Revealing Helitron signatures in *Cænorhabditis elegans* by the Complex Morlet Analysis based on the Frequency Chaos Game Signals

Imen Messaoudi[1], Afef Elloumi Oueslati[1] and Zied Lachiri[2]

[1]Université de Tunis El Manar, Ecole Nationale d'Ingénieurs de Tunis, Laboratoire Signal, Images et Technologies de l'Information, LR-11-ES17, BP 37, le Belvédère, 1002, Tunis, Tunisie.

imen_messaoudi@ymail.com
afefelloumi@gmail.com

[2]Département de Génie Physique et Instrumentation,
Département physique et Instrumentation, Institut National des Sciences Appliquées et de Technologie (INSAT), BP. 676, Centre Urbain Cedex, 1080, Tunis, Tunisie.

Zied.lachiri@enit.rnu.tn

Abstract.
Helitrons are a specific class of DNA which belongs to the category of transposable elements (TEs). The specificity of these elements resides in their interaction with transposition proteins; which considerably indicates their impact on the organisms' evolution. Since the helitron family possesses variable and complex structures, their identification is a veritable challenge. Therefore, the emergency to find new computational tools that can efficiently identify such structures.
Here, we propose the complex Morlet wavelet analysis to investigate helitrons' characterization in the *Caenorhabditis elegans* genome. To be able to compute the continuous wavelet transform, a DNA data set must be converted into a numerical sequence. Thus, we propose a new coding technique based on the Chaos Game Representation that we name the Frequency Chaos Game Signal (FCGS). DNA scalogram representations reveal very specific periodic patterns that characterize helitrons independently from their nature.

Keywords: Continuous Wavelet Transform; complex Morlet wavelet; Frequency Chaos Game Signal; Helitron; periodic motifs; *Caenorhabditis elegans*.

1 Introduction

The DNA sequences hold all the required instruction to rule the biological processes in the living beings' cells. Its complex nature is still poorly understood. This made it the center of interest for many scientific studies and thus leads to several discoveries. Nowadays, one recognize some structures like genes (which are responsible of proteins production), "junk" DNA within and around genes, repeating units and transposable elements. Only some of the

molecular constituents as well as the inherent features of DNA sequences are revealed. Therefore, much work remains to be done about the DNA characterization.

Here, we direct our attention towards to helitrons, which are a distinguished class of transposable elements. Transposable elements (TEs) are DNA sequences which can change their position along the genome through a specific mechanism called transposition. The transposition of TEs can cause deletions and genome size's cell alteration. In the other hand, they can capture gene fragments and transpose them within the genome, which gives rise to genes diversity. This has a great effect on the evolution of organisms [21]. Thus, helitron analysis is an important topic. Helitrons represent a new class of TEs that is found in all eukaryotic genomes. The transposition mode adopted by these elements is known as the rolling-circle mechanism.

The discovery of the helitron DNA category was made by Vladimir Kapitonov and Jerzy Jurka in 2001 [1]. Their researches gave birth to the Repbase database which consists on a universal classification of diverse eukaryotic transposable elements and repetitive elements [2]. Several helitrons are annotated in Repbase and accompanied by a short description; the related consensus sequences are also available in the database. However, these descriptions are not sufficient to characterize enough the helitronic sequences. In fact, helitrons practically have no typical structural features, as in the case of the other TEs. Their analysis is, therefore, a difficult task.

Usually, analyzing helitronic sequences is performed by manual comparisons. For example, in [24] a vertical comparison was performed to identify helitrons in maize. But the method is computationally expensive; thus the urgent need to find new automated tool for helitrons' searching. In [22, 23, 26], a number of automated algorithms are presented. Most of these approaches are based on a multiple alignment created by ClustalW; which is not the optimal way because of gaps. However, in [25, 26] a model of Helitrons identification introduced a classification step (based on HMM profiles) which allows describing the differences between helitron family members in terms of domains content. All of these approaches do not permit a visual identification of helitrons, which forms the crucial motivation of this work. Then, our main goal is not to provide a platform for helitron identification, but to present a new way for visually recognition of the helitron structure.

In this paper, we focus on helitrons sequences analysis in the *Cænorhabditis elegans* (*C.elegans*) genome as part of the genomic signal processing discipline. For this purpose, we propose the continuous wavelet transform to analyze the DNA data sets. Nevertheless, the biological sequences must be converted into digital ones, which defines the coding DNA principle. As a coding technique, we introduce a new method based on the Chaos Game Theory; which is the Frequency Chaos Game Signal (FCGS). The approach allows the representation of a given DNA sequence by assigning the frequency of oligomers occurrence. The usage of the statistical properties of a genome as a base of our coding allows revealing the core features in DNA sequences. The proposed model is shown to efficiently characterize different genomic sites in *C.elegans* through the complex Morlet wavelet scalograms independently from the size or the biological function. The novelty in this work resides not only in the coding technique, but also in providing a new way to represent the long genomic data using color scalograms. Although previous studies have investigated the role played by wavelets in identifying regions with biological interest, our results are shown to be unique.

In fact, in [12], the fractal analysis was used to study the long-range correlations of chromatin. Thus, the Hölder exponent using the modulus maxima of the continuous wavelet transform was calculated. On one hand, the Gabor wavelet transform was modified by fixing its basic frequency at 1/3 at the aim to predict exons [9]. On the other hand, the continuous wavelet transform was used to reveal common characteristic frequencies of different oncogenes[1] [13, 14]. Besides, in [15] different wavelet transforms were used to denoise gene sequences which aims at determining the coding regions. A number of other studies have been developed in this sense [10, 11, 16, 17]. Here, our work is devoted to examine the inherent helitron sequence features in the *C.elegans* genome. Thus, an exploration of the periodic structures within five types of helitron DNA will be presented. These helitrons are :{Helitron1, HelitronY1, Helitron Y1A, Helitron Y2 and Helitron Y4}.

This paper is organized as follow. After an introduction, we introduce the Frequency Chaos Game Signal technique. Then, we give an overview on the Continuous Wavelet Transform and the Complex Morlet mother wavelet. In the fourth section, we expose some results of the helitron representation by complex Morlet scalograms. Finally, in section 5 we conclude the content of the paper.

2 Introduction to the Frequency Chaos Game Signals

The DNA graphical representations are of great importance, since they play a key role in visualizing and characterizing information contained in such sequences. In addition, they bridge the gap between the presence of DNA in the form of letter strings within public databases and applying digital signal processing techniques.

The DNA sequences consist of a chain made of four types of nucleotides: Adenine 'A', Cytosine 'C', Guanine 'G' and Thymine 'T'. The principle of coding consists on converting the DNA symbols into numerical values. In this context, we outline the construction of a new DNA representation based on the Chaos Game Representation scheme; which is the Frequency Chaos Game Signal (FCGS). The Chaos Game Representation (CGR) is an algorithm based on recurrent iterated function system that can allow producing pictures of fractal nature. It consists on mapping a DNA sequence into a unit square in such way that one obtains a scatter dots plot. Thus as a first step, the four nucleotides must be placed at the corners of the square and an initial point P_0 must be placed at the center of the representation's space [18]. Then, considering that a DNA sequence S can be regarded as: $S=\{S_1,S_2, \ldots S_N\}$, each letter is represented into the square by a point P_i, with i varies from 1 to N. The point coordinates are calculated iteratively using the formula (1). That consists on repeatedly placing the current point P_i on the half distance between the previous point P_{i-1} and the segment joining the vertex corresponding to the letter S_i [8, 18]. Note that the CGR map consists only on the final dots set.

$$P_i = \frac{1}{2}\left(P_{i-1} + \ell_{S_i}\right) \; where \; \ell_{S_i} can \; be: \ell_A, \; \ell_C, \ell_G \; or \; \ell_T \qquad (1)$$

$$with: \ell_A = (0,0), \ell_C = (0,1), \ell_G = (1,1) \; and \; \ell_T = (1,0)$$

[1] Oncogenes are genes whose expression promotes the cancer occurrence

In the following, figure 1 illustrates how to represent the sequence "GCATT" by the CGR technique.

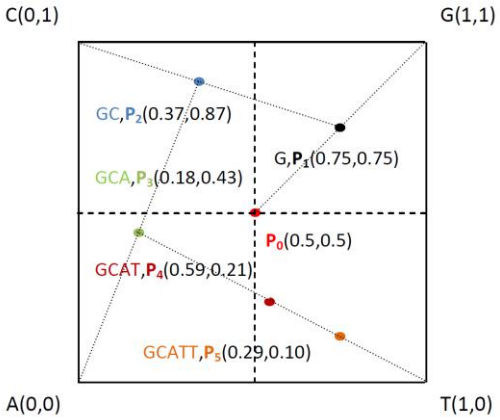

Fig. 1. Mapping the sequence "GCATT" by the CGR technique. The first point P1 is drawn halfway between P0 and the G corner, thus its coordinates are=1/2[(0.5, 0.5)+(1,1)]=(0.75, 0.75). Likewise, P2 is placed in the middle between P1 and the C corner and so on. The dot set {P1, P2, P3, P4 and P5} represents the CGR plot.

The CGR approach provides unique representations that characterize genomes by revealing specific fractal patterns. They have also the advantage to represent the frequency of oligonucleotides. This can be realized by dividing the CGR into $2^k \times 2^k$ quadrants, where specific k-lengthen patterns are associated. Thus, the frequency of a sub-pattern occurrence can be estimated by counting dots in the correspondent quadrant divided by the complete length of the DNA sequence. Then, all the frequency values are coded according to a color scale so that the darker the color is the higher the occurrence frequency is [19, 20].

The figure 2, exposes the Frequency Chaos Game Representations for k={1, 2 and 3} of the *C.elegans* chromosome III, as well as the arrangement of all possible k-patterns into the related quadrants.

Fig. 2. FCGRs of the *C.elegans* chromosome III with order {1, 2 and 3} as well as the distribution of the related k-mers.

Based on the statistical properties of the FCGRs, we develop a particular scheme to which we gave the name of the Frequency Chaos Game Signal (FCGS).

The basic concept of the FCGS is to transforms the FCGR representation of a genomic sequence into a one dimensional representation. In other words, we give a temporal aspect to the FCGR by assigning the frequency of occurrence of a given k-mer when it appears in the DNA sequence. Thus to represent a DNA sequence by FCGS, we have to follow these steps:

- Considering oligomers with length of k bases, we first generate the k^{th}-order FCGR ($FCGR_k$) for the whole sequence.
- Second, we start by reading the first k-mer in the input sequence and replacing it by its frequency of occurrence that we extract from the $FCGR_k$ matrix.
- Third, we slide the pointer to the next position and we repeat the procedure for the second oligomer.
- Then, we redo the same operation till reaching the position: Pos= N-k+1, where N is the DNA sequence.

For example, we consider a sequence from the *C.elegans* chromosome III: {AGATCCGGTGTTTTGGGGAG}. Then, we enumerate the oligomers for k={1, 2 and 3}:

- Monomers={A, G, A, T, C, C, G, G, T, G, T, T, T, T, G, G, G, G, A, G} ;
- Dimers={AG, GA, AT, TC, CC, CG, GG, GT, TG, GT, TT, TT, TT, TG, GG, GG, GG, GA, AG};
- Trimers={AGA, GAT, ATC, TCC, CCG, CGG, GGT, GTG, TGT, GTT, TTT,TTT, TTG, TGG, GGG, GGG, GGA, GAG}.

After that, we compute $FCGR_1$, $FCGR_2$ and $FCGR_3$ for the whole chromosome III. According to the frequency matrices calculation, we assign the correspondent value to each of the monomers, dimers and trimers enumerated above. The figure 3 outlines the resulted signals: $FCGS_1$, $FCGS_2$ and $FCGS_3$.

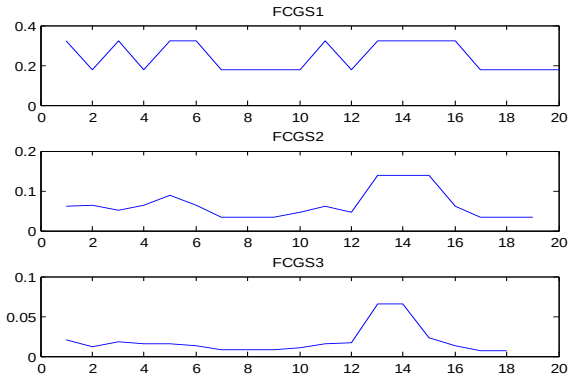

Fig. 3. Plotting $FCGS_1$, $FCGS_2$ and $FCGS_3$ for the sequence {AGATCCGGTGTTTTGGGGAG} which is taken from the *C.elegans* chromosome III.

In this way, we are able to generate a wealth of signals that map the same input sequence depending on a desired scale k. This allows investigation of a large spectrum of characteristic information within the DNA sequences, since we are based on different statistical properties of the genome.
To prove the ability of the FCGS coding in reflecting interesting information in DNA, we opt for the complex Morlet wavelet analysis.

3 Genomic sequences analysis by the Complex Morlet wavelet

Insights into DNA sequences can be gained by investigation of the time-frequency analyses. However, such sequences contain a broad scope of regular patterns with varying size and structure. Out of this fact, classical approaches based on the average of DNA contents within a fixed window (such Fourier transform), appear to be inadequate to readily and efficiently unveil information of biological interest. This is with the astonishing advent of wavelets that one can deal with this shortcoming [4].

The wavelet analysis consists on reducing a given function X(t) into its frequency and location elements by convolution with a specific small wave called the mother wavelet ψ(t). The mother wavelet dilates or compresses depending on a scale parameter "a". It also shifts accordingly through time using a shift parameter "b". Translations and expansions or dilations of ψ(t) engender a set of daughter wavelets $\psi_{a,b}(t)$. The analysis function for daughter wavelet is expressed as:

$$\psi_{a,b}(t) = \frac{1}{\sqrt{a}}\psi\left(\frac{t-b}{a}\right),\ a \succ 0,\ b \in \mathbb{R} \tag{2}$$

The layout of the time-frequency analysis of a function X(t) through the continuous wavelet transform (CWT) is defined by the following equation :

$$T_{\psi}(X)(a,b) = \frac{1}{\sqrt{a}}\int_{-\infty}^{\infty} X(t)\psi^{*}\left(\frac{t-b}{a}\right)\mathrm{d}t \tag{3}$$

The * indicates the operation of complex conjugate. Whereas, the elements $T_{\psi}(a, b)$ are called wavelet coefficients. Note that changing the scale a parameter implies variation of the analysis window length as well as the central frequency f_0 of the basic wavelet. Thus, the frequency set is proportional to scale one. Therefore, it is possible to use the frequency for representing the wavelet coefficients instead of the scale [3, 5].

Overall, CWT offers a good time-frequency resolution, which the quality is substantially influenced by the choice of the wavelet. In our analysis, we choose the Complex Morlet wavelet as a wavelet function ψ(t). This wavelet is a Gaussian-windowed complex sinusoid; which is defined as supplied from (4):

$$\psi(t) = \pi^{-\frac{1}{4}}\left(e^{i\omega_0 t} - e^{-\frac{1}{2}\omega_0^2}\right)e^{-\frac{1}{2}t^2} \tag{4}$$

Here $\omega_0=2\Pi f_0$; which corresponds to the number of oscillations of the wavelet. Note that ω_0 must be greater than 5 to satisfy the admissibility condition which is required by the CWT [6]. The absolute wavelet coefficients: $|T\psi(a,b)|$ is named amplitude scalogram. The Wavelet scalograms communicate the time frequency localization property of signals, which al-

lows easy detection of regions having specific behavior. Thus, analysis of DNA sequences using complex Morlet scalograms might be promising and fruitful direction.

4 Results and discussion

It is well known that DNA sequences reflect specific properties within and among genomes, which are defined as genomic signatures like in the case of CGR and FCGR images [19, 20]. Besides, while exploring a given genome, we can found that some sub-sequences display invariant information; which can be defined as a local signature. This exploration can be done in the time-frequency plan, which is the focus of our work. From this perspective, we consider that a DNA class, with a distinct biological function or structure, has to possess a local signature. This was demonstrated through studying the behavior of the *C.elegans* sequences with large DNA classes, from which we have chosen to expose the results relating to helitrons. Thus, we turn our attention to five types of helitron DNA which are :{ Helitron1, HelitronY1, Helitron Y1A, Helitron Y2 and Helitron Y4}. The DNA sequence and helitrons' annotations were extracted from the NCBI database [7]. As for coding, we considered the $FCGS_1$, $FCGS_2$ and $FCGS_3$ based on the FCGRs matrices of the whole chromosomes. For the complex Morlet analysis, we took a Morlet wavelet having the parameter ω_0 fixed at 5.4285 and a temporal support of 1420 points. Then, we performed the continuous wavelet transform along 64 scales. As we have calculated the wavelet coefficients for the whole chromosome, we proceeded by a mere zooming into the resulting scalograms using a window of 10^3 bp. The most striking feature observed while looking into scalograms is the characterization of helitron by sharp and distinguished periodic structures. In figure 4, we investigate the behavior of an example of helitron1_CE in the time-frequency domain when coding by the three FCGSs. This helitron belongs to the class of the autonomous TEs, which means that it encodes for a protein (CEHEL1p) [2]. As it can be seen from the related subfigures, the helitron1 is characterized by several motifs including different periodicities (like the 10 bp, 5 bp and the 4 bp periodicities). The overall periodicities are shown to be highlighted with the $FCGS_2$ coding; while the $FCGS_3$ gives more smoothed scalograms whilst keeping the main characteristic features.

In figure 5, we provide the scalograms of an helitronY1_CE (which is an non-autonomous helitron1_CE [2]) when we code with $FCGS_1$, $FCGS_2$ and $FCGS_3$. The helitronY1 scalograms exhibit common features with those of helitron1 (like the 5 bp and the 4 bp periodicities); which may indicate the existence of a paternity relationship between them. When comparing the contribution of the different FCGSs in expressing the main features of the considered helitron, we can note that $FCGS_2$ best visualizes the periodic patterns that characterize the helitronY1 sequence. Like in the case of the helitron1 sequence, the $FCGS_3$ allows a general smoothing while the main characteristics persist.

In figure 6, we present the scalograms of an helitronY1A_CE (which is also a non-autonomous TEs [2]) when it is coded by $FCGS_1$, $FCGS_2$ and $FCGS_3$. Upon the three levels of FCGS coding, we can see similar behavior. Nevertheless, it is with the $FCGS_3$ that we can clearly distinguish the periodic motifs that possess a high level of energy around frequencies 0.2 and 0.15. The latter frequencies are also noted in the previous example of heli-

tron (figure 5). This led us to think about the existence of common repetitions like DNA between the two classes of helitrons.

As for the figure 7, it shows the $FCGS_1$, $FCGS_2$ and $FCGS_3$ scalograms of an helitronY2_CE (an non-autonomous helitron_CE [2]). Through our works we have found that helitronY2 is the only sub-class that is characterized by an invariant behavior along the whole genome. These elements are shown to be easily recognized through the $FCGS_1$ scalograms. This is because this level exhibits some periodic motif with high energy on the left upper part of the scalogram. Note that these patterns become more smoothed in the level of the $FCGS_3$ scalogram.

Finally, in figure 8, we present the $FCGS_1$, $FCGS_2$ and $FCGS_3$ scalograms of an helitronY4_CE (which is also an non-autonomous helitron in *C.elegans* [2]). From the related sub-figures, we can remark the characterization of this type of DNA by more regular shaped patterns. These patterns are essentially situated on the upper frequency sub-band of the scalograms (which is limited by the frequency 0.15). In addition, even though the energy of these motifs fades when we increase the FCGS order, the main features remain remarkable.

Through these examples, we can clearly note the role played by the FCGS coding as well as he complex Morlet wavelet in characterizing helitron DNA sequences. In fact, even if in some cases the information is accentuated in one FCGS level than others; it is so easy to distinguish the information about helitrons. Thus, we can approve that an helitron structure possesses unique behavior that would be considered as a local signature.

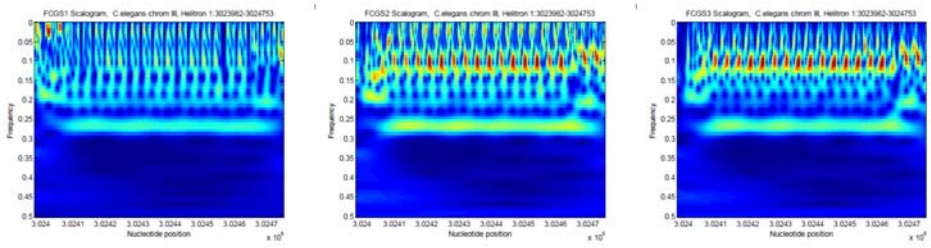

Fig. 4. Visualization of the helitron1_CE scalograms when coded by $FCGS_1$, $FCGS_2$ and $FCGS_3$.

Proceedings IWBBIO 2014. Granada 7-9 April, 2014 1441

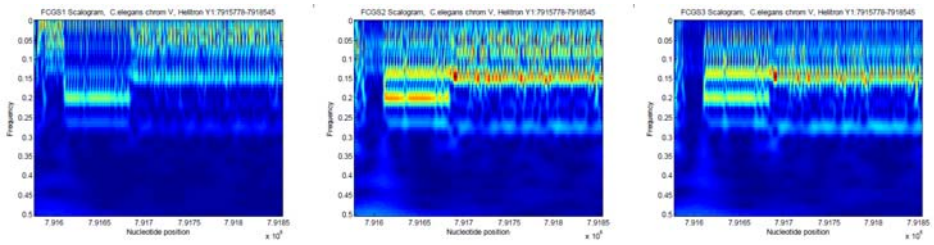

Fig. 5. Visualization of the helitronY1_CE scalograms when coded by $FCGS_1$, $FCGS_2$ and $FCGS_3$.

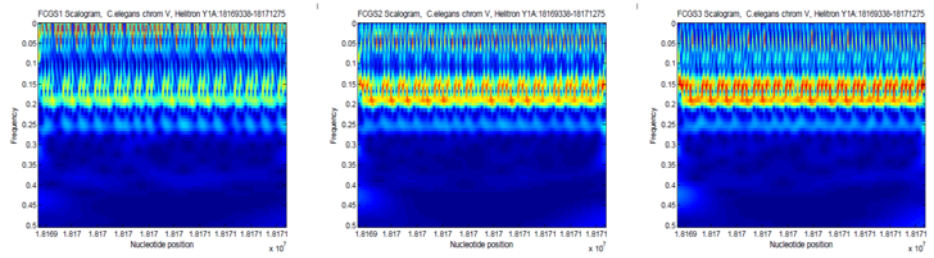

Fig. 6. Visualization of the helitronY1A_CE scalograms when coded by $FCGS_1$, $FCGS_2$ and $FCGS_3$.

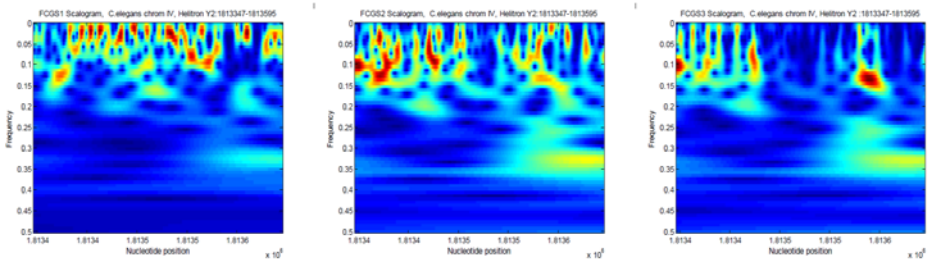

Fig. 7. Visualization of the helitronY2_CE scalograms when coded by $FCGS_1$, $FCGS_2$ and $FCGS_3$.

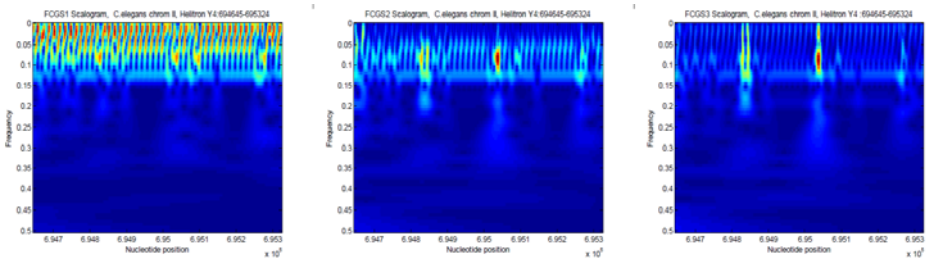

Fig. 8. Visualization of the helitronY4_CE scalograms when coded by $FCGS_1$, $FCGS_2$ and $FCGS_3$.

Proceedings IWBBIO 2014. Granada 7-9 April, 2014 1442

5 Conclusion

Many studies have been carried out to unveil hidden structures, having biological interest, within the DNA sequences. Thus, different signal processing algorithms have been employed to reach this goal. In this paper, we reviewed the role of the continuous wavelet transform in exploring specific structural entities among DNA sequences; which are helitrons. To further characterize helitron DNA in *C.elegans*, we have applied the complex Morlet wavelet transform to the Frequency Chaos Game Signals (FCGSs). The Frequency Chaos Game Signal is a new coding technique which we designed based on the chaotic dynamical systems.

The basic concept of our coding (FCGS) consists on assigning numerical values to each of the DNA letters based on the Frequency Chaos Game Representations (FCGRs). The particularity of this technique resides in using the statistical properties of the genome itself, which may strongly reflect main interesting features of the specific DNA structures. Application of the complex Morlet analysis has proven its highly aptitude to depict helitron features independently of the adopted FCGS's scale; which attest the validity of the genomic signature concept in the helitrons Frequency Chaos Game Signals. In fact, the associated color scalograms were demonstrated to easily distinguish helitron elements by very special textures. These textures include a wide number of periodic patterns surrounding different frequencies. In the other hand, each of these five helitron classes {Helitron1, HelitronY1, Helitron Y1A, Helitron Y2 and Helitron Y4} is shown to possess a unique signature. Hence, it is interesting to use these results to establish new algorithms which aim at classifying and recognizing helitrons in the genomes.

6 References

1. V. Kapitonov, J. Jurka,). "Helitrons on a roll: eukaryotic rolling-circle transposons". Trends in Genetics, Vol. 23, No. 10, 2007.
2. (2013) The Repbase database. [Online]. Available: http://www.girinst.org/repbase/.
3. M. Steinbuch and M.J.G. van de Molengraft, Eindhoven University of Technology, Control Systems Technology Group Eindhoven, “Wavelet Theory and Applications”, a literature study, R.J.E. Merry, DCT, 2005.
4. A .Grossmann and J. Morlet, “Decomposition of Hardy functions into square integrable wavelets of constant shape,” SIAM: Journal on Mathematical Analysis,” Vol. 15, pp.723-736, 1984.
5. N.C.F. Tse and L.L. Lai, “Wavelet-Based Algorithm for Signal Analysis,” EURASIP Journal on Advances in Signal Processing, 2007.
6. A.H. Najmi and J. Sadowsky, “The Continuous Wavelet Transform and Variable Resolution Time–Frequency Analysis,” Johns Hopkins Apl Technical Digest, Vol. 18, No.1, 1997.
7. (2013) The NCBI GenBank database. [Online]. Available: http://www.ncbi.nlm.nih.gov/Genbank/.
8. A. Fiser, G. E. Tusnady and I. Simon, “Chaos game representation of protein structures,” J.Mol Graphics, Vol. 12, pp. 295-304, 1994.
9. P. J. Mena-Chalco, H. Carrer, Y. Zana, and R. M. Cesar, “Identification of Protein Coding Regions Using the Modified Gabor-Wavelet Transform,” IEEE, ACM /TCBB, Vol. 5, No.2, 2008.

10. http://bioinform.genetika.ru/projects/reg/wavelets/wavelets.html .
11. Y. Qi, N. Jin and D. Ai, "Wavelet Analysis of DNA Walks on the Human and Chimpanzee MAGE/CSAG-palindromes," Elsevier, Genomics Proteomics Bioinformatics, Vol.10, pp. 230-236, 2012.
12. A. Arneodo, F. Argoul, E. Bacry, J. Elezgaray, and J. F. Muzy, "Ondelettes, Multifractales et Turbulence : de l'ADN aux Croissances Cristallines, " Art et Sciences, Paris, 1995.
13. K. D. Rao and M. N. S. Swamy, "Analysis of Genomics and Proteomics Using DSP Techniques," IEEE Transactions on circuits and systems, Vol. 55, No. 1, 2008.
14. E. Pirogova, Q. Fang, M. Akay and I. Cosic, "Investigation of the Structural and Functional Relationships of Oncogene Proteins," Proceedings of the IEEE, Vol. 90, No. 12, pp. 1859-– 1867, 2002.
15. H. K. Kwan, R. Atwal and B. Y. M. Kwan, "Wavelet Analysis of DNA Sequences," Proceedings of International Conference on Communications, Circuits and Systems (ICCCAS), pp. 917-921, 2008.
16. A. Arneodo, "Long-range correlations in genomic DNA: a signature of the nucleosomal structure", Phys Rev Lett 86:2471–4, 2001.
17. Berger, J.A., Mitra, S.K., Carli, M.,"Visualization and analysis of DNA sequences using DNA walks," J. Franklin Inst. 341, 2004, pp. 37-53.
18. Y. W. Wang, K. Hill, S. Singh and L. Kari, "The spectrum of genomic signatures: from dinucleotides to chaos game representation," Gene, Vol. 346, pp. 73-185, 2005.
19. J. S. Almeida, J. A. Carrico, A. Maretzek, P. A. Noble, M. Fletcher M, "Analysis of genomic sequences by Chaos Game Representation," Bioinformatics, Vol. 17, No. 5, pp.429-37, 2001.
20. P. Deschavanne, A. Giron, J. Vilain, CH. Dufraigneand and B. Fertil, "Genomic Signature Is Preserved in Short DNA Fragment," International Symposium on Bio-Informatics and Biomedical Engineering, IEEE, 2000, pp. 161-167.
21. L. Yang and J.L. Bennetzen, "Structure-based discovery and description of plant and animal Helitrons", PNAS Vol. 106 no. 31, pp. 12832–12837.
22. M.Sweredoski, L. DeRose-Wilson and B.S. Gaut, "A comparative computational analysis of nonautonomous Helitron elements between maize and rice, BMC Genomics 9, pp. 467- 479, 2008.
23. Ch. Du, J. Caronna, L. He and H.K. Dooner, "Computational prediction and molecular confirmation of Helitron transposons in the maize genome", BMC Genomics 9, pp. 51-60, 2008.
24. Q. Wang and H.K. Dooner, "Remarkable variation in maize genome structure inferred from haplotype diversity at the bz locus", Proc Natl Acad Sci, 103, pp.17644-17649, 2006.
25. S.Tempel, J. Nicolas,A.E. Amranl and I. Couee, "Model-based identification of Helitrons results in a new classification of their families in Arabidopsis thaliana", Gene 403, pp. 18-28, 2007.
26. S. Tempel, "Dynamique des hélitrons dans le génome d'arabidopsis thaliana: développement de nouvelles stratégies d'analyse des éléments transposables", PHD Thesis, IRISA, Université de Rennes I, 2007.