Whole-Genome Based Molecular Phylogeny of Crimean-Congo Hemorrhagic Fever Virus for More Accurate Identification of the Virus Origin

Sepideh Eisazaei [1], Mohammad Tohidlou [2], Milad Lagzian [3]*

Dept. of Biology, Faculty of Science, University of Sistan and Baluchestan, Zahedan, Iran [1, 3]

Dept. of Biophysic, Faculty of Biological Science, Tarbiat Modares University, Tehran, Iran [2]

* m.lagzian@science.usb.ac.ir

Abstract: Crimean-Congo hemorrhagic fever virus (CCHFV) is one of the deadliest zoonotic viruses with widespread distribution around the world. In Iran, this virus is more prevalent in Sistan and Baluchestan provenance due to a large number of livestock imports and presence of the intermediate host [1]. Therefore, epidemiological studies seem crucial for identification of the virus origin to take countermeasures. However, due to high rate of mutation in the virus genome and also segment reassortment between the viruses, finding of their exact origins proved to be very problematic by a phylogenetic tree constructed on a single gene [2, 3]. In this study, we explore the power of the evolutionary tree constructed on the whole genome of the virus instead of its S-segment for elucidation of relationship among the virus population. Initially, a deep search was accomplished to find all completely sequenced S, M and L segments of the virus based on the isolation country. Subsequently, the complete genome of the virus was constructed by joining appropriate segments together for each country. In the next step, phylogenetic tree construction was conducted by two methods on the whole genomes: K-mer based distance estimation, which is very fast algorithm and does not require sequence multiple alignments or high computational power. The second was RAxML maximum likelihood tree with rapid bootstrapping algorithms (1000 bootstrap) which is relayed on sequence multiple alignments and high calculation power. In order to compare the effectiveness of the method, another RAxML tree was constructed on the complete S-segment of the same virus genomes. The results are interesting and showed the robustness of the whole genome tree for better determination of the virus origin. For example, according to the S-segment tree, the virus first entered to Pakistan and then reached to Afghanistan. Subsequently, it was spread out to Iran and India at a same time which is very unlikely because Afghanistan is only an importer of livestocks. In the other side, whole genome based tree completely gives a different pattern which is more realistic. According to this tree, the virus initially entered to India (a major livestock breeder) then spread to Pakistan and subsequently reach to Iran and Afghanistan, which are major customers of Pakistan's livestocks. In addition, clades in the whole genome tree have significantly higher bootstrap values in compare to S-segment tree. Surprisingly, the tree constructed by K-mer based distance estimation is completely similar to the RAxML whole genome tree. This tree can be drawn in less than 10 minutes in compare to the RAxML tree which takes more than 10 hours to finish in our study.

Keywords : CCHFV; Phylogenetic Tree; RaxML; K-mer; Whole genome

References:

[1] Lukashev, A.N., et al., Phylogeography of Crimean Congo Hemorrhagic Fever Virus. PLoS One, 2016. 11(11): p. e0166744.

[2] Bente, D.A., et al., Crimean-Congo hemorrhagic fever: history, epidemiology, pathogenesis, clinical syndrome and genetic diversity. Antiviral Res, 2013. 100(1): p. 159-89.

[3] Zehender, G., et al., Bayesian phylogeography of Crimean-Congo hemorrhagic fever virus in Europe. PLoS One, 2013. 8(11): p. e79663.