Comparative Study of Middle East respiratory syndrome coronavirus using Bioinformatics Techniques

-Based on the Codon Variation Patterns -

Insung Ahn and Jin-Hwa Jang
Biomedical Prediction Technology Lab.,
Korea Institute of Science and Technology Information,
Daejeon, Korea
e-mail: isahn@kisti.re.kr

Insung Ahn
Department of Big Data Science,
University of Science & Technology,
Daejeon, Korea
e-mail: isahn@ust.ac.kr

Abstract— **In this study, we collected 141 coding sequences of MERS-CoV from the National Center for Biotechnology and Information (NCBI), including the sequences isolated in Korea in 2015. We conducted the phylogenetic analysis using the Maximum Likelihood method to examine the overall variation patterns among the target sequences, and then, we divided the target sequences into 4 different groups according to occurred countries and host species. Using the codon analyzer named SimFluVar program, we analyzed the codon variation patterns in the wobble position among 4 groups. In order to investigate the effect of codon variations that can change the phenotype of target genes, we compared the transversional substitutions between the Korean-origin sequences and other groups. As a result, we found that the Korea-origin sequences showed very minor differences with those collected from the Saudi Arabia in 2015, whereas other groups which were collected from USA and UK in 2013 and 2014 showed more complicated differences. We also compared the Korea-origin sequences with those of camel-origin sequences, and we found that the substitution pattern was somewhat different with that of human-origin viruses.**

Keywords- Middle East respiratory syndrome coronavirust; codon variation; wobble position; bioinformatics

I. INTRODUCTION

Since the first case of the Middle East respiratory coronavirus (MERS-CoV) was reported in 2012, this novel virus has been spread among the Middle-Eastern countries such as Egypt, Oman, Qatar, and Saudi Arabia. The route of transmission is not clearly understood, but, according to the phylogenetic studies, MERS-CoV showed the high similarities with the camel-origin MERS-CoV. In May 20, 2015, MERS-CoV was transmitted to Korea through a person who visited the Arabian Peninsula. Because of the high pathogenicity and infectivity, the MERS-CoV which was collected in Korea has been doubted as a novel variant of MERS-CoVs. In this study, we determined the outline of the variation patterns among 141 MERS-CoVs sequences using phylogenetic analysis method, and then, performed more intensive analysis using the codon analyzer named SimFluVar (http://lcbbsnu.ac.kr/simfluvar) to identify the minor, but might be important, differences among them.

.

II. MATERIALS & METHODS

A. Sequences

In the present study, 141 nucleotide sequences (coding sequences) of MERS-CoVs were collected from the National Center for Biotechnology and Information in the United States. We used the 'ORF1a-ORF1an-S-ORF3-ORF4a-ORF5-E-M-N-ORF8b' region of MERS-CoVs whose nucleotide length was 4062 base pairs. The data used spanned over a period of 4 years, from 2012 to 2015, including the sequences occurred in Korea.

B. Phylogenetic analysis

For phylogenetic analysis, multiple sequence alignment was performed on the collated sequences using the ClustalW program (gap open penalty = 40, gap extension penalty = 0.8) [1]. Using the aligned sequences that included the gaps generated from multiple sequence alignment, phylogenetic correlation was performed using MEGA (version 6.06) program [2] to create the maximum-likelihood tree (circular cladogram) based on Tamura-Nei substitution model [3].

C. Comparison of codon variation patterns

To ascertain the codon substitution pattern of MERS-CoVs, we extracted 25 sequences which can represent the minor differences among 141 MERS-CoVs. We divided the target sequences into 4 different groups according to their epidemiological characteristics. Analysis was performed using the SimFluVar (http://lcbb.snu.ac.kr/simfluvar) program, which enabled sequence analysis between different genome groups. The SimFluVar is an analytical tool for calculating the codon substitution patterns of influenza virus. Designed to compare a large number of nucleotide sequences, SimFluVar provides precise patterns of codon variations between two viral groups, especially for the influenza virus. SimFluVar also provides the useful functions, such as editing and visualization of the result matrix [4].

I. RESULTS

In the early stage of our analysis, we performed the phylogenetic analysis to identify the approximate differences among MERS-CoVs. This is the traditional method commonly used worldwide when the disease outbreak occurs.

Here, we suggest another method which can detect the minor, but might be important, patterns among the very similar viral sequences using the SimFluVar program. Using this software, we found that the Korea-origin sequences showed very minor differences with those collected from the Saudi Arabia in 2015, whereas other groups which were collected from USA and UK in 2013 and 2014 showed more complicated differences. We also compared the Korea-origin sequences with those of camel-origin sequences, and we found that the substitution pattern was somewhat different with that of human-origin viruses.

In conclusion, codon variation analysis, especially the transversional substitution analysis, can be used as a good method to find a minor, but possibly important, difference among similar viral groups.

ACKNOWLEDGMENT

The research was supported by Biomedical Prediction Technology Laboratory at the Korea Institute of Science and Technology Information (K-15-L03-C02-S03).

REFERENCES

[1] M. A. Larkin, G. Blackshields, N. P. Brown, R. P. A. Chenna, McGettigan, H. McWilliam, F. Valentin, I. M. Wallace, A. Wilm, R. Lopez, J. D. Thompson, T. J. Gibson, D. G. Higgins, "Clustal W and Clustal X version 2.0", Bioinformatics, vol. 23, 2007, p2947

[2] K. Tamura, G. Stecher, D. Peterson, A. Filipski, S. Kumar, "MEGA6: Molecular Evolutionary Genetics Analysis version 6.0", Mol. Biol. Evol., vol. 30, 2013, p2725

[3] K. Tamura, M. Nei, "Estimation of the number of nucleotide substitutions in the control region of mitochondrial DNA in humans and chimpanzees", Mol. Biol. Evol., vol. 10, 1993, p512

[4] Ahn, J. H. Jang, H. Y. Kim, J. H. Lee, H. S. Son, "A Visualization Tool for Calculating the Genetic Substitution Patterns Between Two Different Groups", Evol. Bioinformatics, vol. 11, 2015, p179

[5]