

# Analysis and Comparison of Polyproteins of Foot-and-Mouth Disease Virus (FMDV) Type O, A, and Asia-1, Using Apriori Algorithm and Decision Tree

Minkyong Kang, Jiwon Park and Taeseon Yoon<sup>+</sup>

International, Hankuk Academy of Foreign Studies, Wangsanri Yongin, Republic of Korea

**Abstract.** The Foot-and-Mouth Disease (FMD) is a highly infectious viral disease that is common among not only domesticated animals but also wild animals. Its several symptoms include high fever and blisters. FMD is caused by Foot-and-Mouth Disease Virus (FMDV), and it has seven serotypes: O, A, C, SAT-1, SAT-2, SAT-3, Asia-1. Among these types, type O, A, and Asia-1 are endemic to Southeast Asian region. Recently, in South Korea, FMDV type O and A occurred simultaneously for the first time, bringing big disorder in stockbreeding farmhouses. FMDV type O and A showed significant difference in their spreading speed. In addition to type O and A, it was found that type Asia-1 also shows different spreading speed. We hypothesized that the cause of different spreading speed lies on the different amino acid structures of FMDV type O, A, and Asia-1, and conducted experiments to prove heterogeneity between three types of FMDV. Our experiments include two methods: Apriori algorithm and decision tree. We extracted results that showed the significant differences of the three types of FMDV, mainly focusing on positions and frequencies of amino acids of polyproteins in each of the 3 categories we have chosen. We divided the amino acid sequences in polyprotein up into 17, 13, and 9 windows. Through these experiments, we proved our hypothesis and presented a very probable conclusion that one of the possible causes of different spreading speed can lie on the different amino acid structures. Moreover, we suggested necessity of further research about vaccines that are specialized to different types of FMDV in order to increase the efficiency of vaccines.

**Keywords:** FMD(Foot-and-Mouth Disease), FMDV(Foot-and-Mouth Disease Virus), FMDV type O(Foot-and-Mouth Disease Virus type O), FMDV type A(Foot-and-Mouth Disease Virus type A), FMDV type Asia-1(Foot-and-Mouth Disease Virus type Asia-1), Apriori algorithm, Decision Tree

## 1. Introduction

Since 2000, stockbreeding farmhouses in South Korea have suffered from sporadic outbreak of foot-and-mouth disease (FMD). Until 2015, about 3,800,000 stocks (e.g. cow, pigs, goats) were culled, and economic losses reached about 3.5 trillion won. The social anxiety about FMD was increased, and FMD occurred again throughout Gyeonggi-do and Chungcheongbuk-do in 2017. Especially, this outbreak of FMD has great implications because FMDV type O and A occurred at the same time for the first time in South Korea. Whereas 4 additional outbreaks of type O FMDV in Boeun, Chungcheongbuk-do were found, no more A type FMD was detected in Yeoncheon, Gyeonggi-do. We came up with the question about the cause of difference of spreading speed between type O and A. Therefore we researched about the endemic types of FMDV in South Korea and its surrounding region Southeast Asia. We found out that FMDV type O, A, and Asia-1 are endemic to Southeast Asia, and their spreading speed shows differences. Until now, the exact cause of their different spreading speed is unknown, and we hypothesized that its cause lies in their different protein sequences. Using Apriori algorithm and decision tree, we analyzed and compared the polyprotein amino acid sequences of FMDV type O, A, and Asia-1.

---

<sup>+</sup> Corresponding author  
E-mail address: [tsyoon@hafs.hs.kr](mailto:tsyoon@hafs.hs.kr).

## 2. Materials

### FMDV

Foot-and-mouth disease (FMD, *Apathae epizooticae*) is a highly infectious viral disease. It affects domesticated animals and wild animals in Asia, Africa, and South America, especially cloven-hoofed animals. It causes high fever and blisters inside the mouth and foot. FMD is caused by Foot-and-Mouth Disease Virus (FMDV), which is the picornavirus, the member of the Aphthovirus genus. There are seven serotypes of FMDV: O, A, C, SAT-1, SAT-2, SAT-3, and Asia-1. These serotypes differ in their aspects of regions where they mainly occur, spreading speed, and etc. Serotype O occurs most commonly, and type O, A, and Asia-1 are endemic to Southeast Asia.

### FMDV type O, A, and Asia-1

In Southeast Asia, type O, A, and Asia-1 mainly occur. Among them, type O occurs most frequently, but type A and Asia-1 are also meaningful that substantial amounts of animals are affected by those types. They differ in their spreading speeds. Type O is the most common type of FMDV and also the most rapid-spreading type of FMDV. Type A spreads slower than type O, as shown in the case of South Korea. Type Asia-1 is currently restricted to Asia, so their occurrence data is not enough to come up with information about its spreading speed. However, certainly its speed is slower than Type O.

## 3. Materials

### Apriori Algorithm

Apriori algorithm is an algorithm used in data mining for learning association rules. It finds out frequent items and extends them into larger subsets as those items are identified repeatedly in the database. Typically, marketing strategies apply association rules to analyze the consumption patterns of consumers. Markets place their items that are frequently sold together next to each other; Amazon recommends products that are purchased often with items that the consumer chose. Apriori thus discovers the general pattern between each data in the whole database.

### Decision Tree

A decision tree is a branching method to predict every potential outcomes of a decision. The tree always begins with a specific decision. Lines are drawn right next to the decision to indicate all possible options, and results are shown at the end of each option. If the result turns out as another decision, the same process is reiterated. Thus, the outermost branch becomes the ultimate result of the decision.

## 4. Experiments

### Apriori Algorithm

We used the Apriori Algorithm to compare and contrast the amino acid features of three types of FMDV: type O, type A, and type Asia-1. We first figured out the frequencies of amino acids in each type of the viruses. Through Apriori algorithm, we gained data for each window of the division in 17, 13 and 9 windows. The results are as follows.

Table 1 17, 13, 9 window Apriori Experiment Results (amino acid)

	Window-17			Window-13			Window-9				Window-17			Window-13			Window-9		
	Type O	Tyoc A	Type Asia-1	Type O	Tyoc A	Type Asia-1	Type O	Tyoc A	Type Asia-1		Type O	Tyoc A	Type Asia-1	Type O	Tyoc A	Type Asia-1	Type O	Tyoc A	Type Asia-1
Amino Acid A									Amino Acid L										
Amino 1							27		29	Amino 1			14						
Amino 2		16								Amino 2								26	
Amino 3	17	15					18		28	Amino 3	18								
Amino 4			17							Amino 4		15							
Amino 5				18						Amino 5			19	20					
Amino 6				20	19	20				Amino 6	15		15	20				27	
Amino 7	15									Amino 7		15	14	19					
Amino 8				19				27		Amino 8			18				33	28	
Amino 9	14			18						Amino 9			15	18					
Amino 10	16			20	24					Amino 10			15	18					
Amino 11			15							Amino 11			19	21					
Amino 13		15		18	20					Amino 12	14	14	17						
Amino 16		16	14							Amino 13		15		18					
Amino 17			24							Amino 15	15	15							
Amino Acid V									Amino Acid K										
Amino 1		15	17			18				Amino 5				20	19			30	
Amino 2	18					20			27	Amino 7									
Amino 3						22				Amino Acid G									
Amino 4	15			20	22					Amino 2		15				18			
Amino 5		16								Amino 5	15								
Amino 7			16							Amino 6				18					
Amino 8								27		Amino 9	14								
Amino 9		15	17					26		Amino 12			18						
Amino 12				23		19				Amino 13		15							
Amino Acid T									Amino Acid J										
Amino 1	15									Amino 14									
Amino 4						23				Amino 15			14						
Amino 6								29		Amino 16									
Amino 8			14			20													
Amino 9																			
Amino 11		17				19													
Amino 12																			
Amino 13			15																
Amino 15		15																	
Amino 16		16																	

Through this, we can see that it is hard to find any similarities among FMDV type O, A, and Asia-1. Amino acid alanine, valine, threonine, and leucine are found commonly in three types, but their position and frequency were completely different. Also, even there were amino acids such as glycine, lysine, and aspartic acid that were found only in some types, rather than in every type. Frequencies of amino acids were all spread out as if they are absolutely different viruses. Thus, we can deduce that three types of FMDV should be regarded as different viruses and treat them with different vaccines which can effectively work with each type of FMDV.

However, type O+A vaccine that is recently used in South Korea to treat with newly found type A FMDV is not perfectly verified yet. In other words, it is not sure that O+A vaccine can prevent the occurrence of FMD perfectly. We hypothesized that this is because not enough researches about distinctive features between types of FMDV are conducted. Thus, to increase the prevention percentage of vaccines, further research about distinctive features of amino acids of different types is needed.

### Decision Tree

Table 2 Decision Tree Experiment Results

Window 17	Class 1	Rule 1	Class 2	Rule 1	Rule 2	Class 3	Rule 1	Rule 2	Rule 3	Rule 4	Rule 5	Rule 6	Rule 7	Rule 8	Rule 9		
	Amino Acid	Pos 2=E Pos 4=V		Pos 4=V Pos 13=A	Pos 2=C		Pos 2=C	Pos 2=A Pos 10=L	Pos 4=D Pos 17=E	Pos 4=R	Pos 4=H	Pos 2=Y	Pos 2=E Pos 9=D	Pos 2=D Pos 5=K	Pos 2=E Pos 4=K		
Window 13	Class 1	Rule 1	Rule 2	Rule 3	Class 2	Rule 1	Class 3	Rule 1	Rule 2	Rule 3	Rule 4	Rule 5					
	Amino Acid	Pos 1=F Pos 4=P	Pos 4=D Pos 11=A	Pos 4=A Pos 11=V		Pos 4=D Pos 11=H		Pos 4=I Pos 7=K	Pos 4=D Pos 11=E	Pos 4=W	Pos 1=A Pos 4=P	Pos 3=N Pos 4+V					
Window 9	Class 1	Rule 1	Rule 2	Rule 3	Rule 4	Rule 5	Class 2	Rule 1	Rule 2	Rule 3	Class 3	Rule 1	Rule 2	Rule 3	Rule 4	Rule 5	Rule 6
	Amino Acid	Pos 2=G Pos 5=A	Pos 3=K Pos 4=H	Pos 8=D Pos 9=A	Pos 1=E Pos 7=K	Pos 3=G Pos 5=A		Pos 4=A Pos 8=I	Pos 3=G Pos 5=V	Pos 3=K Pos 5=V		Pos 4=T Pos 8=Y	Pos 6=S Pos 8=K	Pos 6=W	Pos 1=G Pos 8=V	Pos 3=Y Pos 8=E	Pos 2=D Pos 3=I

In these experiments, class 1 refers to FMDV type O, class 2 refers to FMDV type A, and class 3 refers to FMDV type Asia-1. Rules that have frequency higher than 0.8 are selected in order to make sure that the results are highly reliable. In contrast to apriori algorithm, an experiment using decision tree provides more specific information, especially about amino acids in the inner segments of viruses. According to the results above, class 3 shows its distinctive features. First of all, it is noticeable that in experiments under 13 and 17 windows, rules containing aspartic acid and glutamic acid at the same time are extracted in class 3 while no rule with amino acid aspartic acid and glutamic acid is observed in any of other two classes. Second, class 3, which refers to FMDV type Asia-1, has its unique type of amino acid, as observed in every experiment under 9, 13, and 17 windows. For instance, rule extraction under windows 17 shows that amino acid types arginine, histidine, tyrosine, and more are observed only in class 3, which is FMDV type Asia. Therefore, it is hypothesized that the structure of amino acid of FMDV type Asia-1 is different from that of FMDV type O and A.

In similar method, amino acid structures of type O and A are differentiated. According to the rule extraction under 9 windows, it is evident that rules with glycine and alanine are extracted in class 1, which is FMDV type O, whereas no rule with any of amino acid type glycine and alanine is observed in class 2. Also, each of FMDV type O and A has its distinctive type of amino acid. According to rule extractions under windows 9 and 17, in case of FMDV type O, rules containing glutamic acid are found while no rule with glutamic acid is observed in FMDV type A. Likewise, rules containing amino acid type isoleucine are extracted in class 2, FMDV type A, unlike FMDV type O which has no rule with amino acid type isoleucine. Thus, it can be concluded that FMDV type O, A, and Asia-1 have different amino acid structures.

## 5. Conclusion and Discussion

To conclude, the main goal of our research is to find out the cause of different spreading speeds of FMDV type O, A, and Asia-1. In order to reach that goal, we hypothesized that the cause lies on the different amino acid structures of FMDV type O, A, and Asia-1. Thus, our experiments were focused on proving the heterogeneity between the three types of FMDV. We carried out experiments with Apriori algorithm and Decision tree algorithm. For the Apriori algorithm, we found out that no similar trend exists between the amino acid structures of the three FMDV at all; the positions and frequencies of commonly observed amino acids such as alanine, valine, threonine, and leucine were absolutely different. Further experiments using Decision Tree algorithm offered much more detailed results; as expected, the different types of FMDV showed their distinctive rules of amino acids. Therefore, we concluded that our results provide valid justifications to our hypothesis.

Based on our hypothesis, we further assumed that different vaccines are needed to deal with different types of FMDV, since type O+A vaccine which is recently used in South Korea in order to treat the newly found FMDV type A is not perfectly verified yet and seems to have no significant efficiency. We predicted that not enough researches on different types of FMDV induced such situation. Thus, to increase the prevention percentage of vaccines, further research on heterogeneity between each type of FMDV is needed.

## 6. References

- [1]. Bae, Junhyung, Jongjun Lee, Yeji Jang, Sooyoon Boo, and Sookyoung Lee. "Comparison of HIV-1, SIV-1, and SHIV in Protein Translation Using Apriori Algorithm." *Journal of Medical and Bioengineering* 4.2 (2015): 130-34. Web.
- [2]. Knowles, Nick J., Jijun He, Youjun Shang, Jemma Wadsworth, Begoña Valdazo-González, Hiroyuki Onosato, Katsuhiko Fukai, Kazuki Morioka, Kazuo Yoshida, In-Soo Cho, Su-Mi Kim, Jong-Hyeon Park, Kwang-Nyeong Lee, Geraldine Luk, Vladimir Borisov, Alexey Scherbakov, Anna Timina, Dashzeveg Bold, Tung Nguyen, David J. Paton, Jef M. Hammond, Xiangtao Liu, and Donald P. King. "Southeast Asian Foot-and-Mouth Disease Viruses in Eastern Asia." *Emerging Infectious Diseases* 18.3 (2012): 499-501. Web.
- [3]. "Analysis of Knowledge Set Discovery in Mining Items with Enhanced Apriori Association Algorithm." *International Journal of Science and Research (IJSR)* 4.11 (2015): 160-62. Web.
- [4]. "Simplifying Decision Trees." *Simplifying Decision Trees - ScienceDirect*. N.p., n.d. Web. 29 Mar. 2017.

- [5]. Stenfeldt, Carolina, Michael Eschbaumer, Juan M. Pacheco, Steven I. Rekant, Luis L. Rodriguez, and Jonathan Arzt. "Pathogenesis of Primary Foot-and-Mouth Disease Virus Infection in the Nasopharynx of Vaccinated and Non-Vaccinated Cattle." *PLOS ONE*. Public Library of Science, n.d. Web. 29 Mar. 2017..
- [6]. Tsai, Cheng-Fa, Yi-Chau Lin, and Chi-Pin Chen. "A New Fast Algorithms for Mining Association Rules in Large Databases." *IEEE International Conference on Systems, Man and Cybernetics* (n.d.): n. pag. Web.
- [7]. Kim, Mi Hyung, Chang-Ryong Ko, and Geonha Kim. "Costs Analysis of Carcass Burial Site Construction: Focused on the Foot and Mouth Disease 2011, South Korea." *Environmental Engineering Research* 20.4 (2015): 356-62. Web.