathbf{X} = \{01010101\}$, the first 4-palindrome of type ‘T’ is written as $\mathbf{X}_4(1) = \{0101\}$, with $\mathbf{X}_2(1) = \{01\}$ and
 121 $\Gamma_T[\mathbf{X}_2(1)] = \Gamma_I[\mathbf{X}_2(1)] = \{01\}$.

122 If the objective is to measure the level of symmetry of a binary sequence through the presence of palindromes
 123 of type $j \in \{T, R, I, G\}$, then we can define the following measure:

$$124 \quad v_j^*(m) = \frac{v_j(m)}{2(M-m+1)(M-1)} \quad (5)$$

125 with $v_j(m)$ being the total number of palindromes of size m , $v_j(0) = M$ and the palindrome vector of type j by
 126 the following:

$$127 \quad \mathbf{V}_j^* = [v_j^*(0), v_j^*(1), \dots, v_j^*(m), \dots, v_j^*(M)]^t. \quad (6)$$

128 In order to propose a scalar measure of the level of symmetry of a given type, it seems judicious not to take
 129 into account the non-trivial palindromes because they could mask, for very long sequences, the presence of larger
 130 palindromes in smaller numbers. The total number of non-trivial palindromes σ_j^* of type $j \in \{T, R, I, G\}$, for
 131 the whole range of sizes m , is obtained by computing

$$132 \quad \sigma_j^* = \sum_{m=2}^M v_j^*(m). \quad (7)$$

133 To obtain the global level of symmetry present in a binary sequence, the global palindromic symmetry σ^*
 134 is defined as follows:

$$135 \quad \sigma^* = \sigma_T^* + \sigma_R^* + \sigma_I^* + \sigma_G^*, \quad (8)$$

136 where $\sigma_R^* = \bar{\sigma}$ is defined in Section 2. Note that, for binary sequences where the level of symmetry is the
 137 maximum as for example for the sequences $\mathbf{X} = \{01010101\}$ and $\mathbf{X} = \{111111\}$, the symmetry is maximum
 138 with $\sigma^* = 1$.

139 To quantify the “diversity” of different types of palindromes, the overall palindromic **symmentropy** \mathcal{E} can
 140 be defined as follows:

$$141 \quad \mathcal{E} = -\mathbf{P}^t \log_4 \mathbf{P}, \quad (9)$$

142 where \mathbf{P} is the *quarte* probability \mathbf{P} defined as follows:

$$143 \quad \mathbf{P} = [p_T, p_R, p_I, p_G]^t, \quad (10)$$

144 with $p_j = \sigma_j / \sigma^*$. Note that the values of the symmentropy are between 1/2 and 1. When there is equi-probability,
 145 then $\mathcal{E} = 1$. For example, for the sequence $\mathbf{X} = \{01010101\}$ of $M = 8$ bits, the symmentropy is maximal at $\mathcal{E} =$
 146 0.99 , and the value of $\mathcal{E} = \lim_{M \rightarrow \infty} 1$. When two probabilities out of four are null with $\mathbf{P} = [1/2, 1/2, 0, 0]^t$,
 147 as is the case for the 8-bit sequence $\mathbf{X} = \{11111111\}$, then the symmentropy is minimal and is $\mathcal{E} = 1/2$. This
 148 means that, when the symmentropy is minimal, there is always a minimum symmetric information content in the
 149 binary sequences.

150 Finally, it seems appropriate to compute a local palindromic symmentropy $\epsilon(m)$ for each m scale:

$$151 \quad \epsilon(m) = -\mathbf{Q}^t(m) \log_4 \mathbf{Q}(m), \quad (11)$$

152 where $\mathbf{Q}(m) = [q_T(m), q_R(m), q_I(m), q_G(m)]^t$ is the *quarte* probability at scale m , where $q_j(m) = \frac{v_j(m)}{\sigma(m)}$ and
 153 with $\sigma(m) = v_T(m) + v_R(m) + v_I(m) + v_G(m)$.

154 To illustrate, let us consider the binary sequence $\mathbf{X} = \{01101001\}$ of 8 bits. We reported in Table 2 $Dict_j$,
 155 $v_j(m)$, $v_j^*(m)$ and q_j with $j \in \{T, R, I, G\}$.

156 Remark: This measure of symmentropy is similar in idea to the one proposed by Yodogawa [16], who
 157 proposed an entropic measure of the level of symmetry present in the images via a decomposition in the
 158 Walsh–Hadamard basis. (The method of Yodogawa that measures the entropy of symmetric patterns is called
 159 symmetry. From our point of view, it is rather a symmentropy since it is derived from an entropy measure,
 160 which is not the case of symmetry as we define it in Section 3. On the other hand, in Yodogawa’s approach,
 161 the probabilities allowing us to computation the entropy in base 2 are obtained from a decomposition in the
 162 Walsh–Hadamard basis. In Yodogawa’s paper, it is clearly stated that not all symmetries are considered, which is
 163 not the case for our approach based on symmetry friezes.) Here, the proposed definition is different.

Table 2. $Dict_j$, $v_j(m)$, q_j and $\epsilon(m)$ with $j \in \{T, R, I, G\}$ computed from the 8 binary sequence $\mathbf{X} = \{01101001\}$. The non-trivial palindromic symmetry is $\sigma^* = 0.44 = 1302/2940$ with $\sigma_T^* = 102/2940$, $\sigma_R^* = 214/2940$, $\sigma_I^* = 472/2940$ and $\sigma_G^* = 514/2940$, and the global palindromic symmetry is $\mathcal{E} = 0.89 = -\left(\frac{102}{1302} \log_4\left(\frac{102}{1302}\right) + \frac{214}{1302} \log_4\left(\frac{214}{1302}\right) + \frac{472}{1302} \log_4\left(\frac{472}{1302}\right) + \frac{514}{1302} \log_4\left(\frac{514}{1302}\right)\right)$.

m	0	1	2	3	4	5	6	7	8
$Dict_T$	e	0,1	00,11	-	1010	-	-	-	-
$v_T(m)$	8	8	2	0	1	0	0	0	0
$v_T^*(m)$	-	-	$\frac{2}{2 \times 7 \times 7}$	$\frac{0}{2 \times 6 \times 7}$	$\frac{1}{2 \times 5 \times 7}$	$\frac{0}{2 \times 4 \times 7}$	$\frac{0}{2 \times 3 \times 7}$	$\frac{0}{2 \times 2 \times 7}$	$\frac{0}{2 \times 1 \times 7}$
$q_T(m)$	-	-	2/14	0	1/6	-	0	-	0
$Dict_R$	e	0,1	00,11	101,010	0110,1001	-	-	-	-
$v_R(m)$	8	8	2	2	2	0	0	0	0
$v_R^*(m)$	-	-	$\frac{2}{2 \times 7 \times 7}$	$\frac{2}{2 \times 6 \times 7}$	$\frac{2}{2 \times 5 \times 7}$	$\frac{0}{2 \times 4 \times 7}$	$\frac{0}{2 \times 3 \times 7}$	$\frac{0}{2 \times 2 \times 7}$	$\frac{0}{2 \times 1 \times 7}$
$q_R(m)$	-	-	2/14	1/2	2/6	-	0	-	0
$Dict_I$	e	0,1	01,10	-	1010	-	110100	-	01101001
$v_I(m)$	8	8	5	0	1	0	1	0	1
$v_I^*(m)$	-	-	$\frac{5}{2 \times 7 \times 7}$	$\frac{0}{2 \times 6 \times 7}$	$\frac{1}{2 \times 5 \times 7}$	$\frac{0}{2 \times 4 \times 7}$	$\frac{1}{2 \times 3 \times 7}$	$\frac{0}{2 \times 2 \times 7}$	$\frac{1}{2 \times 1 \times 7}$
$q_I(m)$	-	-	5/14	0	1/6	-	1	0	1/2
$Dict_G$	e	0,1	01,10	010,101	0110,1001	-	-	-	01101001
$v_G(m)$	8	8	5	2	2	0	0	0	1
$v_G^*(m)$	-	-	$\frac{5}{2 \times 7 \times 7}$	$\frac{2}{2 \times 6 \times 7}$	$\frac{2}{2 \times 5 \times 7}$	$\frac{0}{2 \times 4 \times 7}$	$\frac{0}{2 \times 3 \times 7}$	$\frac{0}{2 \times 2 \times 7}$	$\frac{1}{2 \times 1 \times 7}$
$q_G(m)$	-	-	5/14	1/2	2/6	-	0	-	1/2
$\epsilon(m)$	-	-	0.93	0.50	0.96	-	0	-	0.50

164 4. Results

165 In this section, we wish to show the interest of these new scalar and vector descriptors in the study of
 166 binarized sequences. We propose to compute the different proposed descriptors (palindromic vectors, symmetry
 167 and symmetry) for binarized sequences taken from $1/f$ noises and 2 DNA sequences.

168 4.1. Binarized $1/f$ Noise

169 One way to study complexity, in which the meaning here is reduced to that of irregularity as reported in
 170 [17], is to vary the exponent β of the noise in f^β . For $\beta = 0$, the generated noise is white noise, and for $\beta = -2$,
 171 the generated noise is a Brownian motion, with the integral of a white noise being a Brownian motion.

172 Here, in order to stay within the framework of our study, the time series are binarized. All values above
 173 the median are replaced by '1', otherwise '0'. Moreover, in order to compare the different scalar and vector
 174 descriptors, the Lempel–Ziv complexity \mathcal{C}_{Lz} is proposed as a reference and is computed as presented in [18]. This
 175 normalized complexity is almost zero for periodic binary sequences and close to unity for random sequences
 176 such as white noise.

177 In Figures 2–4, the scalar and vector descriptors obtained for noises in f^β with $\beta \in \{-2 : 0\}$ by step of 0.2
 178 are presented. For a same value of β , 300 binarized noises composed of 1000 bits are generated.

179 In Figure 2, the different scalar palindromic descriptors are computed and plotted as whisker boxes. From
 180 Figure 2, we observe that all scalar palindromic descriptors describe monotonic curves increasing for Lempel–Ziv
 181 complexity and symmetry and decreasing for symmetry (as well as these components through the *quarte*
 182 probability \mathbf{P}). This monotonicity property can be auspicious for tracing the values of β knowing the value of the
 183 descriptor. Indeed, it is possible to discriminate binarized noises in f^β on larger or smaller regions depending on
 184 the descriptor considered. For example, for the Lempel–Ziv complexity, the body (second and third quartile) of
 185 the non-overlapping whisker boxes in the region $-1.2 < \beta < -0.6$ allows us, from a Lempel–Ziv complexity
 186 of 0.62, to go back to a value of $\beta = -1.2$ without much error. When $\beta = 0$, the complexity is maximal
 187 and tends to unity; when $\beta = -2$, the complexity is less and is 0.2 for a Brownian motion. For symmetry,
 188 the non-overlapping boxes for $-1.2 < \beta < -0.4$ also allows us to find the value of β from the symmetry
 189 measures. Note that the discrimination range ($\beta > -1.2$) of symmetry is much larger than those obtained by

190 Lempel–Ziv complexity and symmetry. We also check that the values of the symmentropy are well between
 191 $1/2$ and 1 . Finally in Figure 2, we observe a decrease in the probabilities drawn from the *quarte* \mathbf{P} . Indeed,
 192 it decreases as β approaches zero for types ‘T’ and ‘R’ to go from 50% to 25% and 33%, respectively, and it
 193 increases progressively for types ‘I’ and ‘G’ to go from 0% to 17% and 25%, respectively. At the maximum
 194 complexity $\beta = 0$, we observe that the reflection symmetry level is always higher than the translation/glide
 195 reflection and inversion: $\sigma_R^* > \sigma_T^* = \sigma_G^* > \sigma_I^*$.

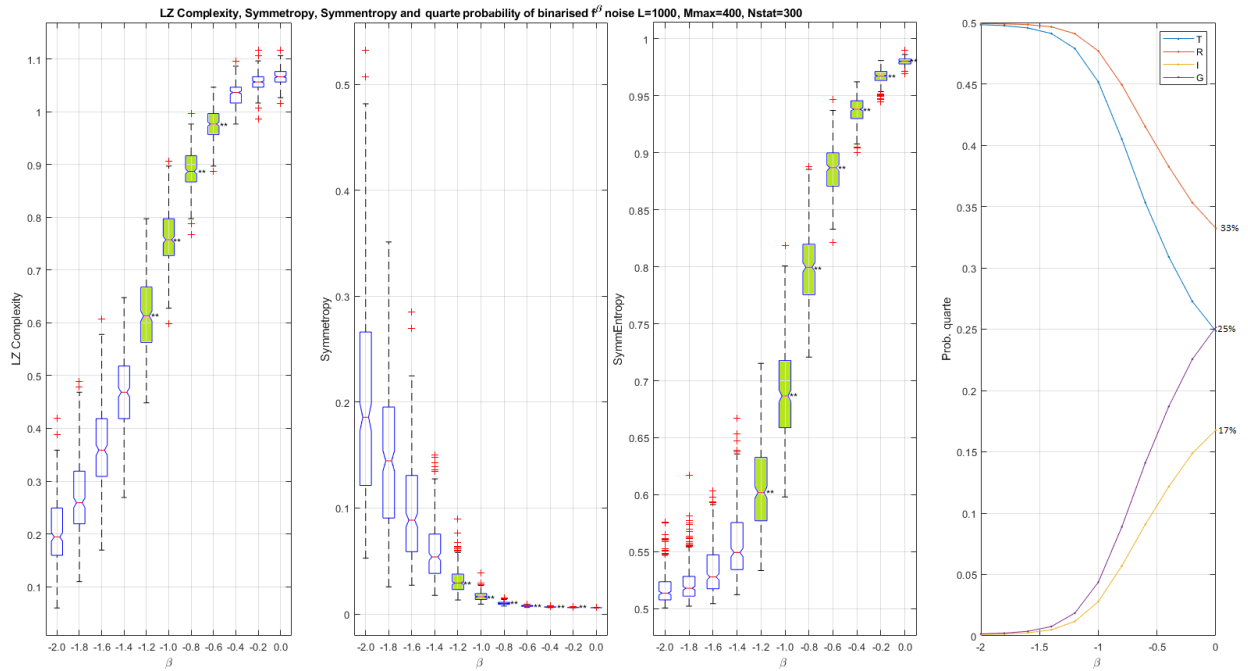


Figure 2. Scalar palindromic descriptors obtained from binarized f^β noises and for different values of β . Left, Lempel–Ziv complexity C_{Lz} , in which the boxes do not overlap for $-1.2 < \beta < -0.6$. Left middle, symmetry σ^* , in which the boxes do not overlap for $-1.2 < \beta < -0.4$. Right middle, symmentropy \mathcal{E} , in which boxes do not overlap for $\beta > -1.2$. Left, *quarte* probability \mathbf{P} versus β . When $\beta = -2$, the *quarte* probability is $\mathbf{P} = [0.50, 0.50, 0.00, 0.00]$. When $\beta = 0$, the *quarte* probability is $\mathbf{P} = [0.25, 0.33, 0.17, 0.25]$ and the level of reflection symmetry is higher than the translation/ glide reflection and the inversion: $\sigma_R^* > \sigma_T^* = \sigma_G^* > \sigma_I^*$. The closer β is to zero, the higher the complexity. We notice that both C_{Lz} and \mathcal{E} increase as the complexity increases. On the contrary σ^* decreases as the complexity increases.

196 In Figure 3, the palindromic vectors obtained for $\beta = -2, -1, 0$, which correspond to Brownian motion,
 197 pink noise and white noise, respectively, are presented. From Figure 3, we observe that all of the average
 198 palindromic vectors (obtained by averaging 300 palindromic vectors) decrease as the palindromic size m
 199 increases and this decrease is all the more marked as β approaches zero, i.e., when the correlations between
 200 samples are almost non-existent. Note that, for Brownian motions ($\beta = -2$), there are large palindromes up to
 201 about 450. On the contrary, for white noise, we note that the size of the palindromes does not exceed 20 bits.
 202 Moreover, the palindromic vector obtained for $\beta = 0$ is very similar to the one obtained in the case of binary iid
 203 (independent and identically distributed) sequences, as shown in Figure 5.

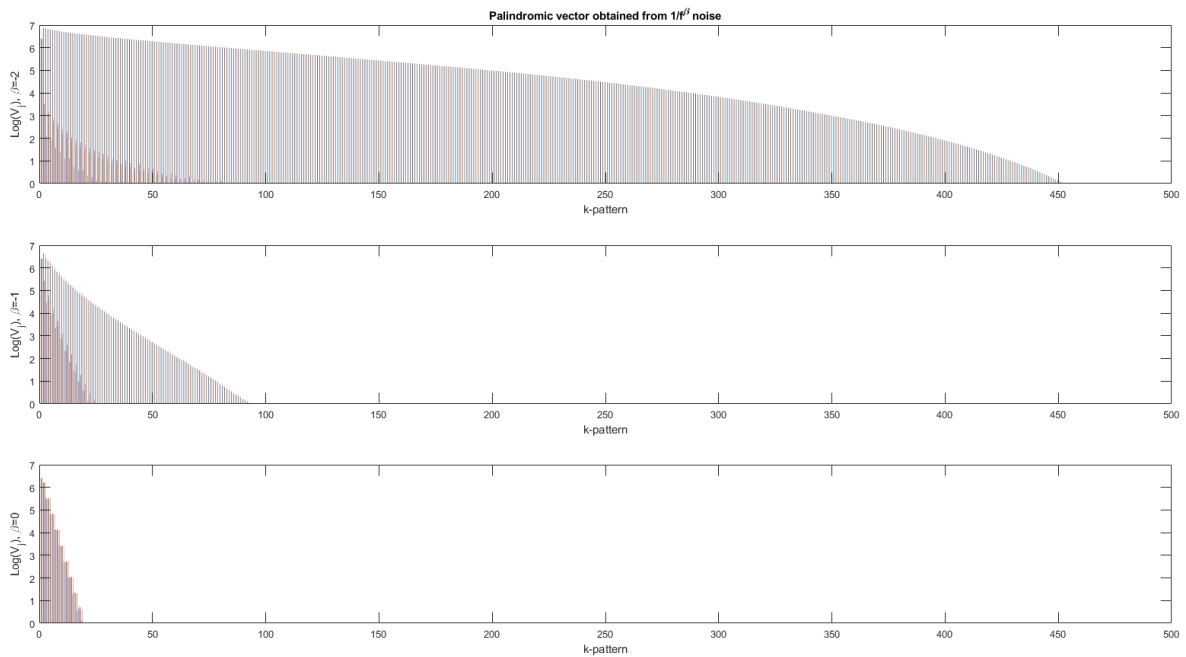


Figure 3. Average palindromic vectors obtained for binarized f^β noises of length 1000, with $\beta = -2.0, -1.0, -0.0$ (from top to bottom) and $m \in \{1 : 500\}$. Top, average palindromic vectors obtained after averaging 300 vectors for $\beta = -2.0$. Middle, average palindromic vectors obtained after averaging 300 vectors for $\beta = -1.0$. Bottom, average palindromic vectors obtained after averaging 300 vectors for $\beta = 0.0$. The more irregular the sequence (strong negative value of β) and the larger the spread of the palindromic vector descriptors.

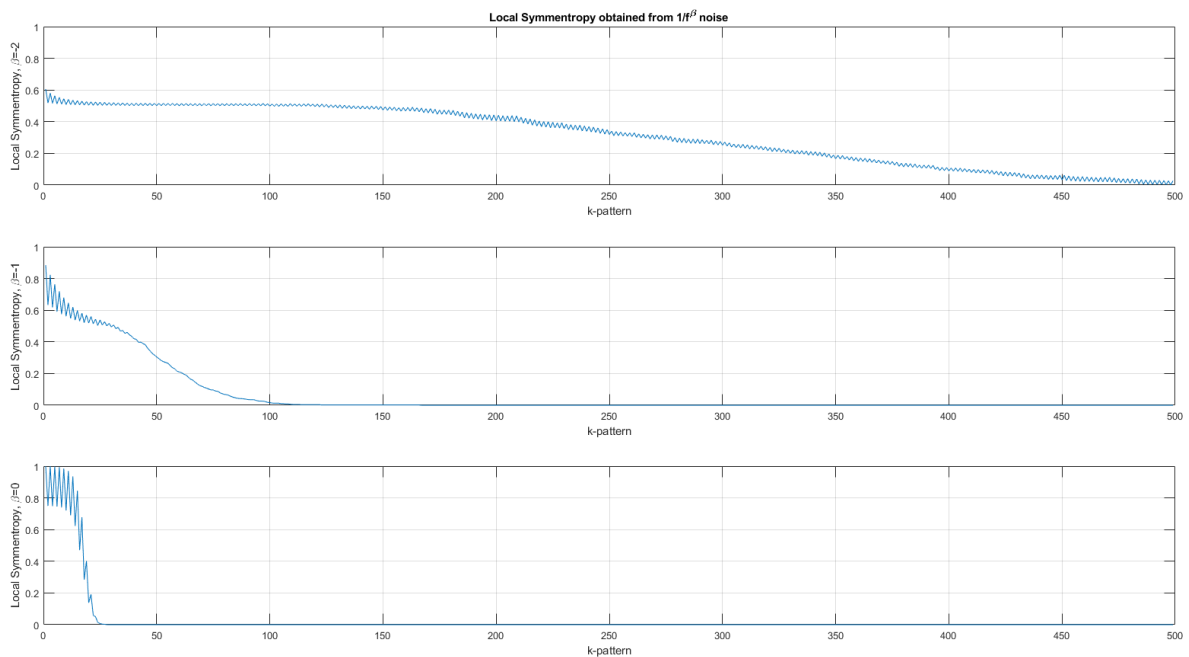


Figure 4. Average local symmentropy (with 300 trials) computed for three types of noises. Top, local symmentropy of a Brownian motion ($\beta = -2$). Middle, local symmentropy of a pink noise ($\beta = -1$). Bottom, local symmentropy of a white noise ($\beta = 0$). The sawtooth fluctuation comes from the fact that the symmentropy values are slightly different for even and odd palindromes. The local symmentropy synthesizes the information carried by the four palindromic vectors into only one.

204 In Figure 4, the local symmentropy (averaged from 300 trials) $\epsilon(m)$ computed for three different types of
 205 noise (Brownian motion, pink noise and white noise) is plotted. As for the palindromic vectors, the symmentropy

206 decreases as the size of the palindromes increases. The spread out of the symmentropy depends on the type of
 207 noise and thus on the correlations between samples. The range in size is very small for white noise with no
 208 correlation between samples/bits compared with Brownian motion. Moreover, the value of the symmetry is
 209 close to unity for the white noise and close to half for the Brownian motion.

210 In Figure 5, the palindromic vectors obtained from binary sequences independent and identically distributed
 211 are plotted. We observe in Figure 5 a different distribution between even and odd palindromes. There is an
 212 equi-distribution between the different types of symmetry for the even palindromes. For odd palindromes, we
 213 also note the non-presence of palindromes of type 'I'. Note that there are no palindromes with sizes exceeding 40
 214 bits. On average, the proportion of palindromes is $P_T = 25\%$, $P_R = 33\%$, $P_I = 17\%$, $P_G = 25\%$. We notice a
 215 decrease in the symmetry levels as the size of the palindromes increases. In logarithmic scale, the decrease in the
 216 symmetry level (and thus of the number of symmetrical palindromes) is linear. Indeed, for a fixed length of the
 217 binary sequence, the more the palindrome size increases, the smaller the number of palindromes composing the
 218 binary sequence. For example, a sequence of 8 bits can only be composed of one palindrome of size $m = 8$,
 219 of two palindromes of size $m = 4$, of four palindromes of size $m = 2$ and of eight palindromes of size $m = 1$.
 220 This decrease is therefore inversely proportional to the size m . If we suppose that, for a given type of symmetry,
 221 the palindrome vector is expressed by $V_j(m) = K_j/m$, then $\log(V_j(m)) = -1 \times \log(m) + \log(K_j)$. This is
 222 indeed the affine line observed in Figure 5.

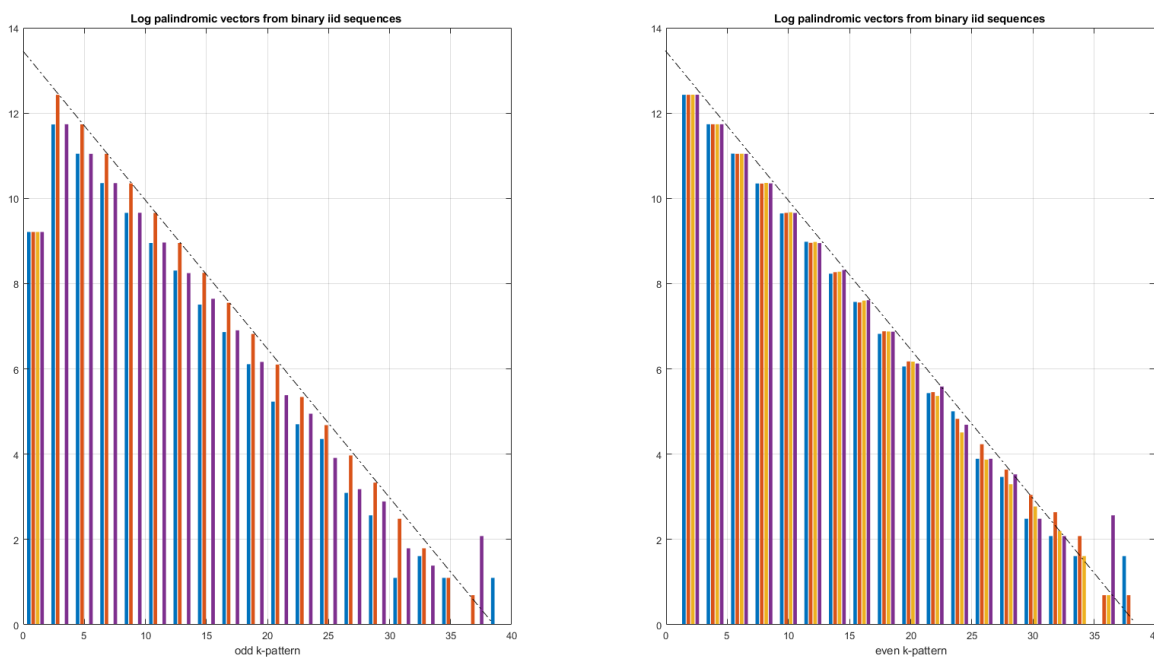


Figure 5. Logarithm of the four average palindromic vectors computed from 100 binary sequences iid (independent and identically distributed) of 5000 bits. We note a different distribution between even and odd palindromes. There is an equi-distribution between the different types of symmetry for even palindromes. For odd palindromes, we also note the non-presence of palindromes of 'I' type. We note a decrease in the symmetry levels as the size of the palindromes increases. Note that there are no palindromes with sizes exceeding 40. Finally, on average, the proportion of palindromes is $P_T = 25\%$, $P_R = 33\%$, $P_I = 17\%$ and $P_G = 25\%$.

223 4.2. Biological Sequences: DNA

224 To show the relevance of the different symmetry descriptors proposed in a practical case, let us consider
 225 two DNA sequences. The objective is to identify descriptors that allow us to differentiate the two sequences:
 226 HUMHBB (human β -region, chromosome 11) with 73308 bases and YEAST1 (Saccharomyces cerevisiae yeast,
 227 chromosome 1) with 230209 bases obtained from (<http://ncbi.nlm.nih.gov> (accessed on 30 December 2021)).
 228 The DNA sequences is binarized, 'A' and 'G' are coded by 1, and 'T' and 'C' are coded by 0. For example, the
 229 sequence 'ATATGCATTTCC...' is coded '101010100000'.

At first, it seems interesting to indicate that, although the sequence “YEAST1” is 3.14 larger than the sequence “HUMHBB”, the total number of palindromes coming from the sequence “YEAST1” is 2.95 times larger than that of the sequence “HUMHBB”, as indicated in Table 3.

Table 3. Distribution in % of the total number of palindromes of different types present in each of the two non-randomized and randomized DNA sequences, $m \in [1, 500]$. For the non-randomized sequences, the most frequent palindromes are reflection palindromes with $N_R > N_T > N_G > N_I$, while for the randomized sequences, the distribution is $N_R > N_T = N_G > N_I$. The distribution of the different types of palindromes is very similar regardless of the type of DNA sequence. The differences between the total number of palindromes from non-randomized and randomized HUMHBB and YEAST1 sequences are $496,028 - 441,299 = 54,729$ and $1,463,633 - 1,384,396 = 79,237$, respectively.

DNA seq	N_T/N_{Total}	N_R/N_{Total}	N_I/N_{Total}	N_G/N_{Total}	N_{Total}
HUMHBB	29.5%	36.8%	13.5%	20.2%	496,028
randomized HUMHBB	24.9%	33.3%	16.7%	25.1%	441,299
Yeast1	27.9%	35.5%	14.5%	22.1%	1 463 633
randomized Yeast1	25.0%	33.3%	16.7%	25.0%	1,384,396

Moreover, we notice in Table 3 that the proportion of palindromes of type “mirror” (i.e., ‘R’ type) is much higher than that in the other types regardless of the DNA sequence considered. This corroborates what has been observed for $1/f$ noises, namely $P_R > P_T > P_G > p_I$, where P_j is the palindromic probability of type j .

In Table 4, the Lempel–Ziv complexity C_{Lz} , the symmentropy \mathcal{E} and the symmetry σ^* are reported. From Table 4, we notice that the scalar descriptors are slightly different for the 2 DNA sequences. We note a relative difference of 4% for the Lempel–Ziv complexity ($4\% = (0.98 - 0.94)/0.94$), of 1% for the symmentropy ($1\% = (0.97 - 0.96)/0.96$) and of 6% for the symmetry ($6\% = (0.85 - 0.80)/0.80$)

Table 4. Scalar palindromic descriptors of binarized DNA sequences. Lempel–Ziv complexity C_{Lz} , symmentropy \mathcal{E} and symmetry σ^* with $m \in \{0, 500\}$. From scalar palindromic descriptors, it seems possible to differentiate the 2 DNA sequences. The values of Lempel–Ziv complexity and symmentropy are close to unity, indicating a high level of complexity. For randomized DNA sequences, Lempel–Ziv complexity and symmentropy tend toward unity.

DNA seq	C_{Lz}	\mathcal{E}	$100 \times \sigma^*$
HUMHBB	0.94	0.96	0.85
randomized HUMHBB	1.02	0.98	0.75
Yeast1	0.98	0.97	0.80
randomized Yeast1	1.01	0.98	0.75

To go further in the analysis of DNA sequences, in Figure 6, the palindromic vector descriptors for each type j for $m \in (0, 100)$ are reported, even if the calculation has been made with $m_{max} = 500$. We notice that the palindromic vectors are rather concentrated in the 0–100 band with some peaks (not shown here) beyond $m = 100$ located in $m = 270, 192$ for “YEAST1” and $m = 124$ for “HUMHBB”. As for the noises in $1/f$, we notice a different distribution of the types of palindromes. For example, there are no more palindromes of type ‘R’ for the sequence “YEAST1” beyond $m = 60$, idem for the palindromes of type ‘I’ for the sequence “YEAST1” beyond $m = 40$. By the way, note that there are no even palindromes of type ‘I’. This shows the importance of taking into account all types of palindromes and not only the “mirror” palindromes of type ‘R’. By superimposing the palindromic vectors obtained after randomization, we can better see the “useful” information. The signature after randomization being similar to that of an independent and identically distributed random variable seems to be less important information and therefore useless for DNA sequence discrimination.

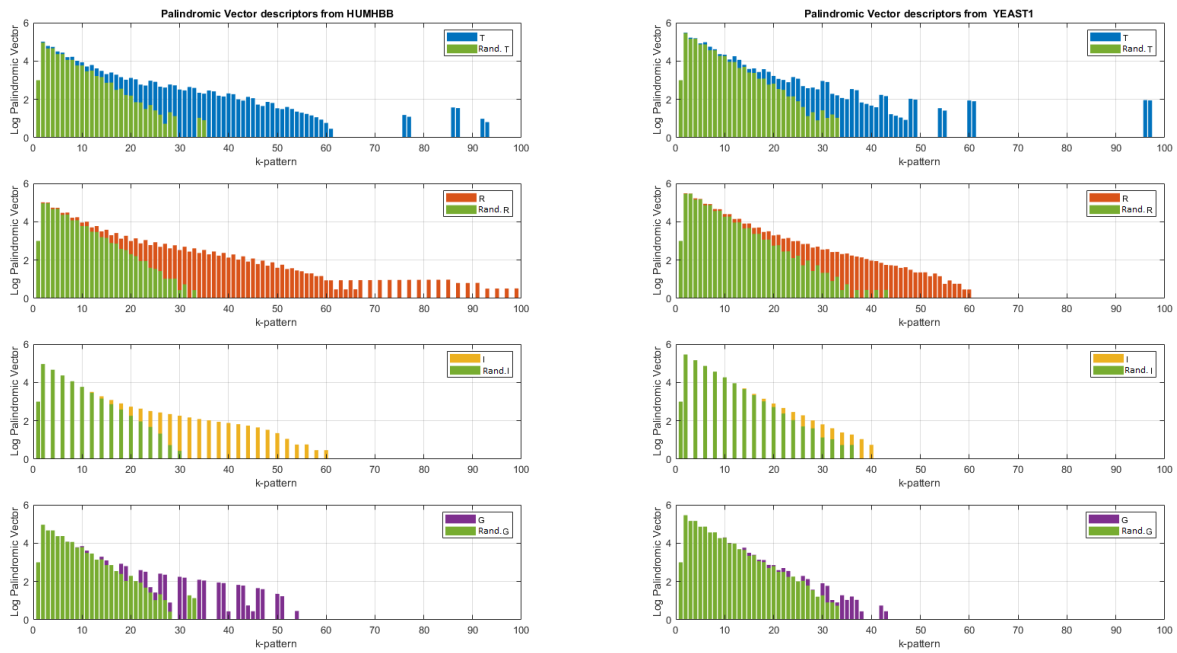


Figure 6. Logarithm of the palindromic vectors obtained from the entirety of the two DNA sequences for $m_{max} = 500$. Zoom for $m \in (1, 100)$. In green, logarithm of the palindromic vectors obtained after randomization of the DNA sequences.

251 Finally, it seems interesting to show how local symmentropies allow us to differentiate each DNA sequence.
 252 In Figure 7, the local symmentropies calculated from “HUMHBB”, randomized “HUMHBB”, “YEAST1” and
 253 randomized “YEAST1” are reported. Straight lines derived from linear fitting from symmentropies show slopes
 254 that are significantly different between each DNA sequence. Indeed, from odd palindromes, the slope derived
 255 from the linear fitting for YEAST1 is 4.41 times the slope obtained from HUMHBB. For even palindromes, the
 256 slope derived from the linear fitting for YEAST1 is 3.61 times the slope obtained from HUMHBB. As expected,
 257 symmentropies obtained from randomized DNA sequences are similar while $m < 20$ and close to unity. Indeed,
 258 the binary sequence obtained after randomization is very similar to an independent and identically distributed
 259 random variable for which the symmentropy is maximal and worth unity. For $m > 30$, as shown in Figure 7, the
 260 symmentropies between the 2 DNA sequences are different.

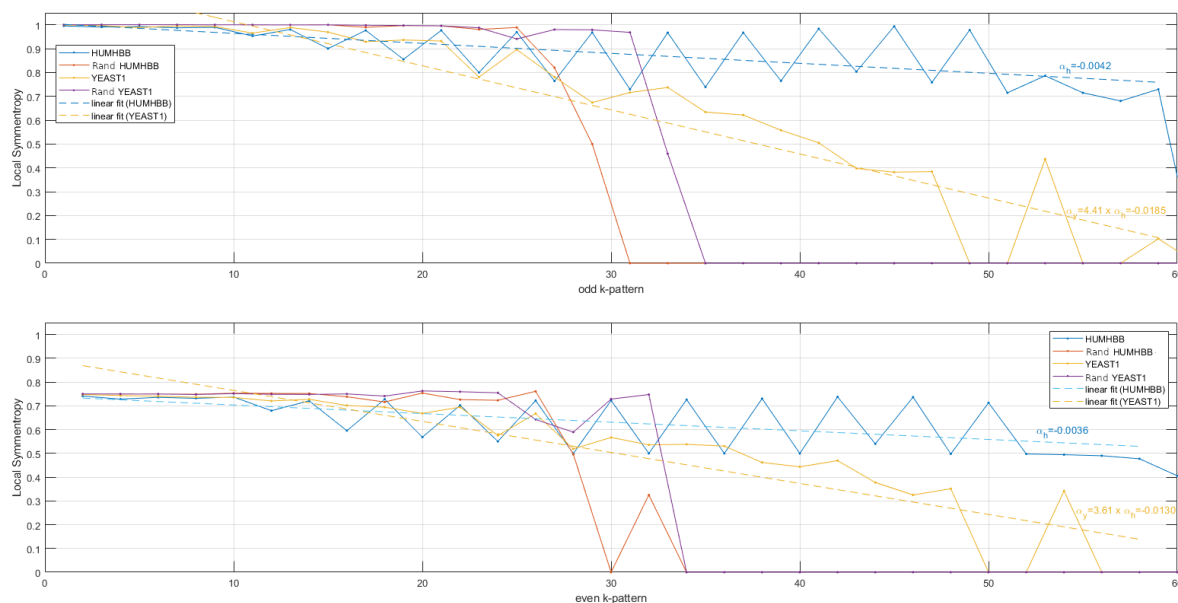


Figure 7. Local symmetries obtained from binarized DNA sequences in the scale range $m \in \{1, 60\}$. Top, odd palindromes. In blue, local symmetry obtained from HUMHBB and straight line fitting. In orange, local symmetry obtained from YEAST1 and straight line fitting. In magenta, local symmetry obtained from randomized HUMHBB. In red, local symmetry obtained from randomized Yeast. The slope α_y derived from the linear fitting for YEAST1 is 4.41 times the slope α_h obtained from HUMHBB. Bottom, even palindromes. The slope α_y derived from the linear fitting for YEAST1 is 3.61 times the slope α_h obtained from HUMHBB.

261 5. Discussion and Conclusions

262 In this work, we proposed new palindromic descriptors (scalar and vector). The notions of palindromic
 263 vectors, palindromic symmetry and palindromic symmetry have been tested with binarized $1/f$ noises
 264 and 2 DNA sequences. For f^β noises for which the “complexity” level is adjustable via β , we showed that
 265 palindromic symmetry as well as palindromic symmetry allows us to better discriminate the different
 266 f^β noises on a larger range than the Lempel–Ziv complexity. Moreover, we showed that symmetry is
 267 a complexity descriptor very similar to the Lempel–Ziv complexity. However, the palindromic symmetry
 268 indicates the level of symmetry and is a descriptor of “anti-complexity”.

269 From this preliminary study, we notice that the “mirror” symmetry is more present than the other types of
 270 symmetries regardless of the level of complexity (see Figure 2). This is probably why only the “mirror” symmetry
 271 through the classical notion of palindrome has been considered so far. However, we showed (see Figure 6) that
 272 the four types of palindromes are necessary to better discriminate the binary sequences. Moreover, we showed
 273 that the distribution of the types of palindrome evolves with complexity. It goes from 50% for ‘T’ and ‘R’ types
 274 and 0% for ‘I’ and ‘G’ types when the complexity is low to 25%, 33%, 17% and 25% for ‘T’, ‘R’, ‘I’ and ‘G’
 275 types when the complexity is maximal. These values are found when the binarized DNA sequences have been
 276 randomized.

277 Multiscale palindromic exploration, i.e., for the whole m size range of palindromes, through palindromic
 278 vectors and local symmetry, allows us to go further in the analysis of binary sequences. In particular, it allows
 279 us to highlight a particular signature of independent and identically distributed random binary sequences found
 280 for white noise ($\beta = 0$) and in the two DNA sequences. This exploration also allows us to clearly identify regions
 281 that allow us to discriminate the two DNA sequences. Furthermore, a factor of 4 between the slopes of the linear
 282 fits of the local symmetries calculated from the two DNA sequences shows the discriminative capacity of the
 283 local symmetry.

284 It seems obvious, as in the article by Tibatán and Sarisaman [9], that symmetry properties, insufficiently
 285 exploited to date, play a more important role in the exploration of biological sequences, both at the molecular
 286 and sub-molecular levels. The new palindromic descriptors presented in this work should contribute in a
 287 non-negligible way and should be widely applied in the study of biological sequences.

288 **Author Contributions:** Writing—original draft, J.-M.G. and S.M. All authors have read and agreed to the published version
289 of the manuscript..

290 **Funding:** This research received no external funding.

291 **Conflicts of Interest:** The authors declare no conflict of interest..

292 References

- 293 1. Berry, M.; Nunez, A.M.; Chambon, P. Estrogen-responsive element of the human pS2 gene is an imperfectly palindromic
294 sequence. *Proc. Natl. Acad. Sci. USA* **1989**, *86*, 1218–1222.
- 295 2. Ohno, S. Intrinsic evolution of proteins. The role of peptidic palindromes. *Riv. Biol.*, **1990**, *83*, 405–410.
- 296 3. Cain, D.; Erlwein, O.; Grigg, A.; Russell, R.A.; McClure, M.O. Palindromic sequence plays a critical role in human
297 foamy virus dimerization. *J. Virol.* **2001**, *75*, 3731–3739.
- 298 4. Giel-Pietraszuk, M.; Hoffmann, M.; Dolecka, S.; Rychlewski, J.; Barciszewsk, J. Palindromes in Proteins. *J. Protein*
299 *Chem.* **2003**, 109–113.
- 300 5. Lisnic, B.; Svetec, I.-K.; Saric, H.; Nikolic, I.; Zgaga, Z. Palindrome content of the yeast *Saccharomyces cerevisiae*
301 genome. *Curr. Genet.* **2005**, *47*, 289–297.
- 302 6. Pinotsis, N.; Wilmanns, M. Protein assemblies with palindromic structure motifs. *Cell. Mol. Life Sci.* **2008**, *65*,
303 2953–2956.
- 304 7. Lamprea-Burgunder, E.; Ludin, P.; Mäser, P. Species-specific Typing of DNA Based on Palindrome Frequency Patterns.
305 *DNA Res.* **2011**, *18*, 117–124.
- 306 8. Raykov, V.; Marvin, M.E.; Louis, E.J.; Maringele, L. Telomere dysfunction palindrome formation independently of
307 double-strand break repair mechanisms. *Genetics* **2016**, *203*, 1659–1668.
- 308 9. Tibatan, M.A.; Sarisaman, M. Unitary structure of palindromes in DNA. *BioSystems* **2022**, *211*, 104565.
- 309 10. Gabric, D.; Shallit, J. Borders, palindrome prefixes, and square prefixes. *Inf. Process. Lett.* **2021**, *165*, 106027.
- 310 11. Allouche, J.-P. Sur la complexité des suites infinies. *Bull. Belg. Math.* **1994**, *1*, 133–143.
- 311 12. Allouche, J.-P.; Baake, M.; Cassaigne, J.; Damanik, D. Palindrome complexity. *Theor. Comput. Sci.* **2003**, *292*, 9–31.
- 312 13. Brlek, S.; Reutenauer, C. Complexity and palindromic defect of infinite words. *Theor. Comput. Sci.* **2011**, *412*,
313 493–497.
- 314 14. Cederberg, J.N. *A Course in Modern Geometries*; Springer: New York, NY, USA, 2001.
- 315 15. Grunbaum, B.; Shephard, G.C. *Tilings and Patterns*; W.H. Freeman and Company: New York, NY, USA, 1989.
- 316 16. Yodogawa, E. Symmetry an entropy-like measure of visual symmetry. *Percept. Psychophys.* **1982**, *32*, 230–240.
- 317 17. Girault, J.-M.; Humeau-Heurtier, A. Centered and Averaged Fuzzy Entropy to Improve Fuzzy Entropy Precision.
318 *Entropy*, **2018**, *20*, 287. <https://doi.org/10.3390/e20040287>.
- 319 18. Kaspar, F.; Schuster, H.G. Easily Calculable Measure for the Complexity of Spatiotemporal Patterns. *Phys. Rev. A* **1987**,
320 *36*, 843–848.