2D-Dynamic Representation of DNA/RNA Sequences as a Characterization Tool of Influenza Viruses

Dorota Bielińska-Wąż

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

Piotr Wąż

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

***Abstract*—The aim of this work is an application of the *2D-dynamic representation of DNA/RNA sequences* for a description of influenza viruses. The descriptors (numerical characteristics) for the considered sequences are calculated. We expect that the results will reveal some new features of the considered objects and will be helpful for the creation of a model of time evolution of influenza viruses.**

Keywords–Bioinformatics; Alignment-free methods; Descriptors.

I. INTRODUCTION

Recently, we have introduced and developed a new method of comparison of Deoxyribonucleic acid/Ribonucleic acid (DNA/RNA) sequences called by us *2D-dynamic representation of DNA/RNA sequences* [1]–[5]. In the 2D-dynamic representation, the DNA/RNA sequence is represented as a set of material points in 2D space ("2D-dynamic graph").The distribution of the points in the plane and the way of calculating their masses is described in [1]. This method belongs to a group of methods in bioinformatics known in the literature as *Graphical Representation Methods*. They allow for both graphical and numerical comparison of the objects. The first methods of this kind have been published in the eighties and nineties [6]–[9]. Since then, many other approaches have been constructed, as for example [10]–[13]. Reviews may be found in [14] [15]. Each method describes different aspects of similarity and still new approaches are constructed.

II. METHOD AND EXPECTED RESULTS

2D-dynamic representation is based on shifts in a two dimensional space. The DNA/RNA sequence is represented by material points with different masses in a two dimensional space. This method is an improvement of traditional plots, in which particular bases are represented by two orthogonal pairs of colinear basis vectors. Such a choice of the vectors leads to the possibility of shifts back and forth along the same trace. The so called repetitive walks lead to degeneracy: different sequences may be represented by the same graphs. In order to remove the degeneracy, points with masses which are a multiplicity of the unit mass have been introduced. After a unit shift a point with unit mass is localized. If the ends of the vectors during the shifts coincide, then the mass of this point increases accordingly. The total mass of the graph (the sum of all masses) is equal to the length of the sequence.

Several examples of the 2D-dynamic graphs representing different complete genome sequences of Zika virus are shown

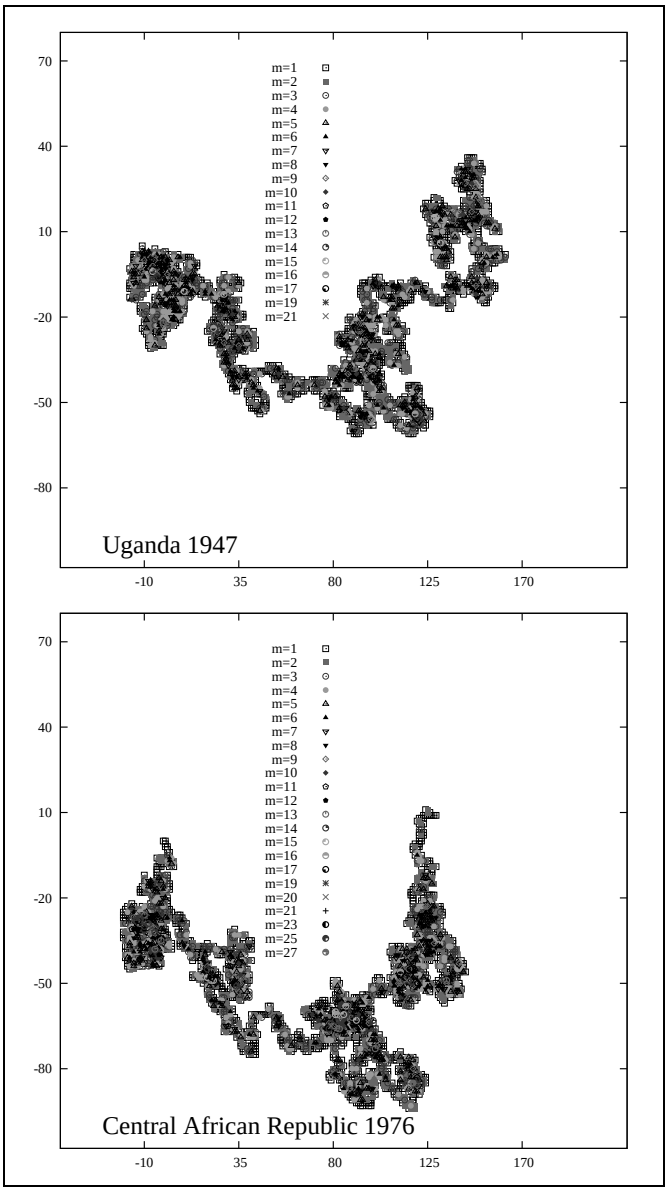

Figure 1. 2D-dynamic graphs.

 ISBN: 978-1-61208-571-5

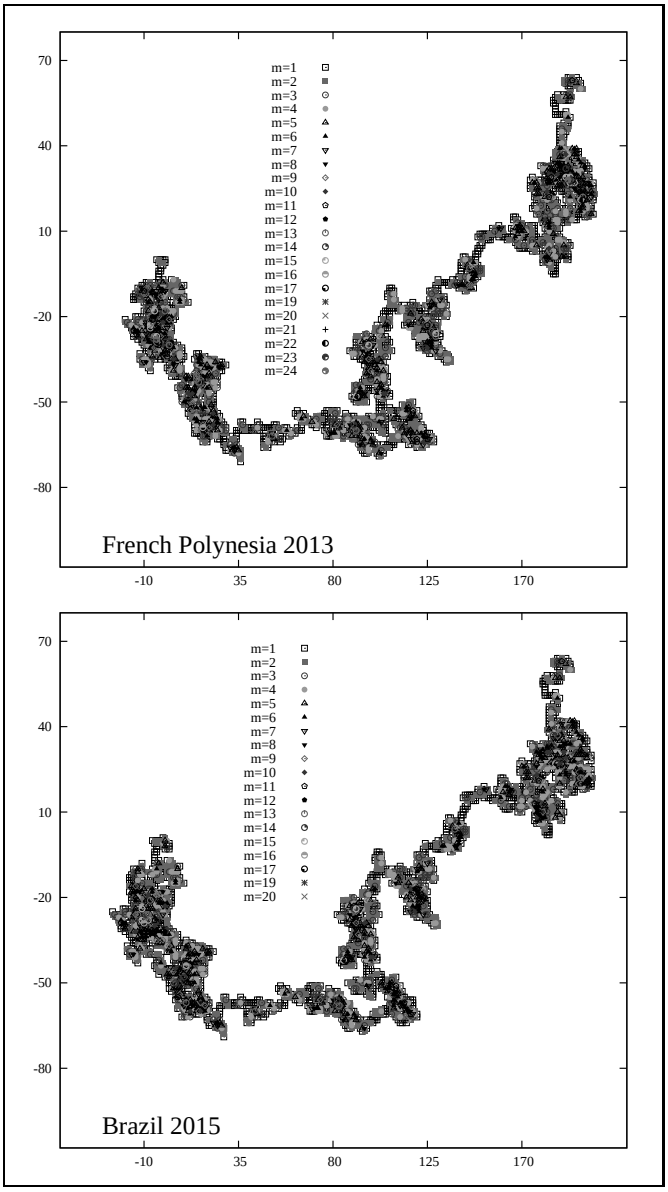

Figure 2. 2D-dynamic graphs.

in Figures 1-2. Similar sequences are represented by similar graphs. The structural forms of the graphs evolve with time. Pairs of graphs are similar to each other: HQ234498 Uganda 1947 is similar to KF268948 Central African Republic 1976 (Fig. 1) and KJ776791 French Polynesia 2013 is similar to KU365777 Brazil 2015 (Fig. 2). This observation is confirmed by the calculations done in [5].

Each graphical object is here described by a set of numerical values called in the theory of molecular similarity *descriptors*. We have shown that dynamical variables of the classical mechanics as, for example, the coordinates of the center of mass are the representative descriptors:

$$\mu_x = \frac{\sum_i m_i x_i}{\sum_i m_i}, \quad \mu_y = \frac{\sum_i m_i y_i}{\sum_i m_i}, \tag{1}$$

where x_i, y_i are the coordinates of mass m_i assigned to the i-th node of the 2D-dynamic graph. We have also used some other descriptors, as for example the principal moments of inertia of the graphs.

In the present work we apply these descriptors to a description of the sequences of influenza viruses. We hope to discover some correlations between the descriptors and time and place. The necessary data are freely accessible in database Genbank. The mathematical description of these data will be helpful for a general knowledge about the viruses and also for a prediction of their evolution. This information is relevant for the designing of the vaccine.

REFERENCES

[1] D. Bielińska-Wąż, T. Clark, P. Wąż, W. Nowak, and A. Nandy, "2D-dynamic representation of DNA sequences", Chem. Phys. Lett. vol. 442, pp. 140–144, 2007.

[2] D. Bielińska-Wąż, W. Nowak, P. Wąż, A. Nandy, and T. Clark, "Distribution moments of 2D-graphs as descriptors of DNA sequences", Chem. Phys. Lett. vol. 443, pp. 408–413, 2007.

[3] D. Bielińska-Wąż, P. Wąż, and T. Clark, "Similarity studies of DNA sequences using genetic methods", Chem. Phys. Lett. vol. 445, pp. 68–73, 2007.

[4] P. Wąż, D. Bielińska-Wąż, and A. Nandy, "Descriptors of 2D-dynamic graphs as a classification tool of DNA sequences", J. Math. Chem. vol. 52, pp. 132–140, 2013.

[5] 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.

[6] 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.

[7] M. A. Gates, "Simpler DNA sequence representations", Nature vol. 316, p. 219, 1985.

[8] A. Nandy, "A new graphical representation and analysis of DNA sequence structure. I: Methodology and application to globin genes", Current Science vol. 66, pp. 309–314, 1994.

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

[10] 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.

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

[12] 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.

[13] 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.

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

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

 ISBN: 978-1-61208-571-5