

## **Analysis of Amino Acid Sequence of SARS Coronavirus, Feline Coronavirus, Rat Coronavirus, Turkey Coronavirus, MERS Coronavirus with Decision Tree, Apriori Algorithm**

Seung Hye Song<sup>1</sup>, Sung jin Kim<sup>1</sup> and Taeseon Yoon<sup>1+</sup>

<sup>1</sup> Hankuk Academy of Foreign Studies, South Korea

**Abstract.** Coronaviruses primarily infect the upper respiratory and gastrointestinal tract of mammals and birds, according to its host, it can be categorized, Porcine coronavirus(TGEV), Infectious bronchitis virus(IBV), Bovine coronavirus(BCV), Feline coronavirus(FCoV) and many other types. Among these diverse kinds, we selected 5 viruses which is: SARS, MERS, Feline coronavirus, Rat coronavirus, Turkey coronavirus. It is a proven fact that H4UK and MERS are almost equal genetically but one factor which is cell-surface protein is what makes these two viruses different. By this, we were able to come up with an idea that there might be a possibility, due to differences in amino acid sequence of cell-surface protein, it may allow the coronavirus to be infectious to humans even if it's not the proper host. So, we selected these 5 coronaviruses mentioned in order to compare their amino acid sequence and predict the possibility by using a proper algorithm: Decision tree, Neural Network, Apriori algorithm.

**Keywords:** Coronavirus, SARS, MERS, Feline coronavirus, Rat coronavirus, Turkey coronavirus, H4UK, Decision Tree, Apriori Algorithm.

### **1. Introduction**

MERS-CoV is a betacoronavirus with single-strand RNA that is transmitted from animals to humans, and it is distinct from SARS coronavirus and common-cold coronavirus. Its genomes are phylogenetically classified into two clades, Clades A and B. Early cases of MERS were of Clade A clusters while new cases are of the other in general. The origins of the virus are not fully understood, but according to the analysis of different virus genomes, it is believed that it originated in bats and camels are likely to be a major reservoir host in its spread to human. It was first identified in Saudi Arabia in 2012. Symptoms include fever, cough, shortness of breath and also gastrointestinal symptoms. Currently, no vaccine or specific treatment is available. Both *Tylosycteris* bat coronavirus HKU4 (Ty-BatCoV HKU4) and *Pipistrellus* bat coronavirus HKU5 (Pi-BatCoV HKU5) are enveloped, positive sense, single-stranded RNA mammalian Group 2 betacoronaviruses that has been found to be genetically related to the MERS-CoV. According to recent study, the spikes in BatCoV HKU4, but not HUK5, bind to the MERS-CoV human receptor site DPP4, which is also known as hCD26, and the receptor is engaged to initiate viral entry. Thus, virus particles pseudotyped with BatCoV HKU4 spike infect cells by engaging hCD26. However, Despite sharing a common receptor, MERS-CoV and HKU4 spikes demonstrate functional differences: whereas MERS-CoV prefers human DPP4 over bat DPP4 as its receptor, HKU4 tends to be the opposite; and in the absence of exogenous proteases, both MERS-CoV and HKU4 spikes mediate pseudovirus entry into bat cells. However, in the case of human cells, only MERS-CoV spike, but not HKU4 spike, mediates pseudovirus entry into the cells. Thus, MERS-CoV, but not HKU4, has adapted to use human DPP4 and human cellular proteases, contributing to the enhanced infection of MERS-CoV in humans.

---

<sup>+</sup> Corresponding author. Tel.: + 821056967108; fax: +82313240700.  
E-mail address: tsyoon@hafs.hs.kr.

## **2. Material and Method**

### **2.1. Feline coronavirus**

FCoV is a positive-stranded RNA virus that is infectious to cats.[1] This virus is part of the Alphacoronavirus 1 species of the Alphacoronavirus genus belonging to the Coronaviridae family of viruses.[2] FCoV is divided into FECV (feline enteric coronavirus) that infects the intestines and the FIPV (feline infectious peritonitis virus) that causes the disease feline infectious peritonitis (FIP) according to its form. The virus becomes feline infectious peritonitis virus (FIPV) by mutation of FECV to FIPV occurred when random errors occur in the virus infecting an enterocyte.[3] Disease caused by FIPV is lethal and incurable which is feline infectious peritonitis (FIP) and there is no known cure; treatment is generally symptomatic and palliative only.

### **2.2. Rat coronavirus**

Sialodacryoadenitis (SDAV), Rat coronavirus (RCV) are very similar and are now considered to be strains of the same virus, namely Rat Coronavirus or RCV. Symptoms for these disease can vary from the subclinical (no detectable illness) to death and they also vary from the age of that rat. In young rats, symptoms such as squinting, photophobia, lacrymation (runny eyes), production of porphyrin around the eyes and nose are seen. While, older rats usually develop sneezing and sometimes swelling of the salivary glands at the angle of the jaw (parotid) and up the sides (submaxillary). There is no specific treatment such as vaccine for this infection but supportive therapy such as blanket administration of pain killers can help.

### **2.3. Turkey coronavirus**

Turkey Coronaviruses include turkey enteritis virus and bluecomb disease virus. Transmission is horizontal via the faeces, and can be direct or indirectly spread by fomites which appears as a form of diarrhoea, anorexia and lethargy. Currently no vaccine for such disease are available but ameliorative therapies which include supplying heat, calf milk replacer, potassium chloride added to milk suspension, antibiotics and copper sulphate can all be used to help the affected birds. In addition if some birds recover from such disease, they are immune for life.

### **2.4. SARS coronavirus**

SARS-CoV, is the virus that causes severe acute respiratory syndrome (SARS).[4] SARS patients share common symptoms which are influenza-like and include fever, malaise, myalgia, headache, diarrhoea, and shivering (rigors)[5] and they commonly show decrease in number of lymphocytes circulating the blood. [6] SARS like many other coronaviruses starts its genome expression with translation of two large ORFs 1a and 1b, which are two polyproteins. Coronavirus including SARS-CoV These viruses have large pleomorphic spherical particles with bulbous surface projections that form a corona around particles. The envelope of the virus contains lipid and appears to consist of a distinct pair of electron dense shells. More research needed to determine the relationship between function and structure of critical enzymes and structural proteins is making vaccine development of SARS which means that until now, no proper vaccine has been developed

### **2.5. MERS coronavirus**

MERS-CoV is a betacoronavirus with single-strand RNA that is transmitted from animals to humans, and it is distinct from SARS coronavirus and common-cold coronavirus[7]. Its genomes are phylogenetically classified into two clades, Clades A and B. Early cases of MERS were of Clade A clusters while new cases are of the other in general. The origins of the virus are not fully understood, but according to the analysis of different virus genomes, it is believed that it originated in bats and camels are likely to be a major reservoir host in its spread to human. It was first identified in Saudi Arabia in 2012. Symptoms include fever, cough, shortness of breath and also gastrointestinal symptoms. Currently, no vaccine or specific treatment is available.

### **2.6. Decision Tree**

This method is a decision support tool that generates a tree-like graph of decisions, in order to apply it into a certain algorithm[8]. In this study uses this method to extract the specific patterns of amino acid

sequences in each virus. The procedure done is called the ‘Data mining.’ Data mining refers to extracting patterns from raw or warehoused data, in other words, efficiently extracting previously unknown but potentially useful patterns from data. The extracted patterns can be used to classify the data, understand the characteristics of each data sets, and predict outcomes for future situations. In this process, decision tree is used as a predictive model which decides the value of given data sets.

## 2.7. Apriori Algorithm

Apriori[9] is an algorithm for frequent item set mining and association rule learning over transactional databases. It proceeds by identifying the frequent individual items in the database and extending them to larger and larger item sets as long as those item sets appear sufficiently often in the database. The frequent item sets determined by Apriori can be used to determine association rules which highlight general trends in the database: this has applications in domains such as market basket analysis

## 3. Experiment

### 3.1 Apriori Algorithm

Table 1: Rule Extraction of Apriori Algorithm

virus	Rule (13window)	Rule (17window)	Rule (19window)
Feline	L 121	L 100	L 89
Rat	L 117	L 96	L 83
Turkey	L 91 / V 76	L 70 / V 59	L 70 / V 49
SARS	L 118	L 95	L 84
MERS	L 100	L 77 / S 62	L 67 / S 55

As mentioned in Table 1, through Apriori Algorithm, we could find some properties among the subject viruses. The very common property was Leucine. Due to this point, we can assume that they are so functionally close enough that a little gene-change in the subjects can cause a significant result as in the case of HKU4 and MERS. However, there were some differences in Turkey and MERS: Turkey had Valine and MERS had Serine

### 3.2 Decision Tree

Class1 refers to Feline coronavirus, class2 to Rat coronavirus, class3 to Turkey coronavirus, class4 to SARS, and class5 to MERS.

Table 2: Rule Extraction of Decision Tree (P=Position)

virus	Rule (13window)	Rule (17window)	Rule (19window)
Feline	P2=C P6=V P9=L	P7 = L P9 = K P11 = L	P3= D P13 = H
Rat	P7=V P9=L P12=L	P7 = P P10 = L	P7 = S P10 = L P13 = L
Turkey	P6=R P8=L P11=L	P7 = L P9 = T	P10 = Q P14 = M
SARS	P6=T P7=D	P4 = Q P7 = D	P3 = D P10 = D
MERS	P6=D P10=Q	P7 = W P9 = D	P10 = D P19 = A

As mentioned in Table 2, in 13 and 17window results, Leucine as the main amino acid was common property among class 1, 2 and 3. However, there were some differences in class 4 and 5: their main amino acid was D which refers to Aspartate rather than Leucine. Due to this point, we could assume that D is the important factor that differs class 4 and 5 from others. Also in 19window results, we found that the main amino acid in class 4 and 5 was D rather than L. Due to this point, we could assume that D is the important factor that differs class 4 and 5 from others. We assumed that D may have critical role in viral entry in human cell, giving the class 4 and 5 its human infectious property which is distinctive from the class 1, 2, and 3.

Interestingly, class 1 also had D as its main amino acid and the position was same as the class 4's. This indicates that Feline has possibility, to some extent, to perform in similar way as the SARS does. However, L, which was dominant in class 1, 2, and 3 in previous windows appeared only in the class 2.

## 4. Conclusion

In Apriori Algorithm, all subjects in 13, 17, and 19window, even though there was a difference in the value itself, every virus commonly included L, while distinctively Turkey coronavirus showed V as well in all three windows, and MERS showed S in 13 and 19 window. Basing on the fact that Leucine is the most frequent amino acid in all subjects, we assumed that this results indicate that they are close enough so that small but appropriate change in their gene can make them perform in very similar way, such as proper attach on human cell surface protein and also the viral entry. In Decision tree, class1 refers to Feline coronavirus, class2 to Rat coronavirus, class3 to Turkey coronavirus, class4 to SARS, and class5 to MERS.

According to the result, class 4, and 5, whose hosts include human, showed high frequency of D in all of the three windows compared to the class1,2, and 3, whose main amino acid was L in 13 and 17windows. Considering that, there is possibility that the ones whose spikes contain D are more appropriate in attaching and entering in human cell. Although the results mostly showed a high frequency of L in class1,2, and 3 in 13 and 17window, D and H was the most frequent property in class 1 in 19window, having the D as the same position as class 4. Same position and amino acid sequence(D) of Feline and SARS in 19window indicate a probability that Feline coronavirus has a potential to act similarly like SARS.

## 5. References

- [1] Taharaguchi, Satoshi; Soma, Takehisa; Hara, Motonobu (2012). "Prevalence of Feline Coronavirus Antibodies in Japanese Domestic Cats during the Past Decade". *Journal of Veterinary Medical Science*
- [2] John Antoniw, Mike Adams. "Members of Genus: Alphacoronavirus".
- [3] Rottier, Peter J. M.; Nakamura, Kazuya; Schellen, Pepijn; Volders, Haukeline; Haijema, Bert Jan (2005). "Acquisition of Macrophage Tropism during the Pathogenesis of Feline Infectious Peritonitis is Determined by Mutations in the Feline Coronavirus Spike Protein". *Journal of Virology* **79**(22): 14122–30.
- [4] Thiel V (editor). (2007). *Coronaviruses: Molecular and Cellular Biology* (1st ed.). Caister Academic Press. ISBN 978-1-904455-16-5.
- [5] Chan-Yeung M, Xu RH (November 2003). "SARS: epidemiology". *Respirology* (Carlton, Vic.) **8** (Suppl): S9–14
- [6] Yang M, Li CK, Li K, Hon KL, Ng MH, Chan PK, Fok TF (August 2004). "Hematological findings in SARS patients and possible mechanisms (review)". *International Journal of Molecular Medicine*
- [7] Wang, Qihui, et al. "Bat origins of MERS-CoV supported by bat coronavirus HKU4 usage of human receptor CD26." *Cell host &microbe* **16.3** (2014): 328-337.
- [8] Quinlan, J. R. "Induction of Decision Trees." *Machine Learning* **1.1** (1986): 81-106
- [9] Rakesh Agrawal and Ramakrishnan Srikant Fast algorithms for mining association rules in large databases. *Proceedings of the 20th International Conference on Very Large Data Bases, VLDB*, pages 487-499, Santiago, Chile, September 1994.