INPE – National Institute for Space Research
São José dos Campos – SP – Brazil – May 16-20, 2016

A GEOGRAPHICALLY-AWARE COMPLEX NETWORK APPROACH FOR FOOT-AND-MOUTH DISEASE PHYLODYNAMICS

Luiz Max F. de Carvalho[1], Leonardo B. L. Santos[2], Paulo E. P. Burke[3], Marcos G. Quiles[3], Waldemir de Castro Silveira[4]

[1]Institute of Evolutionary Biology – University of Edinburgh, Edinburgh, United Kingdom, lmax.procc@gmail.com
[2]Brazilian National Center for Monitoring and Early Warning of Natural Disasters (CEMADEN), São José dos Campos/SP, Brazil, santoslbl@gmail.com
[3]Federal University of São Paulo, São Paulo, Brazil, pauloepburke@gmail.com, quiles@unifesp.br
[4]Trimatrix LTDA, Rio de Janeiro, Brazil, silveira.w@trimatrix.com.br

Abstract: Phylodynamics provides the theoretical foundation for the use of the information contained in genomes to draw inference on temporal and spatial behavior of organisms. In this study we use complex networks to explore the structure of genetic data within a computationally efficient framework. In this work we analyze 167 1D (VP1) gene sequences from Foot-and-Mouth Disease on South-America using a family of Complex Networks embedded on geographic space. Using genetic dissimilarity as connection threshold, our topological analysis show networks with high clustering.

keywords: Complex Networks, Epidemiology and Mathematical Models, Bioinformatics, Phylogeography, clustering.

1. INTRODUCTION

The availability of time-stamped georeferenced genomic sequences made possible the emergence of the new field of phylodynamics [1] in Evolutionary Biology. Phylodynamics provides the theoretical foundation for the use of the information contained in genomes to draw inference on temporal and spatial behavior of organisms. The growing availability of data, however, calls for the development of novel statistical approaches to inference, and new methods for data exploration, visualization and knowledge discovery [2]. Complex Networks are flexible and powerful mathematical modeling tools, allowing for the formal representation of complex physical systems. In Biology, detecting community structure is especially useful in classification problems, such as protein annotation and phylogenetics [3-5].

2. MATERIAL AND METHOD

In this study we use complex networks to explore the structure of genetic data within a computationally fast framework. In this work we analyze 167 1D (VP1) gene sequences from Foot-and-Mouth Disease in South-America (Figure 1) using a family of Complex Networks embedded in geographic space.

Figure 1: Geographical location of each sequence in the dataset.

3. RESULTS AND PERSPECTIVES

       Figure 2 shows a scatterplot of genetic divergence against geographical distance, from which no clear association can be discerned. In contrast to this behavior, we obtained a high clustering complex network obtained using genetic dissimilarity as connection threshold, as shown in Figure 3.
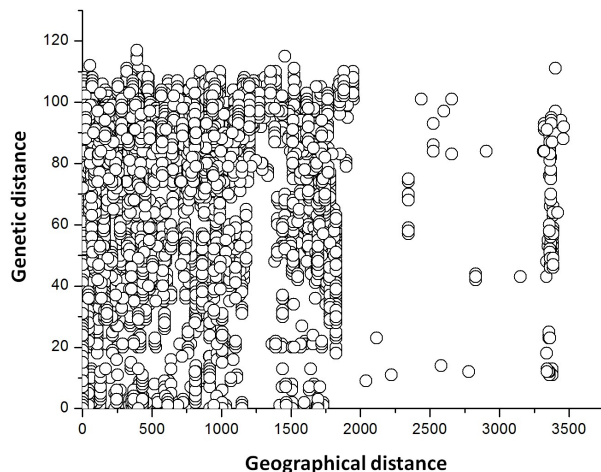

Figure 2: Plot of genetic divergence against geographical distance for each pair of analyzed sequences.

References

[1] B. T. Grenfell, O. G. Pybus, J. R. Gog, J. L. Wood, J. M. Daly, J. A. Mumford, E. C. Holmes. (2004) Unifying the epidemiological and evolutionary dynamics of pathogens. Science, v. 303, p. 327-332.

[2] A. C. Gavin, P. Aloy, P. Grandi, R. Krause, M. Boesche, M. Marzioch, C. Rau, L. J. Jensen, S. Bastuck, B. Dumpelfeld, A. Edelmann, M. A. Heurtier, V. Ho_man, C. Hoefert, K. Klein, M. Hudak, A. M. Michon, M. Schelder, M. Schirle, M. Remor, T. Rudi, S. Hooper, A. Bauer, T. Bouwmeester, G. Casari, G. Drewes, G. Neubauer, J. M. Rick, B. Kuster, P. Bork, R. B. Russell, G. Superti-Furga. (2006) Proteome survey reveals modularity of the yeast cell machinery. Nature, v. 440, p. 631-63.

[3] A. Goes-Neto, M. V. Diniz, L. B. Santos, S. T. Pinho, J. G. Miranda, T. P. Lobao, E. P. Borges, C. N. El-Hani, R. F. Andrade. (2010) Comparative protein analysis of the chitin metabolic pathway in extant organisms: a complex network approach. BioSystems, v. 101, p. 59-66.

[4] R. F. Andrade, I. C. Rocha-Neto, L. B. Santos, C. N. de Santana, M. V. Diniz, T. P. Lobao, A. Goes-Neto, S. T. Pinho, C. N. El-Hani. (2011) Detecting network communities: an application to phylogenetic analysis. PLoS Comput. Biol., vol. 7, p. e1001131.

[5] L. M. Carvalho, L. B. L. Santos, N. R. Faria & W. C. Silveira (2013). Phylogeography of foot-and-mouth disease virus serotype O in Ecuador. *Infection, Genetics and Evolution*, *13*, 76-88.
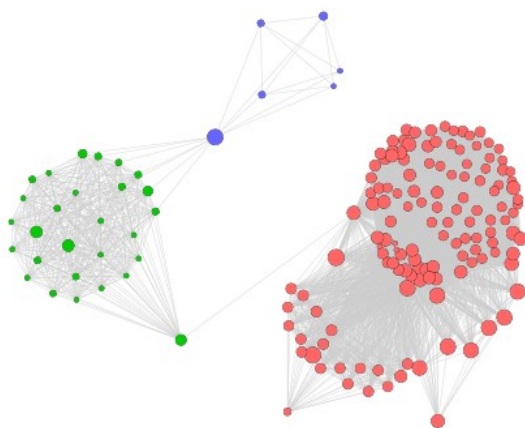
Figure 2: Community structure of the complex network obtained using genetic dissimilarity as connection threshold. Please note node positions are not in geographic space.

       The next step of the research is to investigate the relationship between these communities and the geographical location of the nodes.