

Helitron's Periodicities Identification in C.Elegans based on the Smoothed Spectral Analysis and the Frequency Chaos Game Signal Coding

Rabeb Touati¹, Imen Messaoudi², Afef Elloumi Ouesleti³, Zied Lachiri⁴

SITI Laboratory, National School of Engineers of Tunis (ENIT), University of Tunis El Manar, Tunisia
Tunis, Tunisia

Abstract—Helitrons are typical rolling circle transposons which make up the bulk of eukaryotic genomes. Unlike of other DNA transposons, these transposable elements (TEs), don't create target site duplications or end in inverted repeats, which make them particular challenge to identify and more difficult to annotate. To date, these elements are not well studied; they only attracted the interest of researchers in biology. The focus of this paper is oriented towards identifying the helitrons in C.elegans genome in the perspective of signal processing. Aiming at the helitron's identification, a novel methodology including two steps is proposed: the coding and the spectral analysis. The first step consists in converting DNA into a 1-D signal based on the statistical features of the sequence (FCGS coding). As for the second step, it aims to identify the global periodicities in helitrons using the Smoothed Fourier Transform. The resulting spectrum and spectrogram are shown to present a specific signature of each helitron's type.

Keywords—Helitrons; C.elegans; Frequency Chaos Game Signal (FCGS) coding; spectral analysis; tandem periodicities

I. INTRODUCTION

The helitrons are a distinguished type of the transposable elements (TEs) DNA which transposes by a rolling circle replication mechanism. Due to their ability to move rapidly and replicate within genomes, helitrons play a major role in genomes evolution. In fact, by transferring a DNA segment from one genomic site to another, these elements are responsible for intragenomic multiplication [1]. Many types of genetic variation caused by TEs in animals and plants are described in [2].

The helitrons are part of the transposable elements (TEs) class 2 [3]. They were discovered by in silico genome-sequence analysis. Newly discovered in all eukaryotic genomes [4]-[7], helitrons have shown a remarkable ability to capture gene sequences [8], [9]. In fact, on a large scale, these elements are suggested to duplicate and shuffle exon domains [10]-[12].

Helitrons are present in various organisms: like plants (such as corn, maize, rice, Arabidopsis Thaliana) [13], nematodes (such as the worm *Cænorhabditis elegans*) [14], fungus [15], [16] and animals [17] (such as lucifugus [8] and specifically in the vertebrates genomes like the fishes *Danio rerio* and *Sphoeroides nephelus* [16]). Approximately, the helitron DNA constitutes at most 2% of the A. thaliana, the C.elegans and the maize genomes [12], [14], [18] and 4.23% of the silkworm

genome [19]. These elements are mostly represented by non-autonomous elements.

Since their discovery, helitrons have attracted widespread attention. Many computational tools were developed to identify and analyze the helitron in genomes from which are: HelitronFinder [20], HelSearch [21], a combination of BLAST search and hidden Markov models [22] and HelitronScanner [23].

HelitronFinder and HelSearch are similar [24], [25]; they are based on the conserved sequences at the termini 5'-TC and CTAG-3'(R = A or G) of most Helitrons. Both programs look for the hairpin structure and the CTRR 3' terminus.

The users of HelSearch have to manually search for the 5'end of Helitrons whereas users of HelitronFinder can identify the 5'end automatically. The combination of BLAST search and hidden Markov models is a limited method to identify helitrons with more diverse termini. The HelitronScanner, a two-layered local combinational variable (LCV), is a tool that identifies helitrons missed by both HelSearch and HelitronFinder [23]. This tool aims to extract helitron features like the hairpin structure, CTRR at the 3' end and the TC dinucleotide at the 5' end, and the A and T residues flanking the 5' and 3' ends, respectively. In an automated way, this tool uncovers many new helitrons which were missed by other tools. But this method presents also limitations in finding helitrons which don't have hairpins. Given that these transposable elements lack typical transposon features, the helitron's investigation is limited. The automated identification and localization of helitrons remain purely based on previously known sequences.

Helitrons elements are widespread and highly heterogeneous which makes their identification a difficult task. Here, the novelty of this work consists in the use of signal processing techniques to search helitrons in a large genomic database without prior knowledge about the content of the DNA sequences. The idea consists in finding a way to identify each helitron's type based on a signature that characterizes it.

To apply the signal processing methods on biological sequences, it is required to convert DNA into a digital signal known as DNA coding. The numerical representation of DNA using coding techniques is important since it plays a great role in visualizing, characterizing and highlighting the information contained in it. Different coding techniques exist: the binary

coding [26], the structural bending trinucleotide coding (PNUC) [27], the electron-ion interaction pseudo-potential (EIIP) mapping [28], the Frequency Chaos Game Signal [29]-[31], etc.

In the signal processing field, several techniques were applied with success to detect some biological sequences. For example, genes were segmented into coding and non-coding regions using the windowed Fourier Transform and based on the 3-bp periodicity that characterizes exon [32], [33]. In addition, the Fast Fourier Transform (FFT) was used to reflect the correlation properties of the coding and non-coding DNA sequences [32]. In [34], another technique of analysis is used to detect the short exons: the Fourier transform infrared spectroscopy. The latent periodicities in different genome elements including exons and microsatellite DNA sequences were detected and used the Fourier transform [35]. In addition, the modified Gabor-wavelet transform was used to identify the protein coding regions [36]. The Auto-Regressive technique was also used to predict genes and exon location using allpass-based filters [37].

Further, the Wavelet Transform (WT) allowed one to balance resolution at any time and frequency, which gave the ability to automatically capture different periodicities (frequencies inverse): periodicity 3 in exons [38] and periodicity 10 in nucleosomes [39]. The wavelet transform was also shown to reflect the characteristic signature associated to the tandem repeats. Indeed, some biological sequences (such as tandem repeats) were characterized by periodicities; the scalograms served to visualize the way these features appear as well as their locations [40]. Given the efficacy of these signal processing tools, analyzing the helitron DNA category (which is governed by complex latent periodicities) is particularly challenging.

Our main goal is to characterize helitrons within the framework of signal processing. For this, the Frequency Chaos Game Signal (FCGS) is selected to preserve the statistical proprieties of DNA sequences. Secondly, signal processing analysis techniques are applied to identify each helitron type by a genomic signature. The key component of this system is the combination of the DNA coding with the Frequency Chaos Game Signal (FCGS) and the windowed Smoothed Fourier technique which enhances the spectral signature of DNA (the overall periodicities). This paper distinguishes four main sections: the First section introduces the work. The second section describes the methodology adopted for the helitron identification. The third section explains the establishment of the helitron's signal database. The fourth section explains the spectral analysis used to characterize helitrons. Then, the experimental results are provided and discussed. Finally, a summary is put forth describing the effectiveness to capture all the periodicities in helitrons sequences based on this new approach. The rest of this paper end with a perspective that open questions concerning the biology of this type of TEs.

II. METHODOLOGY OF HELITRON'S IDENTIFICATION

With the wealth of genomic sequences now available, the identification of a specific DNA element has to be an automatic task. The following methodology consists of three steps. The flowchart describing this work is given in Fig. 1.

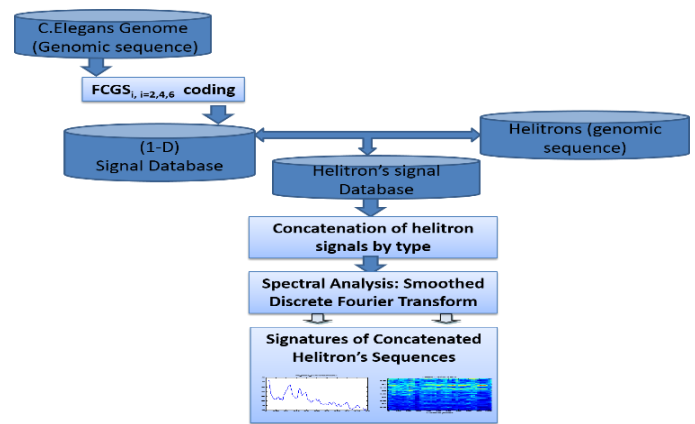


Fig. 1. The Helitron identification establishment flowchart in C.elegans organism.

The first step: A DNA sequence must be converted into a numerical sequence before processing. This step consists in generating different signals for each C.elegans chromosome. For this, chromosomes sequences from the NCBI data base for the C.elegans model (<http://www.ncbi.nlm.nih.gov/Genbank>) are extracted. Then, 1-D signals are generated by applying the Frequency Chaos Game Signal of order i : $FCGS_{i, (i=2, 4, 6)}$ for the whole chromosomes. This coding technique is based on the apparition's probability of N successive nucleotides groups in an entry DNA sequence [29]-[31]. Here, three values of the length: $N=2, 4, 6$ are selected.

The probability (P_{N_nuc}) of given N nucleotides in the chromosome is as follows:

$$P_{N_nuc} = N_{N_nuc} / N_{ch} \quad (1)$$

N_{N_nuc} represents the number of apparition of the N nucleotides in the whole sequence. N_{ch} represents the length in base pairs of the DNA sequence. In other words, the number of occurrence of each of these elements (N nucleotides) in the genome are counted. It is important to note that the FCGS signal depends on the DNA sequence to be encoded since it is based on counting the words contained in it. It reflects, therefore, the statistical-features of the DNA sequence itself. In fact, changing the DNA sequence would affect the probability of apparition of words and the FCGS values. The interesting point here is that this coding technique is useful in terms of enhancing the repetitive DNA (such is the case of helitronic DNA) at any desired scale (i.e. for any word's size). Specifically, the highest level of repetitions in DNA are detected by high order FCGS's coding techniques. It must be noted that hidden information at a certain scale can be highlighted at another scale; that's make the FCGS very suitable to investigate helitrons.

The second step: It consists in the establishment of a 1-D signal database of the helitrons. This is done through the association of the $FCGS_i$ values to each group of letters in the helitron sequence.

In position (k), the oligomer (i), which consists of N nucleotides, is replaced by the correspondent occurrence probability:

$$S_{N_nuc}(k) = \sum_i P_{N_nuc}(i,k) \quad (2)$$

The sum of the N nucleotide indicators (S_{N_nuc}) can be computed as following:

$$H_w[n, k] = H[n]W[n - k\Delta n] \quad (3)$$

Then, a database of helitrons signals regarding different oligomers is prepared. As a result, helitrons are represented by three levels of FCGS ($FCGS_{i,i=2,4,6}$).

As example, Fig. 2 outlines the resulted signals: $FCGS_2$, $FCGS_3$, $FCGS_4$, $FCGS_5$ and $FCGS_6$ of an helitron type NDNAX2 with a size of 341 base pairs and positioned at: [274811bp: 275151bp] in the chromosome II of the C.elegans genome.

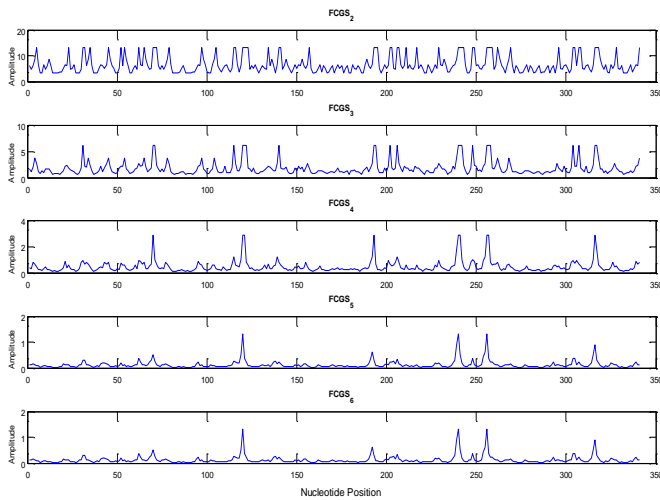


Fig. 2. NDNAX2_CE Helitron presentation by FCGS2 to FCGS6.

After coding process, several numerical signals, denoted by $H[n]$, are obtained for each helitron. A signal database of each helitron's type and for each chromosome in the considered genome is then established. For a desired level i of the FCGS ($FCGS_i$) and for a specific class of the helitron, signals of all existing helitron elements are concatenated in such a way to obtain one global sequence. The resulting database comprises three $FCGS_i$ signals ($i=2, 4, 6$) for each class of helitron contained in the chromosomes of the C.elegans genome.

The third step: The spectral analysis method is adopted to detect the periodicity of each helitron's type. Based on the classic discrete Fourier transform related to numerical sequence (4), a windowed local analysis is used to give more precise location in time and frequency.

The reason for selecting the spectral analysis is to characterize the helitron elements by a global spectral signature.

$$H[k] = \sum_{n=0}^{N-1} H(n)e^{-j\frac{2\pi}{N}nk} \quad (4)$$

At this stage, two types of signatures for each helitron class are revealed: the 1-D spectrum and the 2-D spectrogram. The

spectral and the time-frequency signatures are established for different orders of our coding technique ($FCGS_i, i=2,4,6$).

III. HELITRON'S SIGNALS DATABASE

For experimentation, the C.elegans genome whose chromosomal DNA sequences in the NCBI database [40] are considered. This genome contains ten helitron families diffused through six chromosomes (5 autosomes: ChI, ChII, ..., ChV and 1 gonosome: ChX). These helitron classes are: Helitron1 (H1), Helitron2 (H2), HelitronY1 (Y1), HelitronY1A (Y1A), HelitronY2 (Y2), HelitronY3 (Y3), HelitronY4 (Y4), NDNAX1 (N1), NDNAX2 (N2) and NDNAX3 (N3). The helitron sequences are very heterogeneous. In fact, the apparition number of these elements and the size varies from one chromosome to another as described in Table I. In addition, for a particular type of helitron this size varies from one sequence to another; which makes their identification very difficult.

From Table I, it is obvious that HelitronY1A is the most frequent (occurrence number of 1093) and the longest (size of 425487bp) in the genome.

The least frequent helitron in the genome is Helitron NDNAX1 (with an occurrence number of 77). At the chromosomal level, Helitron NDNAX1 and HelitronY3 are the least frequent elements; they are present in chromosomeII (with an occurrence number of 8). As for the smallest class of helitron is HelitronY3 with a size of 16404 bp; in other words it presents the shortest helitron in the genome.

Because of the variability of helitron's number and size in the six chromosomes, regrouping all the helitron signals in a way to obtain a global signature of each class comes in handy. The idea consists in concatenating all the signals in one vector for a well-defined order of the FCGS, while keeping the apparition order of helitrons in chromosome. This reunification allows linking all the local signatures of helitrons. Our main goal is to find the structure that could be repeated in each helitron.

IV. SMOOTHED DISCRETE FOURIER TRANSFORM

The smoothed spectral analysis is a convenient tool to search for base periodicities within DNA sequences. In fact, the periodically placed motifs may indicate the presence of genes, regulatory elements or other significant hotspots like helitrons; hence the need to locate them and exhibit the correspondent frequency (or periodicity). This technique is applied to investigate the global signature of helitrons within the genome by considering the concatenated FCGS signals of all helitrons which constitutes the novelty of this work. Therefore, two tasks are carried out here:

- Revealing periodicities by enhancing the global periodicities in the helitron sequences by using the Smoothed Fourier Transform.
- Locating the frequencies points using the Smoothed Discrete Fourier Transform applied on sliding window along the DNA sequence [26], [27], [41], [42].

The technique consists in:

a). After converting the DNA sequence into a numerical one, the signal $H[n]$ must be divided into frames of L length with an overlap Δl .

b). Using a sliding analysis window $w[n]$, each L portion is also divided into N overlapped segments with an overlap length Δn :

$$H_w[n,k]=H[n]w[n-k\Delta n] \quad (5)$$

Where, the index of the frequency ($[0, N-1]$) is noted by k .

The choice of the windowing function determines the time-frequency resolution.

c). Using the Discrete Fourier Transform (DFT), each weighted block of the frame $H_w[n]$ is transformed in the spectral domain. The DFT of each segment is expressed as follows:

$$H_w^i[k] = \sum_{n=0}^{N-1} H_w[n, i] e^{-\frac{j2\pi nk}{N}} \quad (6)$$

d). Calculating the DFT mean value for each portion ($1: L$); then, performing the same operation to the N segments. The mean smoothed spectrum is expressed as:

$$Hmean_w^j[k] = \frac{1}{L*N} \sum_{j=0}^{L-1} \sum_{i=0}^{N-1} H_w^i[k] \quad (7)$$

Noted that i corresponds to the index frame of N frames ($[1...N]$), k is the index of the frequency and j corresponds to the index frame of L frames ($[1: L]$).

e). With the obtained values, a matrix containing the joint time frequency information is constructed:

$$\text{Matrix}(j, k) = Hmean_w^j[k] \quad (8)$$

This representation consists of the spectrogram amplitude for a specific index periodicity in a specific nucleotide position in the DNA sequence.

Fig. 3 provides an example of the spectra and the spectrograms of the concatenated helitrons of type Helitrons2_CE that exist in chromosome IV of C.elegans. The frequency and the time-frequency representations are generated by the mean valued technique based on the smoothed Discrete Fourier Transform. As for the sliding window, we suggest using a blackman window with the parameters: $L=1024$, $\Delta l=512$, $N=256$ and $\Delta n=64$. In this example, two levels of the coding technique are used: FCGS2 and FCGS4.

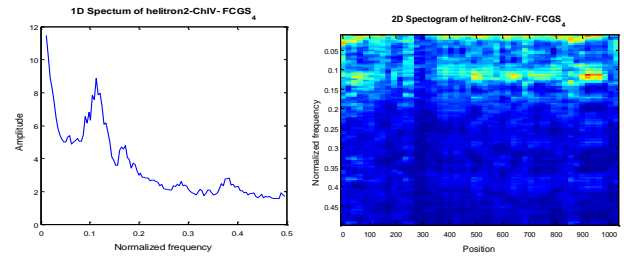
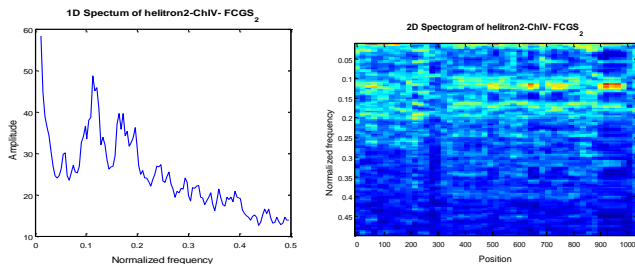
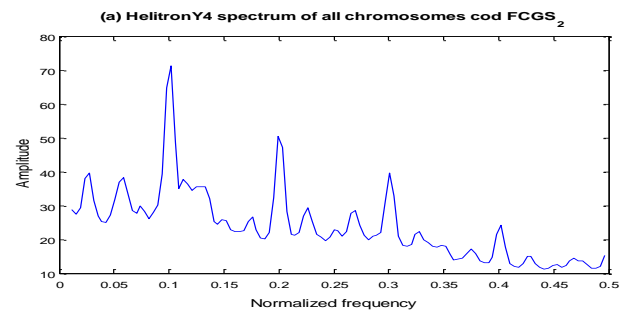


Fig. 3. 1D spectra and 2D spectrograms of helitron2_CE [chromosome IV with size of 25932bp] when coded with FCGS₂ and FCGS₄.

From these figures, it is noticeable that increasing the FCGS order induces a general smoothing of the 1D spectrum and the 2D spectrogram; which allows to enhance the spectral and the time-frequency behaviours of the considered helitron by highlighting its periodicities.

V. HELITRON'S CHARACTERIZATION IN C.ELEGANS: TESTS AND RESULTS

A characterization of helitron DNA at the aim of its identification is the main focus of this paper. Therefore, such a goal can be reached by studying the global resident periodicities of each helitron's class. Thus, considering the helitrons representations in the frequency and the time-frequency plans, a specific signature (pattern or form) characterizing the repetitions of the sequence is allowed by the spectral attributes for each type. The windowed smoothed Fourier analysis allows following the path of existing periodicities (engendered by the repetitions) in each helitron's class. For experimentation, all helitrons existing in the C.elegans genome are encoded by the FCGS_{i(i=2,4,6)} coding technique. Secondly, the FCGS_i fragments which correspond to the helitron sequences in the same chromosome are concatenated for each helitron's type. For comparison, the concatenation of all helitrons contained in the genome (all chromosomes) is considered. Finally, the smoothed spectral analysis is applied to these sequences. In this step, many types of windows with different values of length and overlap are tested. The optimal parameters are fixed to $L=1024$, $\Delta l=512$, $N=256$, $\Delta n=64$ and the most accurate smoothed spectrum is given by a Blackman window. Consequently, two types of spectral representations are obtained: the 1-D spectrum and the 2-D spectrograms. For illustration, the example of HelitronY4 is selected and encoded with FCGS₂. The resulting 1-D spectra are presented in Fig. 4.



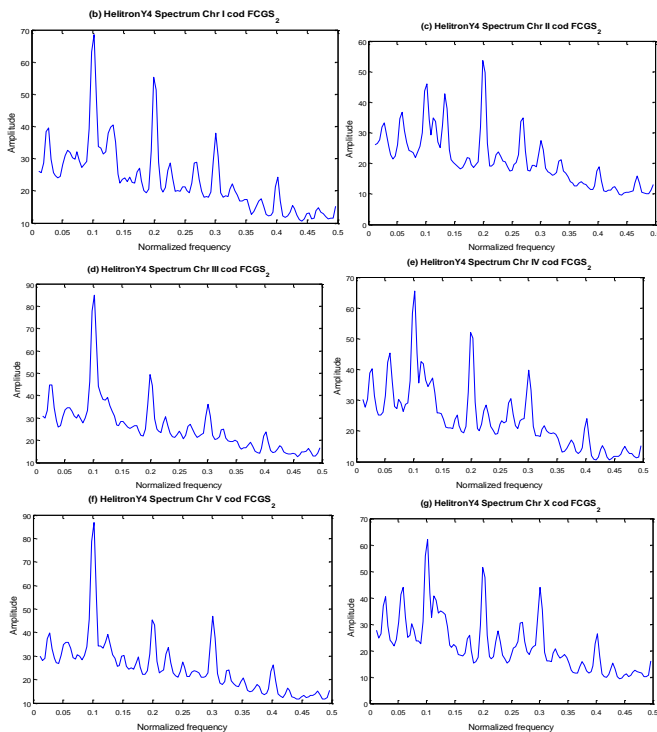


Fig. 4. HelitronY4 spectra- (a) spectrum of the concatenated chromosomes- (b:g) spectrum of each chromosome apart.

The sub-figures (b, c, d, e, f, g) provide the spectrum shape for each chromosome apart; the sub-figure (a) gives the shape spectrum of the overall genome. The horizontal axis of each sub-figure indicates the frequency (which is equivalent to the inverse of periodicity) measured by the Smoothed Fourier Transform and the vertical axis indicates the spectrum's amplitude.

By examining each subfigure closely, it is crystal clear that HelitronY4 is identified by five peaks whose power value differs from one chromosome to another. These peaks are located around remarkable frequencies which are: 0.02 (periodicity 50), 0.1 (periodicity 10), 0.2 (periodicity 5), 0.3 (periodicity 3) and 0.4 (periodicity 2). Since the *C.elegans* genome kept the same behavior as for each chromosome apart, these periodicities characterize the HelitronY4 type. Here, the idea is to identify each helitron class with a specific spectrum since the behavior of all spectra remains almost the same with a small variation in amplitude.

A. Helitron's Signature and Effect of the FCGS Order on the Spectrum

In this section, the role played by the FCGS coding in enhancing the helitron signature is thoroughly examined. For this purpose, the mean values of the Smoothed Discrete Fourier Transform for each type of helitron sequence are computed and thus for different orders of FCGS (order=2,4,6).

The spectrum representations of all helitron classes in *C.elegans* (Helitron1, HelitronY1, HelitronY1A, Helitron2, HelitronY2, HelitronY3, HelitronY4, NDNAX1, NDNAX2 and NDNAX3) are given in Fig. 5.

These representations reflect the specific periodicities (frequencies) that characterize each helitron's type which forms a pertinent tool for their identification. In fact, each helitron class is shown to possess a specific spectral signature.

Further, for each helitron type, the effect of increasing the FCGS order in the spectrum shape evolution is remarkable. But the overall shape remains when increasing the FCGS level. Based on that, the fact that the helitron can be identified by its spectral signature is confirmed. And that is emphasized even in the spectrum shape evolution.

On other hand, a high similarity between these three helitrons is striking:

- Helitron1_CE
- HelitronY1_CE
- HelitronY1A_CE

These helitrons have in common the frequencies; 0.02734 (Periodicity 37), 0.05859 (Periodicity 17) and 0.1 (Periodicity 10); which have the highest amplitudes. These periodicities reflect the presence of hidden minisatellite into the helitrons sequences.

In addition, Helitron2_CE have important amplitudes within the frequencies bands [0.09 0.125] and [0.15: 0.2]. They share several frequencies with the HelitronY2_CE class. The latter helitron (HelitronY2_CE) is shown to be easily recognized through its distinctive spectrum. In fact, during tests, from a chromosome to another the HelitronY2_CE is the only subclass which possesses an invariant behavior along the whole genome.

The HelitronY3 have the main highest amplitudes around the frequencies: 0.015, 0.15 (periodicity 7), 0.09 (periodicity 11) and 0.05 (periodicity 20). Regarding the FCGS signal patterns, this helitron class is shown to have a regular behavior.

The HelitronY4_CE have four remarkable amplitude around the frequencies which are: 0.02734 (periodicity 36), 0.05859 (periodicity 17), 0.1 (periodicity 10), 0.2 (periodicity 5), 0.3 (periodicity 3) and 0.4 (periodicity 2).

The Helitron NDNAX1 have two principal frequency bands which are centered around the frequency 0.1 (periodicity 10) and around the frequency 0.03 (periodicity 30).

The Helitron NDNAX2_CE is characterized by the frequencies: 0.01563 (periodicity 54), 0.04 (periodicity 24), 0.1(periodicity 10), and 0.168 (periodicity 6). Also, their spectrogram contains repetitive patterns around the frequencies band 0.1.

Finally, the Helitron NDNAX3_CE has pronounced amplitude around the frequency 0.1(periodicity 10). Other important frequencies are within reach, such as 0.01953 which corresponds to periodicity 51 and 0.02344 which corresponds to periodicity 42. In addition, the periodicities 10 and 9 are present in this class of helitrons families.

B. Helitron's Spectrograms with different FCGS Orders

The time-frequency presentation allows an energetic interpretation of signals. This representation has shown high-

ability in detecting different hotspots in DNA sequences. In this part, the specific regions that can exist in the helitron sequences are examined.

Fig. 6 presents the spectrogram of each helitron type considering three levels of FCGS: FCGS₂, FCGS₄ and FCGS₆. The vertical axis of each sub-figure indicates the frequency measured by the Smoothed Fourier Transform. As for the horizontal axis, it indicates the position in base-pairs.

Based on these sub-figures, each helitron class seems to possess a unique time-frequency behavior. In other words, it forms a time-frequency signature which will be taken as a basis for the helitron identification.

In this case, the FCGS₂ coding seems to allow the best way of characterizing helitrons since it shows more details about the energy distribution in spectrograms. On the other hand, increasing the coding order has no major impact on the overall signature for all helitron types. Nevertheless, smoothed spectrograms with high levels of the FCGS signals are obtainable.

VI. CONCLUSIONS

Exploring the latent periodicities of helitrons (eukaryotic rolling-circle transposons) can play a key role in the identification of this subclass of DNA. The way these periodicities appear can mark specific regions of helitrons in a unique manner; which constitutes a signature allowing us to distinguish these elements.

This study was done based on the spectral analysis. In fact, the recognition of each class's periodicities can be very useful for helitron classification.

To be able to apply the spectral method, the DNA sequence is converted into a numerical 1-D signal using the FCGS coding. This technique offers the possibility to encode DNA into several signals according to a well-defined order. Three levels of the representation are taken into account: FCGS₂, FCGS₄ and FCGS₆.

To know more about how similar are these helitrons along the genome, all the helitron sequences that exist in the C.elegans chromosomes are associated in order to obtain a global genomic signature.

After that, the Smoothed Fourier Transform Analysis is applied and the 1-D spectra and the 2-D spectrograms are collected as characteristic signatures of the studied elements. In fact, for each helitron type, the periodicities shape was involved in the spectral and the time-frequency representations in a unique manner; which forms a pertinent tool for the helitron characterization and identification.

Comparing the signature of helitrons in a chromosome with the signature of the overall genome, the great similarity between them is plain to see. In addition, increasing the FCGS order has not affected the global behavior in the spectra and the spectrograms: the main periodicities of each helitron class have remained. A smoothing effect is spotted in the representations. This confirms that the helitron adopts a distinctive behavior which characterizes it and thus permits its identification.

In this work, the major advantage of these frequency representations (spectrum) and time-frequency representation (Spectrograms) is that it captures all the periodicities (repetitive sequence) for the helitronic sequences. It is worth mentioning that this approach can be used in detecting periodicities of other transposable elements. The limitation of the approach depends on the limitation of the Fourier transform. Therefore, this approach remains limited as it does not allow the temporal localization using a fixed window size.

VII. FUTURE WORK

In this work, the importance of helitron characterization has been highlighted and a method to perform their identification based on the Smoothed Fourier Analysis has been suggested.

Classification methods based on the feature analysis technique might be investigated to improve the classification accuracy of helitrons as a future work.

TABLE I. THE OCCURRENCE NUMBER AND THE LENGTH (IN BASE PAIRS) OF HELITRONS IN C.ELEGANS.

Type	Characteristics	ChI	ChII	ChIII	ChIV	ChV	ChX	Genome
H1	N helitrons	42	32	37	28	42	14	197
	Size(bp)	12911	59232	11836	9456	11012	3486	107933
H2	N helitrons	106	74	72	77	100	40	469
	Size(bp)	42643	28078	31715	25932	38158	15788	182314
Y1	N helitrons	60	44	82	83	188	26	483
	Size(bp)	21567	21193	35148	32612	89573	12171	212264
Y1A	N helitrons	176	132	211	155	366	53	1093
	Size(bp)	60860	43949	75437	75410	145489	24342	425487
Y2	N helitrons	54	44	39	79	70	51	337
	Size(bp)	11026	8184	6851	15253	12337	10091	63742
Y3	N helitrons	11	8	17	24	25	32	117
	Size(bp)	1578	1264	2521	3496	3167	4378	16404
Y4	N helitrons	68	68	81	114	148	44	523
	Size(bp)	29522	36026	35311	55409	72431	24371	253070
N1	N helitrons	9	8	14	15	19	12	77
	Size(bp)	6197	6701	6896	10422	7761	5730	43707
N2	N helitrons	30	40	22	43	37	16	188
	Size(bp)	7762	19810	12476	16698	21105	7256	85107
N3	N helitrons	20	20	24	33	23	14	134
	Size(bp)	18229	14357	21517	29163	16545	17049	116860

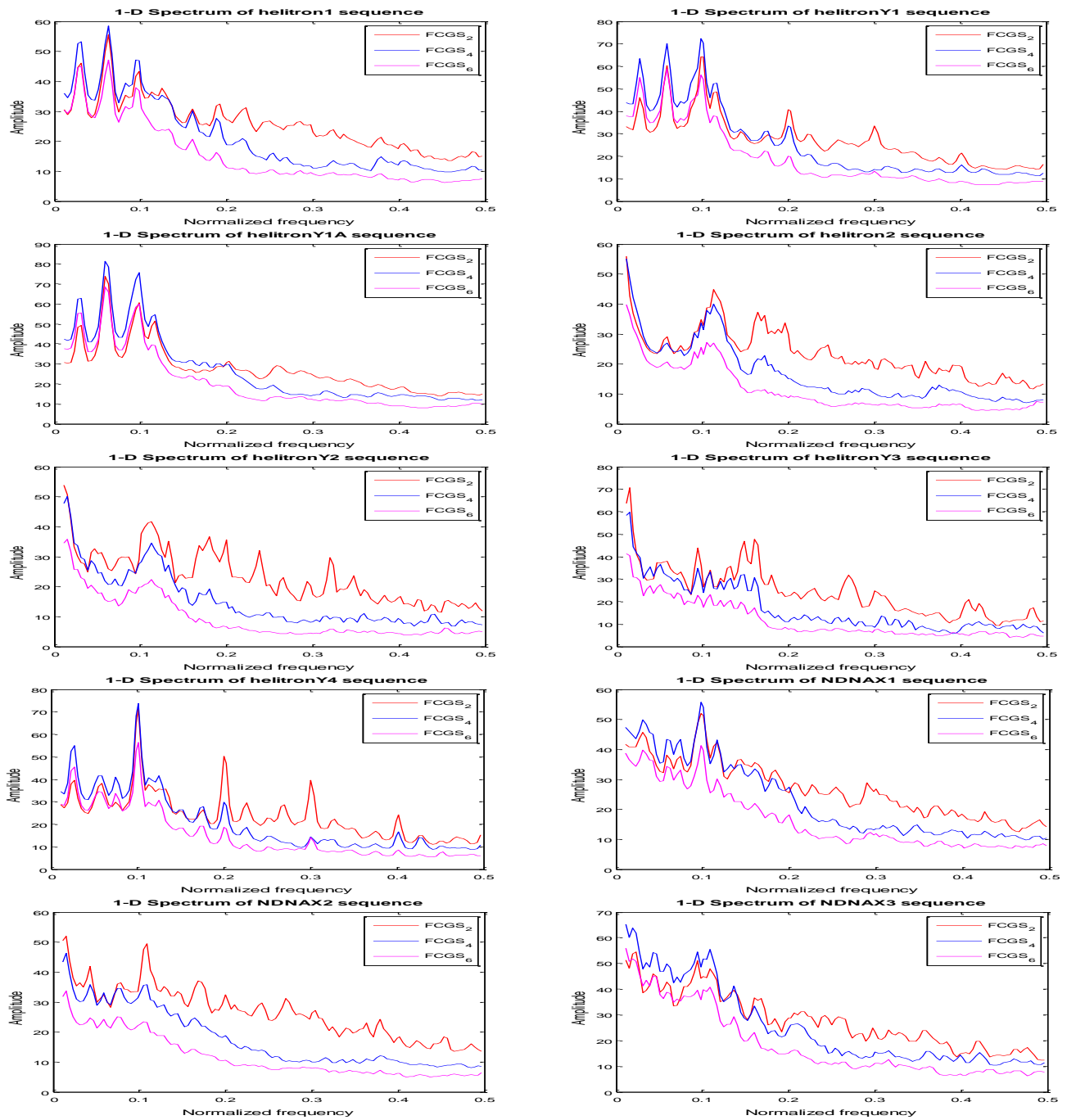
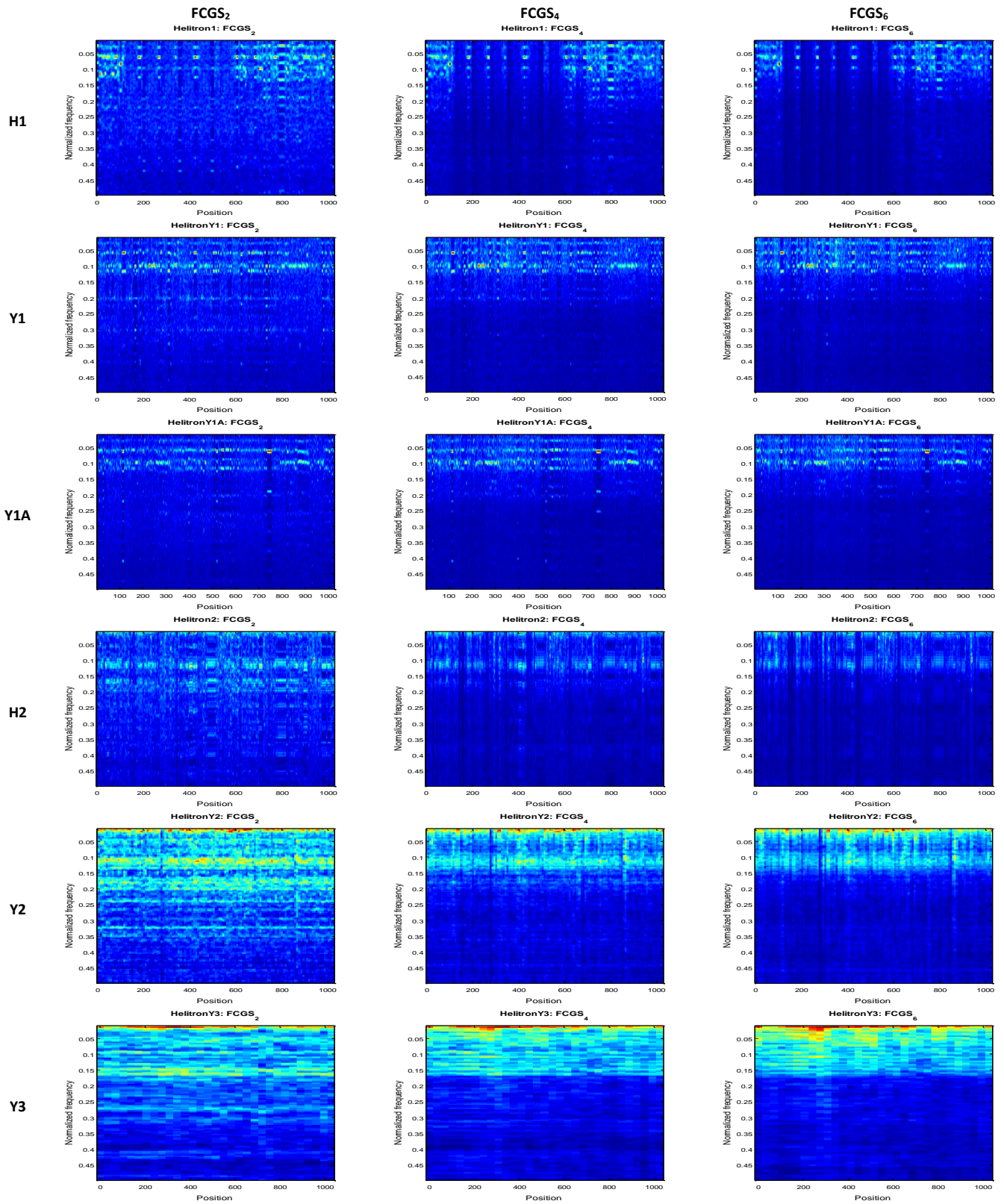


Fig. 5. 1-D Spectrum of each helitron's type in the C.elegans genome when coded by FCGS₂, FCGS₄ and FCGS₆.



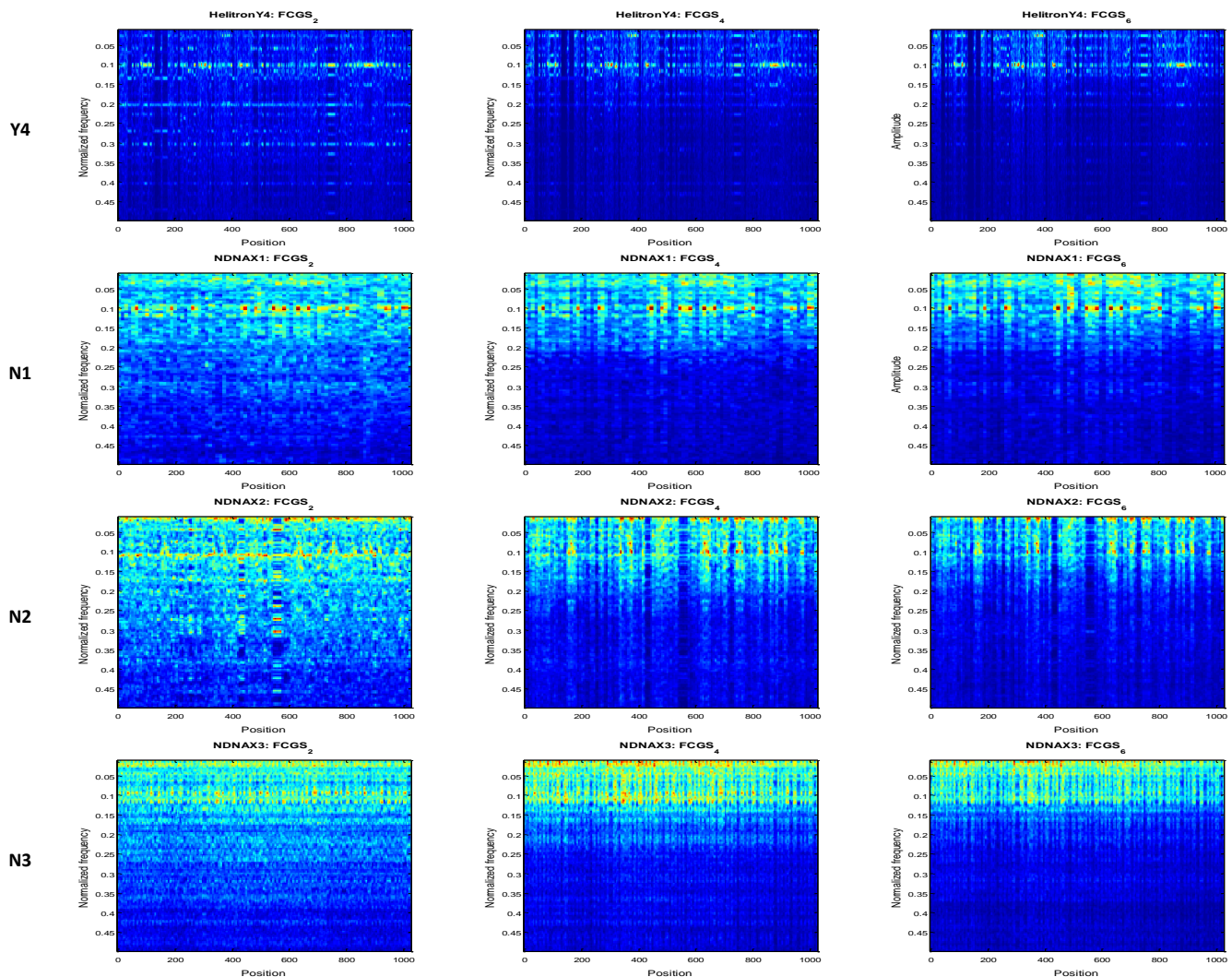


Fig. 6. 2-D Spectrogram of each helitron type in C.elegans when coded with FCGS_{2,4,6}.

REFERENCES

- [1] N. V. Fedoroff, "Transposable elements as a molecular evolutionary force," *Annals. New York Acad. Scien*, vol.870, no. 1, pp. 251-264, 1999.
- [2] M. G. Kidwell and Lisch, D, "Transposable elements as sources of variation in animals and plants", *Proce. National. Acade. Scien*, vol. 94, no. 15, pp. 7704-7711, 1997.
- [3] D. J. Finnegan, "Transposable elements," *Current opinion in genetics & development*, vol. 2, no. 6, pp. 861-867, 1992.
- [4] C. Feschotte and S. R. Wessler, "Treasures in the attic: rolling circle transposons discovered in eukaryotic genomes," *Proceedings of the National Academy of Sciences*, vol. 98, no.16, pp.8923-8924, 2001.
- [5] V. V. Kapitonov and J. Jurka, "Helitrons on a roll: eukaryotic rolling-circle transposons," *TRENDS in Genetics*, vol.23, no.10, pp. 521-529, 2007.
- [6] E. Cocca, S. De Iorio and T. Capriglione, "Identification of a novel helitron transposon in the genome of Antarctic fish," *Molecular phylogenetics and evolution*, vol. 58, no. 3, pp. 439-446, 2011.
- [7] J. Thomas, E. J. Pritham, "Helitrons, the eukaryotic rolling-circle transposable elements," *Microbiology spectrum*, vol. 3, pp. 4, 2015.
- [8] A. Borgognone, R. Castanera, E. Muguerza, A. G. Pisabarro, and L. Ramirez, "Somatic transposition and meiotically driven elimination of an active helitron family in *Pleurotus ostreatus*," *Dna Research*, vol. 24, no.2, pp. 103-115, 2017.
- [9] J. Thomas, C. D. Phillips, R. J. Baker, E. J. Pritham, "Rolling-circle transposons catalyze genomic innovation in a mammalian lineage," *Genome biology and evolution*, vol. 6, no. 10, pp.2595-2610, 2014
- [10] M. Morgante, S. Brunner, G. Pea, K. Fengler, Zuccolo, A., Rafalski A, "Gene duplication and exon shuffling by helitron-like transposons generate intraspecies diversity in maize". *Nature genetics*, vol. 37, no.9, pp. 997-1002, 2005.
- [11] J. Lai, Y. Li, J. Messing, H.K. Dooner, "Gene movement by Helitron transposons contributes to the haplotype variability of maize," *Proceedings of the National Academy of Sciences of the United States of America* vol. 102, no.25, pp.9068-9073, 2005.
- [12] P. S. Schnable, D. Ware, Fulton, J. C. Stein, F. Wei, S. Pasternak, and Minx, P, "The B73 maize genome: complexity, diversity, and dynamics," *science* vol. 326, no.5956, pp. 1112-1115, 2009.
- [13] S. Tempel, J. Nicolas, A. El Amrani, I. Coué, "Model-based identification of Helitrons results in a new classification of their families in *Arabidopsis thaliana*," *Gene*, vol. 403 no. 1, pp.18-28, 2007
- [14] V. V. Kapitonov, J. Jurka, "Rolling-circle transposons in eukaryotes," *Proceedings of the National Academy of Sciences* vol. 98, no. 15, pp. 8714-8719, 2001
- [15] M. E Hood, "Repetitive DNA in the automictic fungus *Microbotryum violaceum*," *Genetica*, vol. 124, no. 1, pp. 1-10, 2005.
- [16] R. T. Poulter, T. J. Goodwin, M. I. Butler, "Vertebrate helitrons and other novel Helitrons," *Gene*, vol. 313, pp.201-212, 2003.

- [17] Zhou Q, et al. Helitron transposons on the sex chromosomes of the Platyfish *Xiphophorus maculatus* and their evolution in animal genomes. *Zebrafish* 3, 39–52 (2006)
- [18] L. Yang, J.L. Bennetzen, “Distribution, diversity, evolution, and survival of Helitrons in the maize genome,” *Proc Natl Acad Sci USA* vol. 106, no. 47, pp. 19922–19927, 2009.
- [19] M. J. Han, Y. H. Shen, M. S. Xu and H. Y. Liang, “Identification and evolution of the silkworm helitrons and their contribution to transcripts,” *DNA research*, dst024, 2013.
- [20] C. Du, J. Caronna, L. He and H. K. Dooner, “Computational prediction and molecular confirmation of Helitron transposons in the maize genome” *Bmc Genomics*, vol. 9, no. 1, pp. 51, 2008
- [21] L. Yang and J. L. Bennetzen, “Structure-based discovery and description of plant and animal Helitrons,” *Proceedings of the National Academy of Sciences* vol. 106, no.31, pp. 12832-12837, 2009.
- [22] M. Sweredoski, L. DeRose-Wilson and B. S. A. Gaut, “comparative computational analysis of nonautonomous helitron elements between maize and rice,” *BMC genomics* vol. 9, no.1, pp. 467, 2008.
- [23] W. Xiong, L. He, J. Lai, H. K. Dooner and C. Du, “HelitronScanner uncovers a large overlooked cache of Helitron transposons in many plant genomes,” *Proceedings of the National Academy of Sciences*, vol. 111, no.28, pp.10263-10268, 2014.
- [24] L. Yang and J. L. Bennetzen, “Distribution, diversity, evolution, and survival of Helitrons in the maize genome” *Proceedings of the National Academy of Sciences*, vol. 106, no.47, pp.19922-19927, 2009.
- [25] C. Du, N. Fefelova, J. Caronna, L. He and H. K. Dooner, “The polychromatic Helitron landscape of the maize genome,” *Proceedings of the National Academy of Sciences* vol. 106, no.47, pp.19916-19921, 2009.
- [26] A. Elloumi-Oueslati, N. Ellouze, and Z. Lachiri, “3D spectrum analysis of DNA sequence: application to *Caenorhabditis elegans* genome” In *Bioinformatics and Bioengineering (BIBE 2007)* pp. 864-871, 2007.
- [27] A. Elloumi-Oueslati, I. Messaoudi, N. Ellouze and Z. Lachiri, “Spectral analysis of global behaviour of *C. elegans* chromosomes,” *INTECH Open Access Publisher*, 2012
- [28] E. Pirogova, Q. Fang, M. Akay, and I. R. E. N. A. Cosic, “Investigation of the structural and functional relationships of oncogene proteins,” *Proceedings of the IEEE*, vol. 90, no. 12, pp. 1859-1867, 2002.
- [29] I. Messaoudi, A. Elloumi-Oueslati, and Z. Lachiri, “2D DNA Representations Generated Using a New Coding and the Time-Frequency Analysis,” *Journal of Medical Imaging and Health Informatics*, vol. 5, no.5, pp.1035-1044, 2015.
- [30] R. Touati, I. Massaoudi, A. Elloumi-Oueslati, and Z. Lachiri, “Classification of Helitron’s Types in the *C. elegans* Genome based on Features Extracted from Wavelet Transform and SVM,” In *Bioinformatics*, pp. 127-134, 2018.
- [31] I. Messaoudi, A. Elloumi-Oueslati, and Z. Lachiri, “Building specific signals from frequency chaos game and revealing periodicities using a smoothed Fourier analysis,” *IEEE/ACM Transactions on Computational Biology and Bioinformatics*, vol. 11, no.5, pp. 863-877, 2014.
- [32] A. A. Tsonis, J. B. Elsner, and P. A. Tsonis, “Periodicity in DNA coding sequences: implications in gene evolution,” *Journal of Theoretical Biology*, vol. 151, no. 3, pp. 323-331, 1991.
- [33] L. Wang, and L. D. Stein, “Localizing triplet periodicity in DNA and Cdna sequences,” *BMC Bioinformatics* vol. 11, pp. 550, 2010.
- [34] M. M. Hlaing, B. R. Wood, D. McNaughton, and D. Ying, “Effect of drying methods on protein and DNA conformation changes in *Lactobacillus rhamnosus* GG by Fourier transform infrared spectroscopy,” *Journal of Agricultural and Food Chemistry*, 2017
- [35] C. Yin, and J. Wang, “Periodic power spectrum with applications in detection of latent periodicities in DNA sequences,” *Journal of mathematical biology*, vol. 73, no. 5, pp. 1053-1079, 2016.
- [36] J. Mena-Chalco, H. Carrer, Y. Zana, and Jr. R. M. Cesar, “Identification of protein coding regions using the modified Gabor-wavelet transform,” *IEEE/ACM Transactions on Computational Biology and Bioinformatics*, vol. 5, no. 2, pp.198-207, 2008.
- [37] P. P. Vaidyanathan, and B. J. Yoon, “Gene and exon prediction using allpass-based filters,” In *Proc. IEEE Workshop on Gen. Sig. Proc and Stat*, 2002.
- [38] X. Zhang, Z. Shen, G. Zhang, Y. Shen, and M. Chen, “Short Exon Detection via Wavelet Transform Modulus Maxima,” *PLoS one*, vol. 11, no. 9, pp. e0163088, 2016.
- [39] R. Touati, I. Massaoudi, A. Elloumi-Oueslati, N. ellouze and Z. Lachiri, Nucleosome location method based on morlet wavelet analysis scalograms investigation. In *Advanced Technologies for Signal and Image Processing (ATSIP)*, pp. 307-312, 2016.
- [40] A. Elloumi-Oueslati, I. Messaoudi, Z. Lachiri and N. Ellouze, “A new way to visualize DNA’s base succession: the *Caenorhabditis elegans* chromosome landscapes,” *Medical & biological engineering & computing*, vol. 53, no. 11, pp. 1165-1176, 2015.
- [41] D. Anastassiou, “Genomic signal processing,” *IEEE signal processing magazine*, vol. 8, no. 4, pp.8-20, 2001.
- [42] A. Fukushima, T. Ikemura, M. Kinouchi, and T. Oshima, “Periodicity in prokaryotic and eukaryotic genomes identified by power spectrum analysis,” *Gene*, 300(1), 203-211 (2002).