

Spread of Middle East Respiratory Coronavirus: Genetic versus Epidemiological Data

Daniel A. Janies*, Colby Ford, Lambodhar Damodaran and Zachary Witter

Bioinformatics and Genomics, University of North Carolina at Charlotte, Charlotte, NC, USA

Objective

Here we use novel methods of phylogenetic transmission graph analysis to reconstruct the geographic spread of MERS-CoV. We compare these results to those derived from text mining and visualization of the World Health Organization's (WHO) Disease Outbreak News.

Introduction

MERS-CoV was discovered in 2012 in the Middle East and human cases around the world have been carefully reported by the WHO. MERS-CoV virus is a novel betacoronavirus closely related to a virus (NeoCov) hosted by a bat, *Neoromicia capensis*. MERS-CoV infects humans and camels. In 2015, MERS-CoV spread from the Middle East to South Korea which sustained an outbreak. Thus, it is clear that the virus can spread among humans in areas in which camels are not husbanded.

Methods

Phylogenetic analyses

We calculated a phylogenetic tree from 100 genomic sequences of MERS-CoV hosted by humans and camels using NeoCov as the outgroup. In order to evaluate the relative order and significance of geographic places in spread of the virus, we generated a transmission graph (Figure 1) based on methods described in 1.

The graph indicates places as nodes and transmission events as edges. Transmission direction and frequency are depicted with directed and weighted edges. Betweenness centrality, represented by node size, measures the number of shortest paths from all nodes to others that pass through the corresponding node. Places with high betweenness represent key hubs for the spread of the disease. In contrast, smaller nodes at the periphery of the network are less important for the spread of the disease.

Web scraping and mapping

Due to the journalistic style of the WHO data, it had to be structured such that mapping software can ingest the data. We used Import.io to build the API. We provided the software a sample page, selected the data that is pertinent, then provided a list of all URLs for the software. We used Tableau to map the information both geographically and temporally.

Results

Geographic spread of Mers-CoV based on transmissions identified in phylogenetic data

Most important among the places in the MERS-CoV epidemic is Saudi Arabia as measured by the betweenness metric applied to a changes in place mapped to a phylogenetic tree. In figure 1, the circle representing Saudi Arabia is slightly larger compared to other location indicating its high importance in the epidemic. Saudi Arabia is the source of virus for Jordan, England, Qatar, South Korea, UAE, Indiana, and Egypt. The United Arab Emirates has a bidirectional connection with Saudi Arabia indicating the virus has spread between the two countries. The United Arab Emirates also has high betweenness. The United Arab Emirates is between Saudi Arabia and Oman and Between Saudi Arabia and France. South Korea, and Qatar have mild betweenness. South Korea is between Saudi Arabia and

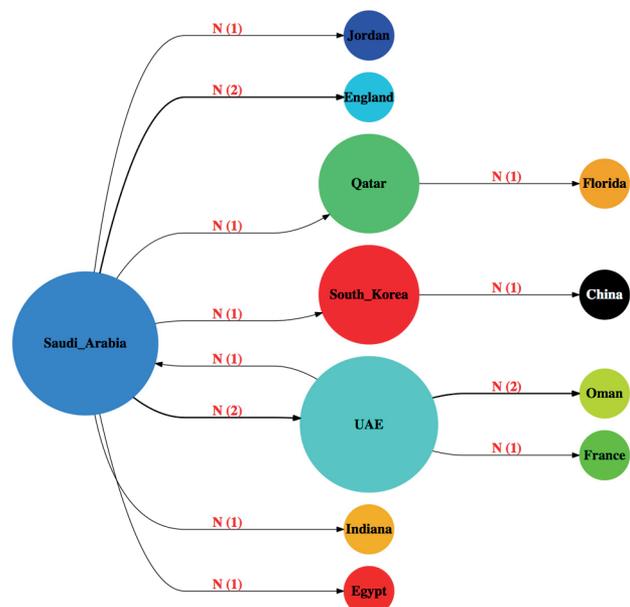
China. Qatar is between Saudi Arabia and Florida. Other locations (Jordan, England, Indiana, and Egypt) have low betweenness as they have no outbound connections.

Visualization of geographical transmissions in WHO Data

Certain articles include the infected individuals' countries of origin. In contrast, many reports are in a lean format that includes a single paragraph that only summarizes the total number of cases for that country. If we build the API in a manner that recognizes features in the detailed reports, we can generate a map that draws lines from origin to reporting country and create visualizations. However, since only some of the articles contain this extra information, mapping in this manner will miss many of the cases that are reported in the lean format.

Conclusions

Our goal is to develop methods for understanding syndromic and pathogen genetic data on the spread of diseases. Drawing parallels between the transmissions events in the WHO data and the genetic data has shown to be challenging. Analyses of the genetic information can be used to imply a transmission pathway but it is hard to find epidemiological data in the public domain to corroborate the transmission pathway. There are rare cases in the WHO data that include travel history (e.g. "The patient is from Riyadh and flew to the UK"). We conclude that epidemiological data combined with genetic data and metadata have strong potential to understand the geographic progression of an infectious disease. However, reporting standards need to be improved where travel history does not impinge on privacy.



A transmission graph for MERS-CoV based on viral genomes and place of isolation metadata. The direction of transmission is represented by the arrow. The frequency of transmission is indicated by the number. The size of the nodes indicates betweenness.



ISDS 2016 Conference Abstracts

Keywords

Phylogenetics; Transmission Graph; MERS-CoV; Syndromic; Genetic

Acknowledgments

Research reported in this publication was supported by a UNC Research Opportunities Initiative grant to UNC Charlotte, NC State University, and UNC-Chapel Hill. The data used in this report are publically available in Genbank and the WHO. We gratefully acknowledge all the researchers and institutions who submitted prepublication data to the public domain.

References

Janies, D. A., Pomeroy, L. W., Krueger, C., Zhang, Y., Senturk, I. F., Kaya, K. and Çatalyürek, Ü. V., Phylogenetic visualization of the spread of H7 influenza A viruses. *Cladistics*. 2015. 31: 679–691. doi:10.1111/cla.12107

*Daniel A. Janies

E-mail: djanies@uncc.edu

