New Mathematical Description of the Zika Virus Genome

Piotr Wąż

Department of Nuclear Medicine
Medical University of Gdańsk, Poland
Email: phwaz@gumed.edu.pl

Dorota Bielińska-Wąż

Department of Radiological Informatics and Statistics
Medical University of Gdańsk, Poland
Email: djwaz@gumed.edu.pl

***Abstract*—A new method of similarity/dissimilarity analysis of Deoxyribonucleic acid/Ribonucleic acid (DNA/RNA) sequences, is briefly outlined. The sequences are represented as a set of material points in a 3D-space. Such a 3D-dynamic graph is characterized numerically by the values analogous to the ones used in the classical dynamics. Application of such an approach for a characterization of the Zika virus genome is also discussed.**

Keywords–Bioinformatics; Alignment-free methods; Descriptors.

I. INTRODUCTION

A fast development of databases stimulated designing of new mathematical methods aiming at similarity/dissimilarity analysis of biosequences (DNA, RNA, protein). One branch of the methods, *Graphical Representations*, was created already in the eighties [1]. Originally, only a visual inspection on the considered objects was the aim of these studies. The DNA/RNA sequences are long and are composed of four letters. Therefore, in a natural way they can be represented as complicated objects in a 4-dimensional space. Since the human perception is limited, the reduction of the space to a lower dimension became desirable. This resulted in developing many different *Graphical Representation* methods [2][3][4][5][6] (for reviews see [7][8]). Alternatively, one can also characterize the sequences numerically, using so called *descriptors* [9]. Assigning descriptors to the graphs is far from being trivial.

In Section 2 we briefly outline a *Graphical Representation* method introduced by us several years ago and its application to a characterization of the Zika virus genome.

II. METHOD AND EXPECTED RESULTS

Recently, we have introduced a new Graphical Representation method called by us *3D-dynamic Representation of DNA/RNA Sequences* [10][11]. The inspiration for the numerical description of the 3D-dynamic-graphs came from the classical dynamics. We treat the graphs as rigid bodies. As descriptors characterizing the 3D-dynamic graphs we took coordinates of the centers of mass and the moments of inertia of these bodies. Two examples of the 3D-dynamic graphs are shown in Figure 1. The shapes and the locations of the graphs in the 3-dimensional space of different sequences are different. The aim of the new studies is an application of this approach to a new mathematical description of the Zika virus genome. Preliminary results are presented in Figures 2-5. Figure 2 and 3 show 3D-dynamic graphs representing the complete genome sequences of Zika virus. As we can see, the time evolution of the complete genome sequence of Zika virus is well represented graphically.

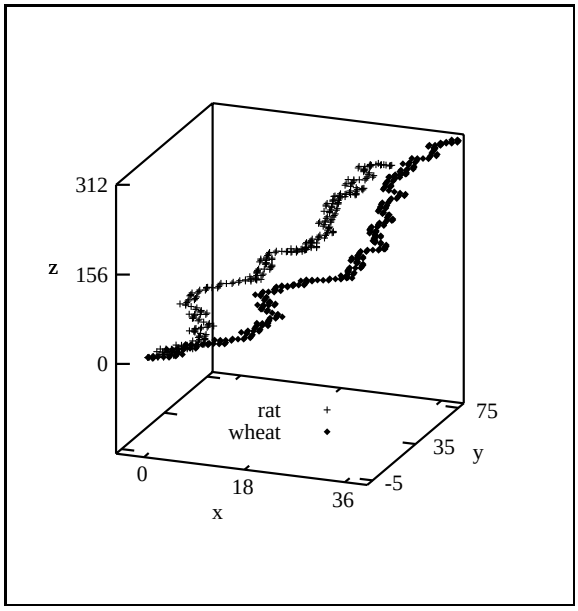

Figure 1. 3D-dynamic graphs.

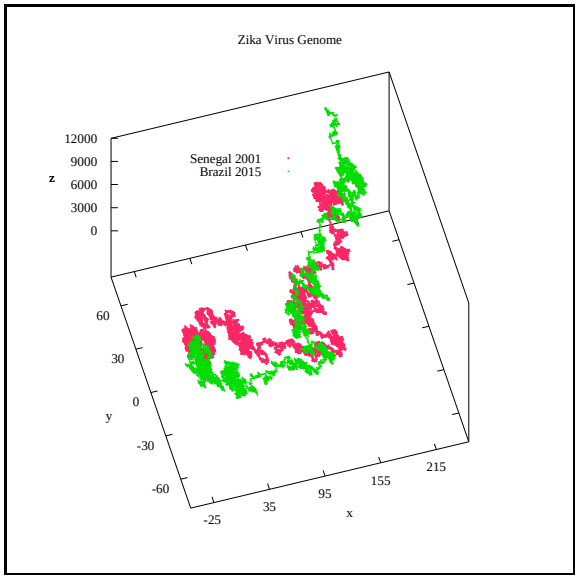

Figure 2. 3D-dynamic graphs representing the genomes of the Zika virus.

 ISBN: 978-1-61208-597-5

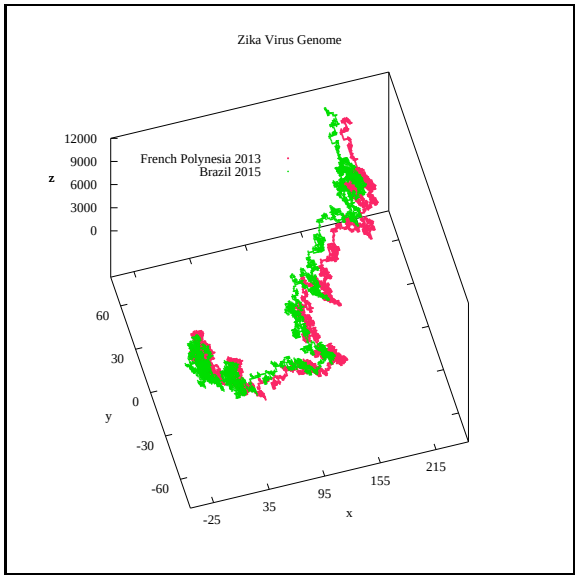

Figure 3. 3D-dynamic graphs representing the genomes of the Zika virus.

A pair of graphs is similar to each other: French Polynesia 2013 and Brazil 2015 (Fig. 2) and different: Senegal 2001 and Brazil 2015 (Fig. 3). This observation is confirmed by the calculations of the descriptors presented in the axes of Figures 4 and 5. μ_x, μ_y, μ_z are the coordinates of the centers of mass of the 3D-dynamic graph, in the $\{X, Y, Z\}$ coordinate system. They are are defined as

$$\mu_x = \frac{\sum_i m_i x_i}{\sum_i m_i}, \quad \mu_y = \frac{\sum_i m_i y_i}{\sum_i m_i}, \quad \mu_z = \frac{\sum_i m_i z_i}{\sum_i m_i}, \quad (1)$$

where x_i, y_i, z_i are the coordinates of the mass m_i. As the descriptors we also select the square roots of the normalized principal moments of inertia:

$$r_1 = \sqrt{\frac{I_1}{N}}, \quad r_2 = \sqrt{\frac{I_2}{N}}, \quad r_3 = \sqrt{\frac{I_3}{N}}, \quad (2)$$

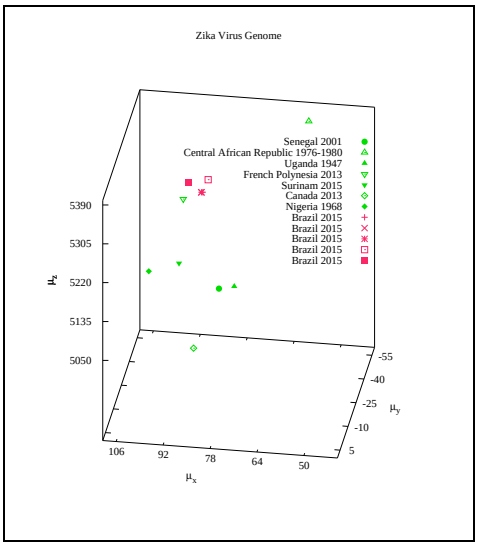

Figure 4. Classification diagram for the genomes of the Zika virus.

where I_1, I_2, I_3 are the principal moments of inertia of the 3D-dynamic graph and N is the length of the sequence.

In the Figures we observe concentrations of points in particular parts of the diagrams. We have already applied the 2D-dynamic representation of DNA/RNA sequences for a characterization of this virus [12]. In the future work, we are going to check if the third dimension supplies any new and relevant information.

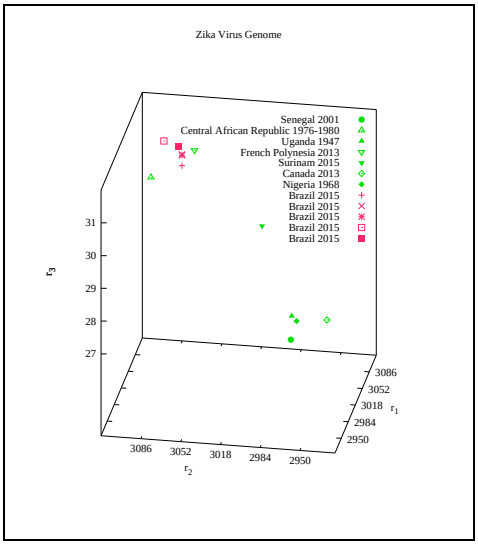

Figure 5. Classification diagram for the genomes of the Zika virus.

REFERENCES

[1] E. Hamori and J. Ruskin, "H Curves, a novel method of representation of nucleotide series especially suited for long DNA sequences", J. Biol. Chem. vol. 258, pp. 1318–1327, 1983.

[2] P. M. Leong and S. Morgenthaler, "Random walk and gap plots of DNA sequences", Comput. Appl. Biosci. vol. 11, pp. 503–507, 1995.

[3] Q. Dai, X. Liu, and T. Wang, "A novel graphical representation of DNA sequences and its application", J. Mol. Graph. Model. vol. 25, pp. 340–344, 2006.

[4] H. González-Díaz et al., "Generalized lattice graphs for 2D-visualization of biological information", J. Theor. Biol. vol. 261, pp. 136–147, 2009.

[5] Z. Zhang, T. Song, X. Zeng, Y. Niu, Y. Jiang, L. Pan, and Y. Ye, "ColorSquare: A colorful square visualization of DNA sequences", MATCH Commun. Math. Comput. Chem. vol. 68, pp. 621–637, 2012.

[6] N. Jafarzadeh and A. Iranmanesh, "C-curve: a novel 3D graphical representation of DNA sequence based on codons", Math Biosci. vol. 241, pp. 217–224, 2013.

[7] D. Bielińska-Wąż, "Graphical and numerical representations of DNA sequences: Statistical aspects of similarity", J. Math. Chem. vol. 49, pp. 2345–2407, 2011.

[8] M. Randić, M. Novič and D. Plavšić, "Milestones in Graphical Bioinformatics", Int. J. Quant. Chem. vol. 113, pp. 2413–2446, 2013.

[9] A. Nandy, M. Harle, and S.C. Basak, "Mathematical descriptors of DNA sequences: development and applications", Arkivoc ix, pp. 211–238, 2006.

[10] P. Wąż and D. Bielińska-Wąż, "3D-dynamic representation of DNA sequences", J. Mol. Model. vol. 20, 2141, 2014.

[11] P. Wąż and D. Bielińska-Wąż, "Non-standard similarity/dissimilarity analysis of DNA sequences", Genomics vol. 104, pp. 464–471, 2014.

[12] D. Panas, P. Wąż, D. Bielińska-Wąż, A. Nandy, and S.C Basak, "2D-Dynamic Representation of DNA/RNA Sequences as a Characterization Tool of the Zika Virus Genome", MATCH Commun. Math. Comput. Chem. vol. 77, pp. 321–332, 2017.

 ISBN: 978-1-61208-597-5