Communications in Information and Systems
Volume 19, Number 4, 375–390, 2019

Comparisons of BRAT1 gene variants

En-Bing Lin and Eric Linton

BRAT1 (BRACA1 Associated ATM Activator 1) is a protein coding gene that interacts with and activates the tumor suppressor gene BRCA1, a protein complex that repairs DNA damage due to ionizing radiation. In this paper, we obtain approximation and detail information of the numerical representation, i.e. wavelet transformation with Daubechies 2 and Coiflet, of the gene variants. We compare the computational and graphical results of these gene variants to each other, as well as to those obtained from BLAST (Basic Local Alignment Search Tool). Thus, we can compare the results of an alignment free search method (wavelet) to that of an alignment dependent search method (BLAST). In general, both methods located the same areas of similarity but the results of the wavelet analysis provide numerical and visual comparisons. We also conclude with some other advantages of using wavelet method.

1. Introduction

Over the past sixty years, deaths from cancer have changed little. In fact, cancer is among the leading causes of death worldwide. Major cancer types are breast, lung, prostate and colorectal cancers. Most inherited cases of breast cancer are associated with mutations in two genes: BRCA1 (BReast CAncer gene one) and BRCA2 (BReast CAncer gene two). The function of the BRCA genes is to repair cell damages and maintain normal tissue growth and function. The BRAT1 is part of the repair DNA (Deoxyribonucleic Acid) mechanism associated with breast cancer suppressor gene BRCA1 and the ATM (ataxia telangiectasia mutated) protein, which helps repair damaged DNA. ATM is thought to be a master controller of cell cycle checkpoint signaling pathways that are required for cellular responses to DNA damage such as double-strand breaks that are induced by ionizing radiation and complexes with BRCA1. BRCA1 expression is reduced or undetectable in the majority of high grade, ductal breast cancers. It has long been noted that loss of BRCA1 activity, either by germ-line mutations or by down-regulation of gene expression, leads to tumor formation in specific target tissues [34].

Our ultimate goal is to extract valuable information from a large amount of biological data. Several clustering and other techniques are applied to DNA and protein sequences [6, 9, 15, 35], by which one can correlate the inherent relationships between DNA or protein sequences. Our approach is to apply digital signal processing techniques which can be used to characterize genomic data more efficiently in comparison to other methods [2–5, 7, 8, 12, 40, 43]. In particular, wavelet techniques in signal processing have been used in various applications in biosciences and medical areas [1, 13–30, 33, 36, 42]. The success of these applications is mainly due to robustness, efficiency and flexibility of wavelets' characteristics [31]. We address the DNA code from humans in the perspective of wavelet signal processing. DNA is a double helix constituted by two polymers connected by hydrogen atoms. The polymers contain three parts to a nucleotide, namely deoxyribose, a phosphate group, and a nitrogenous base. There are four distinct nitrogenous bases: thymine, cytosine, adenine, and guanine, denoted by the symbols {T, C, A, G}. To apply wavelet techniques, we need to map DNA sequences into mathematical representations, which include binary coding [39, 44], complex number [7], integer number [11], EIIP (electron ion interaction potential) [30], graphical representation [45], Z-curves [46–48]. Other models, such as DNA walks [37], are also available. The integer representation appears to be useful and effective [10, 32, 41]. In this paper, we will use the integer representation, namely, we map integer numbers to the four nucleotides as T=0, C=1, A=2, and G=3. We perform wavelet analysis on several gene variants by using Daubechies (db2) wavelet at level 5 and Coiflet 1 at level 3. We will also use, BLAST [38], an alignment tool to work on the gene variants. We will then compare the results obtained by different tools.

Our paper is organized as follows. In Section 2, we provide some mathematical backgrounds in wavelet analysis. We then perform BLAST analysis and wavelet analysis on BRAT1 gene and its variants in Section 3. Some computational comparison results are presented in Section 4. We conclude with several comments in Section 5.

2. Mathematical backgrounds

In this section, we provide a brief review on the wavelet analysis by recalling multiresolution analysis, scaling functions, wavelet functions, as well as continuous and discrete wavelet transforms. A multiresolution analysis of a function space (or an object) is to provide a detailed analysis of that function. It also analyzes the function with detailed decompositions. Moreover, it provides a multi-level of approximation of the function such that one can not

only relate different levels of the approximation but also provide a sequential relationship among different levels of the approximation. In the process of approximating or decomposing the function, one can have different choices of generators which play the role of tools in obtaining the approximations or decompositions as desired. More precisely, the formal definition is described as follows.

A multiresolution analysis (MRA) [31] consists of a sequence of successive approximation spaces $\{V_j\}_{j\in Z}$ of $L^2(R)$ with the following properties:

(i) $V_j \subset V_{j+1}$,
(ii) $\lim_{j\to\infty} V_j = \bigcup_{j\in Z} V_j$ is dense in $L^2(R)$,
(iii) $\bigcap_{j\in Z} V_j = \{0\}$,
(iv) $f(x) \in V_j \Longleftrightarrow f(2x) \in V_{j+1}$,
(v) $f(x) \in V_j \Longleftrightarrow f(x + 2^{-j}k) \in V_j, \ \ \forall k \in Z$,
(vi) There exists a function $\phi \in V_0$ so that $\{\phi(x-j)\}_{j\in Z}$ is an orthonormal basis of V_0.

Where ϕ is called a *scaling function* that generates an MRA with the above properties. Through translation and dilation of ϕ, a Riesz basis $\{\phi_{j,k}(x)\}_{k\in Z}$ is obtained for the subspace $V_j \subset L^2(R)$ by the properties (iv)(v), where

$$\phi_{j,k}(x) = 2^{\frac{j}{2}}\phi(2^j x - k), \quad j,k \in Z. \tag{1}$$

More generally, this family can be expressed as $\phi_{m,n}(x) = \frac{1}{a^{\frac{m}{2}}}\phi(\frac{x-nb}{a^m})$ by using nonzero real numbers a and real number b.

In what follows, we describe dilation and wavelet equations. Since $V_0 \subset V_1$, there is a set of coefficients $\{a_k\}_{k\in Z}$, so that ϕ satisfies the two–scale equation or refinement equation

$$\phi(x) = \sum_k a_k \phi(2x - k). \tag{2}$$

For every $j \in Z$, we define W_j to be the orthonormal complement of V_j in V_{j+1}, we then have

$$V_{j+1} = V_j \bigoplus W_j \tag{3}$$

and

$$W_j \perp W_{j'} \ \ if \ \ j \neq j'. \tag{4}$$

It follows that, for $j > J$,

$$V_j = V_J \bigoplus (\bigoplus_{k=0}^{J-j+1} W_{J-k}). \tag{5}$$

By virtue of (ii) and (iii) above, this implies

$$L^2(R) = \bigoplus_{j \in Z} W_j \tag{6}$$

which is a decomposition of $L^2(R)$ into mutually orthogonal subspaces. It turns out that a basis for W_0 can be obtained by dilating and translating a single function $\psi(x)$ called basic (mother) wavelet which is defined by (wavelet equation)

$$\psi(x) = \sum_k b_k \phi(2x - k) \tag{7}$$

where $b_k = (-1)^k a_{-k+1}$. In fact, $\{\psi_{j,k}(x) = 2^{\frac{j}{2}} \psi(2^j x - k)\}_{k \in Z}$ forms an orthonormal basis for W_j.

Let $P_j,\ Q_j$ denote the orthogonal projection $L^2 \to V_j,\ L^2 \to W_j$, respectively. Then

$$P_j f(x) = \sum_k \alpha_{j,k} \phi_{j,k}(x), \tag{8}$$

$$Q_j f(x) = \sum_k \beta_{j,k} \psi_{j,k}(x), \tag{9}$$

where the coefficients $\alpha_{j,k},\ \beta_{j,k}$ are given by the inner product:

$$\alpha_{j,k} = <f,\ \phi_{j,k}> = \int_{-\infty}^{\infty} f(x) \phi_{j,k}(x) dx, \tag{10}$$

$$\beta_{j,k} = <f,\ \psi_{j,k}> = \int_{-\infty}^{\infty} f(x) \psi_{j,k}(x) dx. \tag{11}$$

$P_j f$ converges to f in the L^2 norm which is the best approximation of f in V_j.

In fact, the above coefficients can also be obtained by applying wavelet transforms which are defined as follows.

The continuous wavelet trasnform is defined as:

$$[w_\psi x(t)](a, b) = \frac{1}{\sqrt{a}} \int_{-\infty}^{\infty} x(t) \psi^*(\frac{t - b}{a}) dt \quad a > 0, b \in R, \tag{12}$$

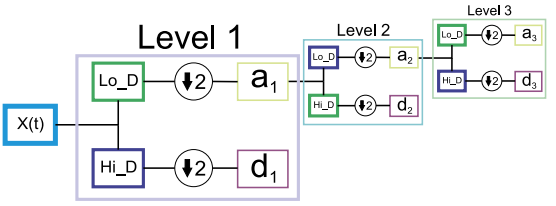

Figure 1: Discrete Wavelet Transform Deconstruction Method.

where the symbol * represents the complex conjugate, $x(t)$ is the given signal (DNA sequence) and ψ is a wavelet.

The discrete wavelet transform is defined as:

$$[Dw_\psi x(n)](a,b) = \sum_{n\in Z} x(n) g_{j,k}(n), \quad a = 2^j, b = k2^j, j \in N, k \in Z, \tag{13}$$

where g's are the coefficients of the wavelet equation associated with ψ. Through the discrete wavelet transform, Figure 1 shows that the approximation and detail information of the given signal are successively decomposed into different levels of resolutions.

Each signal can be represented by its wavelet coefficients by choosing different wavelets. Depending on the nature of the signals and the characteristics of wavelets, the choices of wavelets may give rise to different outcomes for the same applications. In this paper, we use Daubechies wavelet and Coiflet. The Daubechies wavelets are a family of orthogonal wavelets defining a discrete wavelet transform and characterized by a maximal number of vanishing moments for some given support. Daubechies wavelets extend the Haar wavelets by using longer filters, that produce smoother scaling functions and wavelets. A high number of vanishing moments allows bettering compressing regular parts of the signal. However, increasing the number of vanishing moments also increases the size of the support of the wavelets. Coiflets have both scaling functions and wavelet functions with vanishing moments. The wavelet is near symmetric. The main differences between these wavelet functions are due to the vanishing moments and the symmetry.

3. BRAT1 and its variants

The variations of this gene may be associated with breast cancer. It is important to analyze the sequences so that we can find out more about the underlying molecular mechanisms. Recognizing the relevant genetic susceptibility

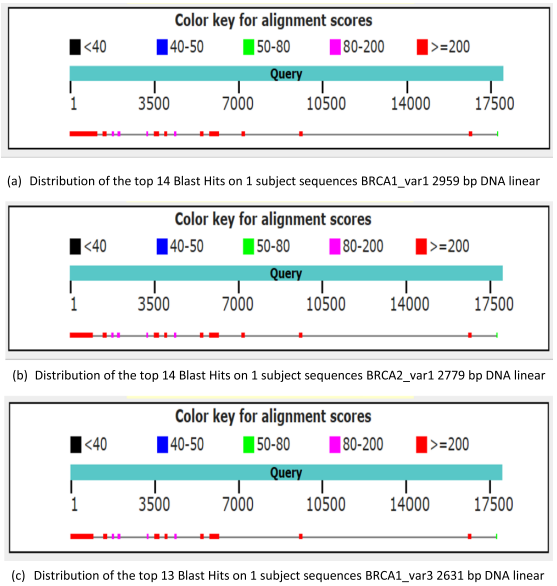

Figure 2: BLAST results.

would help in counseling presymptomatic individuals to adopt preventive and control measures to delay the onset of disease. Therefore it is essential to analyze for gene variants that may be associated with breast cancer. It would help progress faster the diagnostic treatment. Here, we consider three variants. Variant 1 represents the longest transcript and encodes the longest isoform 1. Variant 2 uses an alternate in-frame splice junction in the 3' end of the coding sequence compared to variant 1. The resulting isoform 2 has the same N- and C-termini but is shorter compared to isoform 1. Variant 3 lacks an alternate exon and uses an alternate in-frame splice junction in the 3' coding sequence compared to variant 1. The resulting isoform 3 is shorter at the N-terminus and lacks an internal segment compared to isoform 1.

Searching for similarities between biological sequences is the principal means by which bioinformatics contributes to our understanding of biology. Of the various informatics tools developed to accomplish this task, the most widely used is BLAST, which directly approximates alignments that optimize a measure of local similarity. The results of comparing BRAT1 and variants obtained by BLAST are shown in Figures 2 and 3. In Figure 3 the dot plot shows the sequence of the variants on the y-axis plotted against the sequence of the BRAT1 gene on the x-axis. As the two sequences match in the same direction they form a series of line slanted from the bot-

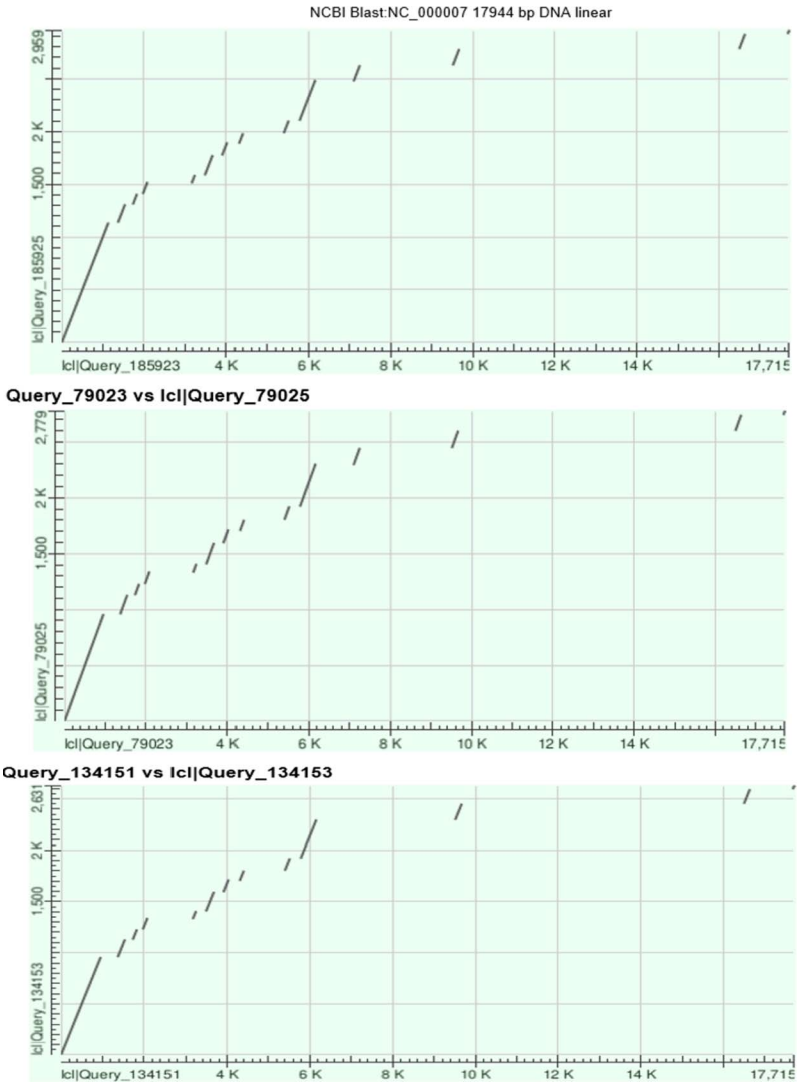

Figure 3: Dot Plot of BLAST results.

tom left to the upper right. The breaks in the lines indicate where there is no match and thus where the exons are located. Next, we performed wavelet analysis of these sequences with different wavelets. Basically, the algorithm uses equation (10) to obtain approximations and uses equation (12) or (13) to obtain continuous and discrete decompositions. Therefore, we provide approximation and detail information of individual genes. Wavelet

analysis and wavelet transform of BRAT1 with Daubechies 2 and Coiflet are shown in Figures 4 and 5 respectively, followed by analysis of three variants in Figures 6–11. Daubechies wavelets are basically smoother than Haar wavelets, while Coiflets posess some symmetry. We also divide the BRAT1 gene (introns and exons) into six segments and perform wavelet analysis for each segment with similar size to those three variants (Figures 12–23). Figures 4–7 are shown below. The rest of the figures are in the supplementary materials http://intlpress.com/site/pub/files/_supp/cis/2019/0019/0004/CIS-2019-0019-0004-s002.pdf. All calculations are done in Matlab. Thus, we have used both alignment and alignment-free methods to analyze sequences, so we can do some camparisons by using different approaches to analyze the signals.

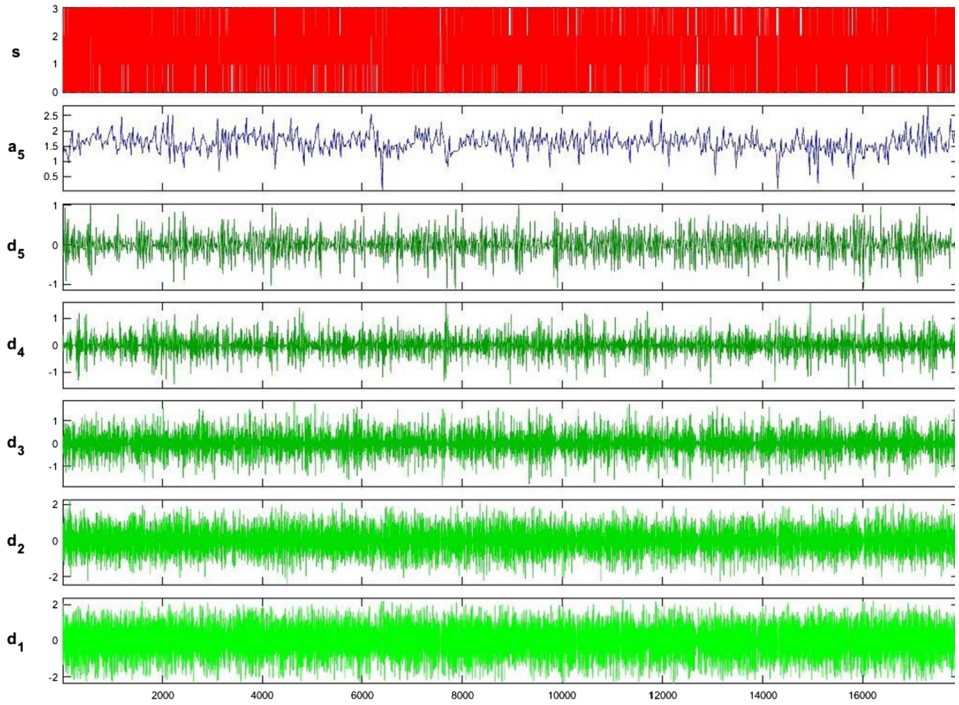

Figure 4: Approximation and Detail of BRAT1 with db2.

4. Comparisons

Each gene's overall wavelet coefficients can be characterized by its nomalized values by using the following global comparison formula [25];

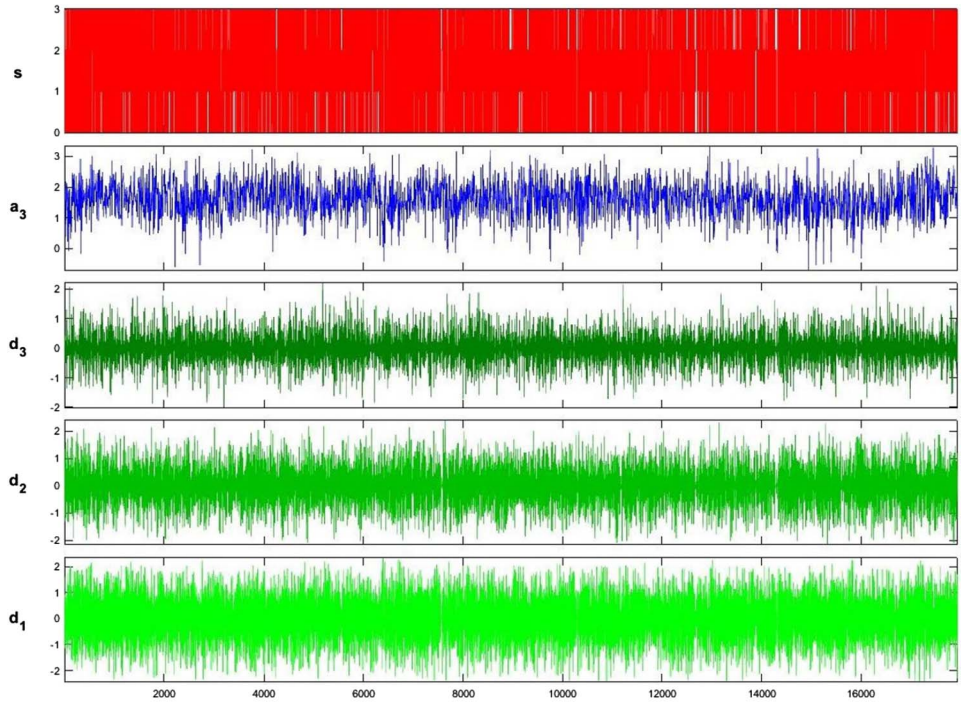

Figure 5: Approximation and Detail of BRAT1 with Coiflet.

$N(a) = \frac{w(a,b)}{max(abs(w(a,b)))}$, where $w(a,b)$ is wavelet transform defined in equation (12). Alternatively, we can calculate wavelet variance to understand the comparisons among different genes. The variance is defined as $V(a) = \frac{1}{n}\sum_{j=1}^{n} w^2(a, x_j)$, where $w(a, x_j)$ are wavelet coefficients [26]. Their values are listed in Table 1.

To measure the disorderliness or randomness in a close system, we use entropy which is considered as a measure of uncertainty. It is defined as $-\sum w_j log_2 w_j$, where w_j are wavelet coefficients [26]. We present the corresponding entropy for each gene in Table 2.

Individuals of a species have similar characteristics but they are rarely identical, the difference between them can be described by variance. On the other hand, entropy is a measure of information content and complexity. It is a standard measure for the order state of symbol sequences and provides the average rate at which information is produced by a stochastic source of data. From the above tables, we can compare the similarities among the genes. In fact, both Table 1 and Table 2 show significant similarities for each wavelet among Variants and Part 1.

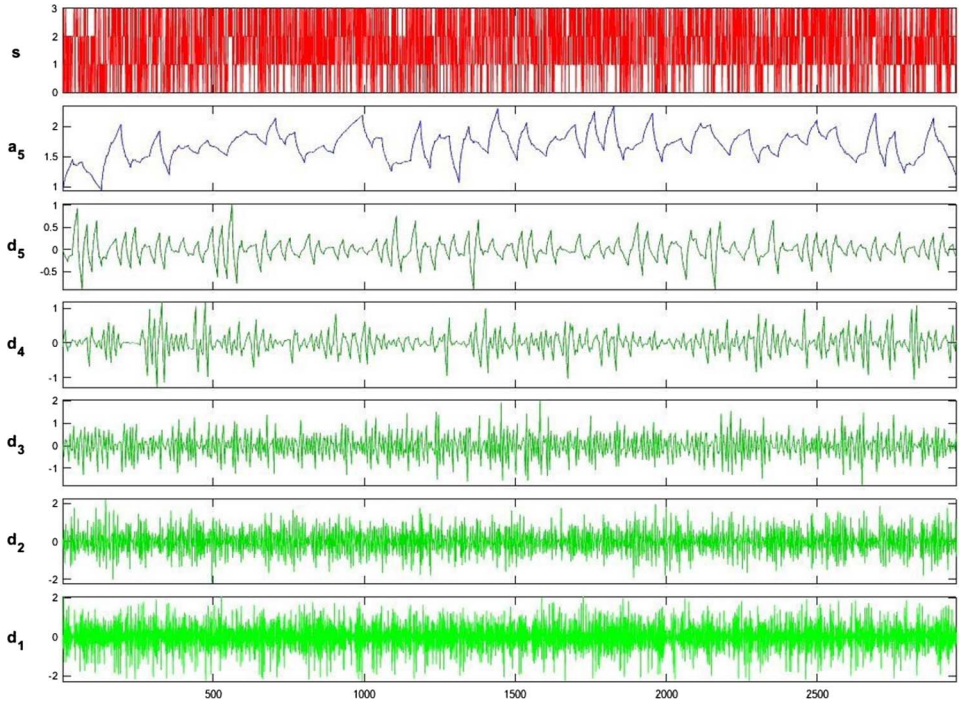

Figure 6: Approximation and Detail of Variant1 with db2.

Table 1: Variance with db2 and Coiflet Wavelet Coefficients of Genes

Genes	Variance with db2	Variance with Coiflet
BRAT1	3.7619	3.7600
Variant 1	4.0900	4.0722
Variant 2	4.0832	4.0691
Variant 3	4.0701	4.0502
Part 1	3.8296	3.8084
Part 2	4.1503	4.0315
Part 3	3.8011	3.7764
Part 4	3.9286	3.8839
Part 5	3.5685	3.5656
Part 6	3.6200	3.6183

5. Conclusions

We have presented DNA analysis by using both alignment and alignment free tools. As we observe the above distributions, figures and data, it turns out that both results agree to some extent. Wavelet representations of the

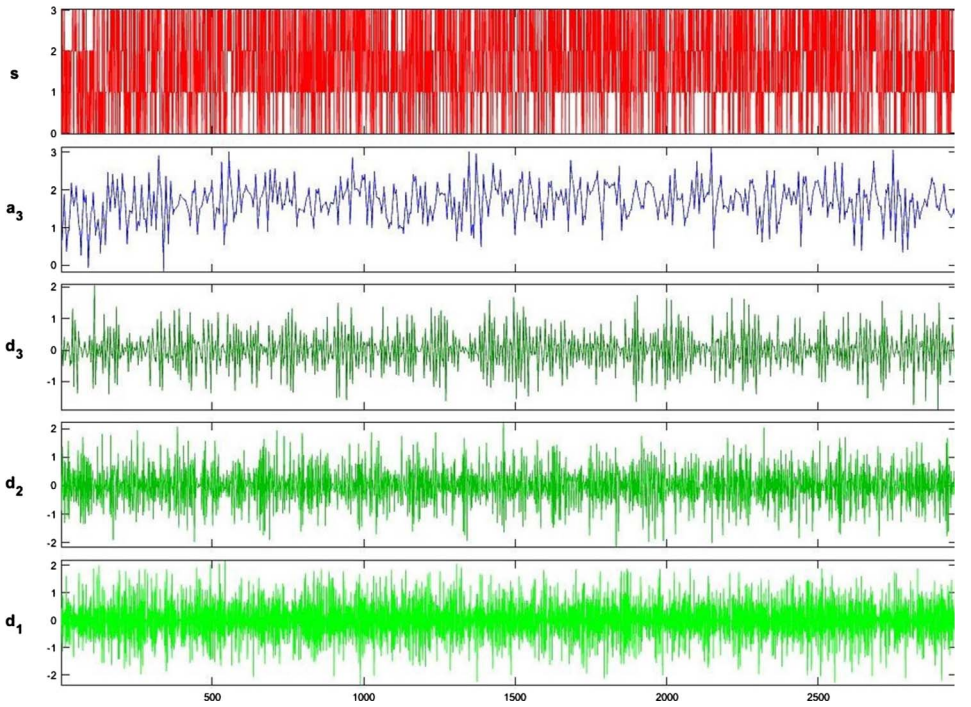

Figure 7: Approximation and Detail of Variant1 with Coiflet.

Table 2: Entropy with db2 and Coiflet Wavelet Coefficients of Genes

Genes	Entropy with db2	Entropy with Coiflet
BRAT1	−11768	−12243
Variant 1	−2136	−2114.6
Variant 2	−1923.9	−2079.1
Variant 3	−1782.9	−1958
Part 1	−1883.1	−1993.8
Part 2	−3926.4	−5316.9
Part 3	−2777.7	−3345
Part 4	−2457	−3783.5
Part 5	−1677.9	−2530.6
Part 6	−1428.5	−979.9

variants shown in the figures, that help the comparisons by visulizations of the original sequences and at different levels of decompositions. There are some limitations of BLAST, for example, each search with BLAST is a single query against a single subject, and if multiple queries need to be

searched then multiple operations are needed. Additionally, BLAST is more computationally intensive when dealing with more complex queries. Wavelet analysis is a useful tool to analyze, decompose and characterize signals. We have presented computational, graphical and conceptional illustrations on several aspects of developments. Moreover, with some advanced techniques, wavelet can perform multiple comparisons without an increase in computational complexity. In general, both BLAST and wavelet methods located the same areas of similarity but the results of the wavelet analysis provide numerical and visual comparisons. These can be further studied, namely, other methods of numerical representations, other wavelets, coding and noncoding regions of different genes, regulation analysis and improvements of diagnosis, treatment and prevention of the disease. Ultimately, such analysis could help establish a more precise diagnosis and treatment for patients.

Acknowledgment

The authors would like to thank the anonymous referees' valuable comments.

References

[1] O. Abbasi, A. Rostami, and G. Karimian, *Identification of exonic regions in DNA sequences using cross-correlation and noise suppression by discrete wavelet transform.* BMC Bioinform, **12**, 1–10, 2011.

[2] M. Akhtar, J. Epps, and E. Ambikairajah, *On DNA numerical representations for period-3 based exon prediction.* In: Proceedings of IEEE International Workshop on Genomic Signal Processing and Statistics, GENSIPS, Tuusula Finland 1–4, 2007.

[3] M. Akhtar, J. Epps, and E. Ambikairajah, *Signal processing in sequence analysis: advances in eukaryotic gene prediction.* IEEE J Sel. Top. Signal Process., **2**(3), 310–321, 2008.

[4] D. Anastassiou, *Genomic signal processing.* IEEE Signal Process. Mag., **18**(4), 8–20, 2001.

[5] A. Arneodo, Y. D. Aubenton-Carafa, B. Audit, E. Bacry, F. F. Muy, and C. Thermes, *What can we learn with wavelets about DNA sequences*? Phys. A., **249**, 439–448, 1998.

[6] J. P. Bao and R. Y. Yuan, *A wavelet-based feature vector model for DNA clustering.* Genetics and Molecular Research, **14**(4), 19163–19172, 2015.

[7] J. A. Berger, S. K. Mitra, M. Carli, and A. Neri, *New approaches to genome sequence analysis based on digital signal processing*. In: Proceedings of IEEE Workshop on Genomic Signal Processing and Statistics (GENSIPS), Raleigh, NC, 1–4, 2002.

[8] J. A. Berger, S. K. Mitra, M. Carli, and A. Neri, *Visualization and analysis of DNA sequences using DNA walks*. J. Framkl. Inst., **341**, 37–53, 2004.

[9] B. Bhosale, B. S. Ahmed, and A. Biswas, *Wavelet Based Analysis in Bio-informatics*. Life Science Journal, **10**(2), 853–859, 2013.

[10] W. Chen, B. Liao, Y. Liu, W. Zhu, and Z. Su, *A Numerical Representation of DNA Sequences and Its Applications*. Communications in Mathematical and in Computer Chemistry, **60**, 291–300, 2008.

[11] P. D. Cristea, *Genetic signals: an emerging concept*. In: Proceedings of 8th International Workshop on Systems, Signals and Image Processing, Bucharest, Romania, 17–22, 2001.

[12] P. D. Cristea, *Genetic signals representation and analysis*. In: Proceedings of SPIE 4623, Functional Monitoring and Drug-Tissue Interaction, **4623**, 77–84, 2002. doi:10.1117/12.491244.

[13] P. D. Cristea, *Genomic Signal Processing and Statistics*. Eurasip Book Series on Signal Processing and Communications, Hindawi Publishing Corporation, 15–65, 2005.

[14] P. D. Cristea, R. Tuduce, C. Banica, and K. Rodewald, *Genomic signals for the study of multiresistance mutations in M Tuberculosis*. In: Proceedings of International Symposium on Signals Circuits and Systems, Romania, **1**, 1–4, 2007. doi:10.1109/ISSCS.2007.4292708.

[15] A. Dakhli, W. Belli, and C. B. Amar, *DNA Sequence Classification Using Power Spectruma and Wavelet Neural Network*. In: Proceedings of the 16th International Conference on Hybrid Intelligent Systems, 391–402, HIS 2016.

[16] G. Dodin, P. Vandergheynst, P. Levoir, C. Cordier, and L. Marcour, *Fourier and wavelet transform analysis a tool for visualising regular patterns in DNA sequences*. J. Theor Biol., **206**(7), 323–326, 2000.

[17] T. P. George and T. Thomas, *Discrete wavelet transform de-noising in eukaryotic gene splicing*. BMC Bioinform., **11**(Supp 1), S50, 2010. doi:10.1186/1471-2105-11-S1-S50.

[18] A. D. Haimovich, B. Byrne, R. Ramaswamy, and W. J. Welsh, *Wavelet analysis of DNA walks*. J. Comput. Biol., **13**(7), 1289–1298, 2006.

[19] Y. Hur and H. Lee, *Wavelet-based identification of DNA focal genomic aberrations from single nucleotide polymorphism arrays*. BMC Bioinform, 2011. doi:10.1186/1471-2105-12-146.

[20] T. M. Inbamalar and R. Sivakumar, *Improved algorithm for analysis of DNA sequences using multiresolution transformation*. Sci. World J., **2015**, 2015. Article ID 786497, doi:10.1155/2015/786497.

[21] A. Krishnan, K. B. Li, and P. Isaac, *Rapid detection of conserved regions in protein sequences using wavelets*. Silico Biol., **4**(2), 133–148, 2004.

[22] H. K. Kwan and S. B. Arniker, *Numerical representation of DNA Sequences*. IEEE International Conference, 307–310, 2009.

[23] E. B. Lin and E. Linton, *Wavelet Packet Analysis of DNA sequences*. In: Proceedings of the 5th International Conference on Bioinformatics and Biomedical Engineering, 2011.

[24] E. B. Lin, M. Haske, M. Smith, and D. Sowards, *Wavelet analysis of ECG Signals, multiscale Signal Analysis and Modeling*. Lecture Notes in Electrical Engineering, Springer, New York, 233–256, 2013.

[25] E. B. Lin and M. Oyapero, *Wavelet analysis of coding and noncoding regions of DNA sequences using integer representation*. Journal of Computational Intelligence in Bioinformatics, **5**(2), 149–159, 2012.

[26] E. B. Lin and M. Oyapero, *Wavelet Variance and Entropy of DNA Sequences International Journal of Computer and Communication Engineering*. IACSIT, Singapore, 70–72, 2012.

[27] E. Linton, P. Albee, P. Kinnicutt, and E. B. Lin, *Time Frequency Analysis of Mitochondrial Genome*. In: Proceedings of the 2010 International Conference on Bioinformatics and Computational Biology, 293–297, 2010.

[28] E. Linton, P. Albee, P. Kinnicutt, and E. B. Lin, *Multiresolution Analysis of DNA Sequences*. In: Proceedings of Second International Conference on Computer Research and Development, 218–221, 2010.

[29] P. Lio and M. Vannucci, *Finding pathogenicity islands and gene transfer events in genome data*. Bioinformatics, **16**(10), 932–940, 2000.

[30] J. A. T. Machado, A. C. Costa, and M. D. Quelhas, *Wavelet analysis of human DNA*. Genomics, **98**(3), 155–163, 2011.

[31] S. Mallat, *A Wavelet Tour of Signal Processing*. Academic Press, 1998.

[32] G. Mendixabal-Ruiz, I. Roman-Godinez, S. Torres-Ramos, R. A. Salido-Ruiz, and J. A. Morales, *On DNA numerical representations for genomic similarity computation*. PLos ONE, **12**(3), e0173288, March 21, 2017.

[33] K. B. Murray, D. Gorse, and J. M. Thornton, *Wavelet transforms for the characterization and detection of repeating motifs*. J. Mol. Biol., **316**(2), 341–363, 2002.

[34] National Center for Biotechnology Information (NCBI). http://www.ncbi.nlm.nih.gov.

[35] N. Nguyen, A. Vo, I. Choi, and K.-J. Won, *A Stationary Wavelet Entropy-Based Clustering Approach Accurately Predicts Gene Expression*. Journal of Computational Biology, **22**(3), 236–249, 2015. doi:10.1089/cmb.2014.0221.

[36] J. Ning, C. N. Moore, and J. C. Nelson, *Preliminary wavelet analysis of genomic sequences*. In: Proceedings of the IEEE Computer Society Conference on Bioinformatics, 509–510, Stanford, CA 2003.

[37] C. K. Peng, S. V. Buldyrev, A. V. Goldberger, S. Havlin, F. Sciortino, M. Simons, and H. E. Stanley, *Long-range correlations in nucleotide sequences*. Nature, **356**, 168–170, 1992.

[38] A. Pertsemlidis and J. W. Fondon III, *Having a BLAST with bioinformatics (and avoiding BLAST phemy)*. Genome Biology, **2**(10), 2010.

[39] R. Ranawana and V. Palade, *A neural network based multi-classifier system for gene identification in DNA sequence*. Neural Comput. Appl., **14**(2), 122–131, 2005.

[40] K. D. Rao and M. n. S. Swamy, *Analysis of genomics and proteomics using DSP techniques*. IEEE Trans. Circuits Syst.-I, **55**(1), 370–378, 2008.

[41] S. Saini and L. Dewan, *Comparison of Numerical Representations of Genomic Sequences: Choosing the Best Mapping for Wavelet Analysis*. International Journal of Applied and Computational Mathematics, **3**(4), 2943–2958, 2017.

[42] A. A. Tsonis, P. Kumar, J. B Elsner, and P. A. Tsonis, *Wavelet analysis of DNA sequences*. Phys. Rev. E Stat. Phys. Plasmas Fluids Relat. Interdiscip. Top., **53**(2), 1828–1834, 1996.

[43] P. P. Vaidyanathan and B. J. Yoon, *The role of signal-processing concepts in genomics and proteomics*. J. Framkl. Inst., **341**, 111–135, 2004.

[44] R. F. Voss, *Evolution of long-range fractal correlations and l/f noise in DNA base sequence*. Phys. Rev. Lett., **68**, 3805–3808, 1992.

[45] S. S.-T. Yau, J. Wang, A. Niknejad, C. Lu, N. Jin, and Y. K. Ho, *DNA sequence representation without degeneracy*. Nucleic acids research, **31**(12), 3078–3080, 2003.

[46] M. Q. Zhang, *Identification of protein coding regions in the human genome by quadratic discriminant analysis*. Proc. Natl Acad. Sci. USA, 565–568.

[47] R. Zhang and C. T. Zhang, *Identification of replication origins in archaeal genomes based on the curve method*. Archaea, **1**, 335–346, 2005.

[48] R. Zhang and C. Zhang, *A Brief Review: The Z-curve Theory and its Application in Genome Analysis*. Current Genomics, **15**, 78–94, 2014.

En-Bing Lin
Department of Mathematics
Central Michigan University
Mt. Pleasant, MI 48859
USA
E-mail address: lin1e@cmich.edu

Eric Linton
Department of Biology
Central Michigan University
Mt. Pleasant, MI 48859
USA
E-mail address: eric.linton@cmich.edu

Received September 30, 2019