

# Analysis and Comparison about the Common Remedy of Respiratory Viruses through Data Mining

Chaewon Ham, Haeun Jung, Taeseon Yoon

Department of International/Natural Science, Hankuk Academy of Foreign Studies, Yongin-si, South Korea  
Email: [codnjs7271@gmail.com](mailto:codnjs7271@gmail.com)

Received May 2015

---

## Abstract

Respiratory diseases have a large proportion among those various diseases. Among those, the main diseases that we are now dealing with are viruses which have no majority vaccine found: Human Rhinovirus 14 (HRV), Human Coronavirus OC43 (HCoV), Respiratory Syncytial Virus (RSV), and Human Para influenza virus 1(HVJ). Even though the body can cure most of these viruses by itself, there are some incidents which end up with death. Starting an experiment with those reasons, we separated viruses by the basic symptoms and appearances, and by using data mining, we found similarities and differences of various sequences. As a result, having a high frequency, decision tree prove that each sequences are too different from each other, but still decision tree only shows the difference of the sequences. According to apriori algorithm, it could be able to find a remedy which can block amino acid L, Leucine.

## Keywords

Rhinovirus, RSV, HVJ, HCoV, Vaccine

---

## 1. Introduction

Respiratory diseases have a large proportion among those various diseases. Among those, the main diseases that we are now dealing with are viruses which have no majority vaccine found: Human Rhinovirus 14 (HRV), Human Coronavirus OC43 (HCoV), Respiratory Syncytial Virus (RSV), and Human Para Influenza Virus 1 (HVJ). Even though the body can cure most of these viruses by itself, there are some incidents which end up with death. Starting an experiment with those reasons, we separated viruses by the basic symptoms and appearances, and by using data mining, we found similarities and differences of various sequences. As a result, having a high frequency, decision tree prove that each sequences are too different from each other, but still decision tree only shows the difference of the sequences. According to apriori algorithm, it could be able to find a remedy which can block amino acid L, Leucine.

## 2. Materials & Methods

### 2.1. Coronavirus

Coronavirus has a single strain RNA for genetic material and symmetric spiral nucleocapsid. It is also a virus with an envelope. Especially, SARS coronavirus among corona virus has S amino acid and hemagglutinin esterase on the envelope. These amino acids help virus to attach on the cell membrane.

Coronavirus spins the vertebrates as hosts and causes various diseases. Among a variety of coronavirus, only 6 of them are known to infect human. This virus usually infects upper airway among respiratory system and gastro-intestinal tract. However, SARS coronavirus infects both upper airway and lower airway for its unique pathogenesis.

Corona virus is known as a major cause of the common cold in adults which appears mostly during spring and winter. However, unfortunately, the culture in the laboratory is difficult to accurately determine the affection to cold. It can cause viral pneumonia or bacterial pneumonia if it gets serious.

There is a vaccine for coronavirus that infects dogs, but for now there is no vaccine or remedy for human. Fortunately, the recent study shows that an inhibitory effect of chemical compound K22 on the proliferation of the coronavirus and the therapeutic agent is likely to be developed.

### 2.2. Human Respiratory Syncytial Virus (RSV)

RSV has a single strain RNA for genetic material it is also a virus with an envelope [1]. On the envelope, there are amino acid F and G. Amino acid F causes syncytia formation by inducing the fusion of the virus and cell. Plus, amino acid G helps RSV to attach on nearby cell's membrane.

Spread mainly by physical contact, and it has an incubation period about 5 days. It infects both upper and lower airway, but to adults, there is no serious symptoms [2]. However, in infants and premature babies with lowered immunity, it is a major cause of pneumonia and bronchiolitis, acute respiratory infections. Some infants with this virus can have asthma even after they grow up.

FI-RSV was made as a vaccine for RSV, but it was found out that it exacerbates the disease.

There are three studies for countermeasures of RSV. The first study uses passive immunization. This uses amino acid F and G on the envelope which has a major role in the initial infection of RSV. We use palivizumab for infants with low immune system, the monoclonal antibody which targets amino acid F.

Second study is about using antiviral. There is Rivavirin the antiviral for RSV, but the effect is not certain.

The last study uses active immunity, and this is still in the middle of the process. Among the vaccines, recombinant vaccine is a vaccine injected into the nasal cavity which combines attenuated recombinant RSV mutants. Moreover it uses amino acid F as an antigen.

### 2.3. Parainfluenza Virus (HVJ)

Parainfluenza virus has a single RNA for genetic material and has an envelope. On the envelope, there are F protein, M protein, P protein and spike protein. It also has 5 serotypes.

When the virus infects adults, it exhibits upper respiratory inflammation and when children are infected they exhibit bronchiolitis and pneumonia. Its main symptom is acute laryngotracheobronchitis but in 2014, as it gained heat, the scale of the virus became wider and showed severe symptoms such as pneumonia, bronchiolitis and degeneration of asthma. When the acute laryngotracheobronchitis is aggravated, it usually ends up in progressive cough that accompanies stridor, hyperventilation and inspiratory retraction. It develops phlegm, but in older age groups, it only develops slight symptoms.

There is no vaccine but it is discovered that ultraviolet rays can deactivate the virus.

### 2.4. Rhinovirus

Rhinovirus has a single RNA for genetic material and does not have an envelope. It has more than 100 serotypes, so it is hard to prevent it by vaccine.

It appears regardless of seasons and it usually appears on spring and autumn. It has a very low degree of heat and acid tolerance and mostly ends up in upper respiratory inflammation. When the pH goes below 6, the virus is deactivated unlike other enteroviruses. Until adulthood the host contains neutralizing antibody of almost all serotypes. Furthermore, regardless of serotypes the immunity last for 2 - 16 weeks. Rhinovirus adheres to spe-

cific cell acceptor to infect the cell. In particular, self-inoculation after contact between hands and touching the conjunctiva and nasal mucosa happen the most. The incubation period is almost zero and its main hosts are mammals. Acute upper respiratory inflammation and the mucus glands in the lower nasal mucosa show hyperactivity state and congest the nasal concha and close the exit of the paranasal sinuses. Children are frequently infected, and last for 4 - 9 days. It does not display lower respiratory inflammation but shows bronchiolitis, pneumonia and asthma. Without side effects, it disappears in a short term. However ear infection, acute sinusitis and complications can occur due to closure of Eustachian tube and exit of paranasal sinuses.

Infection of rhinovirus usually does not require treatment. However, antibacterial antibiotics are required if bacterial complications occur.

### 2.5. Apriori Algorithm

The apriori algorithm is mainly used to find association rules in data mining process. The algorithm drains the elements that are repeated in a section and extends to wider range and find the repetition of the same element [3]. This process shows the overall disposition of the data. It also enables to compare the association rules of the various data groups.

### 2.6. Decision Tree Algorithm

The Decision Tree algorithm is used in rule mining. The algorithm continues to find a common node, the root node, with the categorized data and find features that can bind the respective data into a specific group [4]. Branch points are divided by the binary code and lead to the next branch point until it reaches the last node shown on the screen [5]. Although the whole process is not shown in the result, the apparent rule allows the prediction of the overall structure of the data.

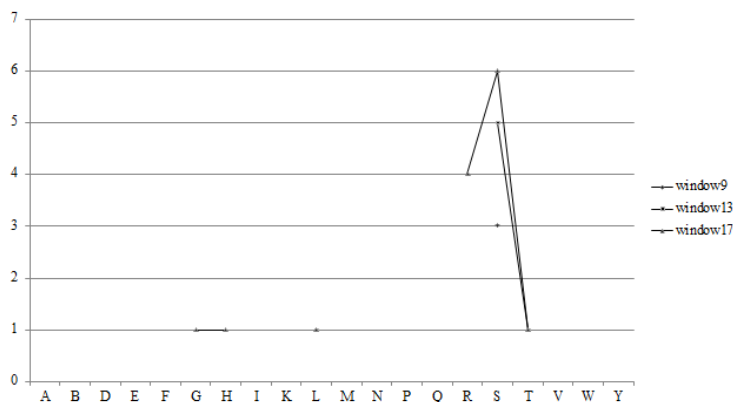
## 3. Results & Discussions

### 3.1. Apriori Results

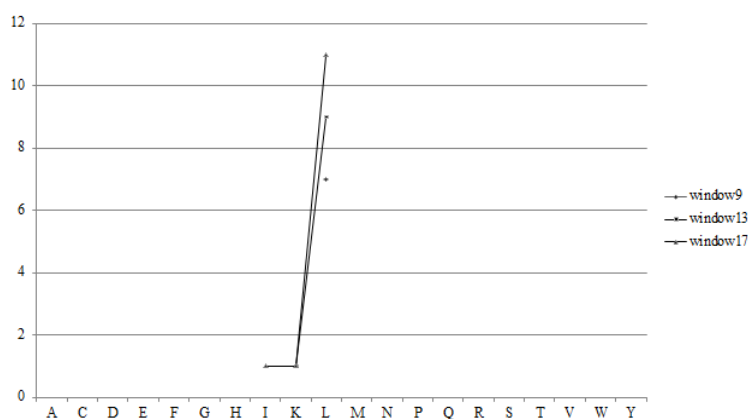
In **Figure 1**, we can know that among the amino acids, leucine received highest level. Isoleucine and lysine emerged as one. Moreover, the rest of the amino acid did not come out. In **Figure 2**, as **Figure 1**, we found out that it has a high level of leucine. Uniquely, it has valine even though its level is low. **Figure 3** has a high level of Serine. Except that, it has lysine, leucine and Isoleucine as **Figure 1** does. Finally, in **Figure 4**, as **Figure 3**, it has high level of Serine. Moreover not like other viruses, it has arginine, threonine and glycine.

### 3.2. Decision Tree Results

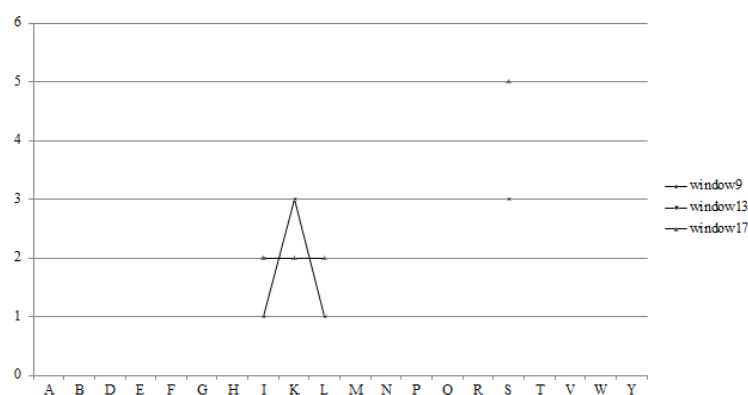
We experimented by using 10 fold cross-validation, and we figure out that each sequences did not follow the rule of other sequences. There was no data set that did not show any rule, but 17window of **Table 2** only showed 2 rules. In contrast, **Table 4** showed a lot of rules compared to other sequences. This is an error occurred from the problem in the method of experiment. **Table 1** (parainfluenza virus), **Table 2** (OC43), **Table 3** (RSV), **Table**



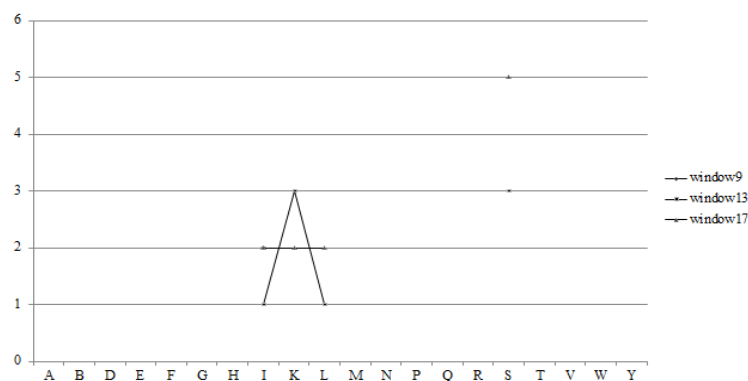
**Figure 1.** Human parainfluenza virus 1.



**Figure 2.** Human coronavirus OC43.



**Figure 3.** Respiratory syncytical virus.



**Figure 4.** Human rhinovirus 14.

**4** (rhinovirus) had different length of sequences and to extract the result we had to amplify the sequence except for **Table 2**. Especially, **Table 4** was amplified 4 times and had advantage in drawing out rules. In contrast, **Table 2** was hard to draw out the rules. Also, unlike the average result, the frequency of class 2 is very high, which means the sequences are very different.

Amino acid T which is shown considerably in **Table 1** is threonine. For the outbreak of parainfluenza, virus L protein which activates NF- $\kappa$ B is needed and this protein requires AKT1 [6].

Amino acid F which is shown considerably in **Table 3** is Phenylalanine. It mediates viral proteins becoming viral particles.

**Table 1.** Human parainfluenza virus 1.

Human parainfluenza virus 1		
Window	Rule	Frequency
Window 9	pos1 = E pos6 = L	0.929
	pos8 = T pos1 = E pos6 = L	0.923
	pos4 = M pos5 = R pos1 = E	0.917
Window 13	pos9 = T	0.909
	pos9 = T pos6 = K pos12 = G	0.917
Window 17	pos5 = L	0.941
	pos15 = Q	0.933
	pos15 = Q	9.292
	pos15 = Q	0.923
	pos2 = Y	0.917
	pos1 = P pos2 = Y	0.909

**Table 2.** Human corona virus OC43.

Human parainfluenza virus 1		
Window	Rule	Frequency
Window 9	pos3 = D pos5 = L pos3 = F	0.917
	pos6 = L pos3 = V pos6 = V	0.917
	pos3 = V pos6 = V pos1 = L	0.909
	pos4 = A pos2 = F pos5 = V	0.909
	pos5 = A pos8 = V pos3 = F	0.909
	pos6 = L	0.909
Window 13	pos2 = I pos7 = V	0.929
	pos2 = I pos7 = V	0.923
	pos7 = A pos12 = L	0.917
	pos7 = V pos12 = F	0.909
Window 17	pos2 = A pos4 = S	0.909

**Table 3.** Respiratory syncytical virus.

Human parainfluenza virus 1		
Window	Rule	Frequency
Window 9	pos6 = Y pos8 = S	0.917
	pos5 = Y pos9 = Q pos6 = Y	0.909
	pos8 = S	0.909
Window 13	pos5 = Y pos12 = Y	0.909
	pos5 = Y pos12 = Y	0.917
Window 17	pos7 = K pos10 = S	0.917
	pos15 = N pos2 = F pos4 = R	0.909

## 4. Conclusion

According to apriori algorithm, there are two main features. Because every virus had that amino acid, it means that it could be the big reason of respiratory diseases. Unfortunately, the studies about the connection between

**Table 4.** Human rhinovirus 14.

Human parainfluenza virus 1		
Window	Rule	Frequency
Window 9	pos3 = C pos7 = W	0.929
	pos3 = C pos7 = W	0.923
	pos3 = S pos9 = W	0.923
	pos8 = C pos3 = C pos7 = W	0.917
	pos3 = C pos7 = W	0.909
Window 13	pos11 = W pos12 = D	0.929
	pos9 = T pos12 = W	0.929
	pos11 = W pos12 = D pos6 = R	0.923
	pos12 = C pos9 = T pos12 = W	0.923
	pos9 = T pos12=W pos11 = W	0.917
	pos12 = D pos13 = I	0.917
	pos1 = N pos2 = S	0.962
Window 17	pos1 = N pos2 = S	0.958
	pos1 = N pos2 = S	0.957
	pos1 = N pos2 = S	0.952
	pos2 = S pos12 = N	0.941
	pos4 = A pos7 = R	0.941
	pos2 = S pos12 = N	0.938
	pos2 = K pos8 = P	0.929
	pos2 = K pos8 = P pos8 = S	0.923
	pos14 = W pos4 = H pos7 = R	0.923
	pos2 = F pos7 = H pos8 = N	0.917
pos5 = C pos9 = F	0.909	

those amino acids and respiratory virus are not in this period. We could not find the main work of those amino acids and how to make remedy that reacts with amino acids and block them. By common features shown in apriori algorithm and considering the difference shown in decision tree algorithm, common vaccine can be found but still, it should be studied more.

## References

- [1] Schijf, M.A., Lukens, M.V., Kruijssen, D., van Uden, N.O.P., Garssen, J., Coenjaerts, F.E.J., van't Land, B. and van Bleek, G.M. (2013) Respiratory Syncytial Virus Induced Type I IFN Production by pDC Is Regulated by RSV-Infected Airway Epithelial Cells, RSV-Exposed Monocytes and Virus Specific Antibodies. *PLoS One*, **8**, e81695. <http://dx.doi.org/10.1371/journal.pone.0081695>
- [2] Costello, H.M., Ray, W.C., Chaiwatpongsakorn, S. and Peeples, M.E. (2012) Targeting RSV with Vaccines and Small Molecule Drugs. *Infect Disord Drug Targets*, **12**, 110-128. <http://dx.doi.org/10.2174/187152612800100143>
- [3] Press, W.H. and Teukolsky, S.A. (2007) Support Vector Machines. In: *Numerical Recipes: The Art of Scientific Computing*, 3rd Edition, Cambridge University Press, New York.
- [4] Pourrut, X., Kumulungui, B., Wittmann, T., Moussavou, G., Délicat, A., Yaba, P., Nkoghe, D., Gonzalez, J.P. and Leroy, E.M. The Natural History of Ebola virus in Africa. *Microbes and Infection*, **7**, 1005-1014. <http://dx.doi.org/10.1016/j.micinf.2005.04.006>
- [5] Pourrut, X., Délicat, A., Rollin, P.E., Ksiazek, T.P., Gonzalez, J.P. and Leroy, E.M. Spatial and Temporal Patterns of Zaire Ebolavirus Antibody Prevalence in the Possible Reservoir Bat Species. *The Journal of Infectious Diseases*, **196**, S176-S183. <http://dx.doi.org/10.1086/520541>
- [6] Shaikh, F.Y., Cox, R.G., Lifland, A.W., Hotard, A.L., Williams, J.V., Moore, M.L. Santangelo, P.J. and Crowe, J.E. (2012) A Critical Phenylalanine Residue in the Respiratory Syncytial Virus Fusion Protein Cytoplasmic Tail Mediates Assembly of Internal Viral Proteins into Viral Filaments and Particles. *MBio*, **3**, e00270-11. <http://dx.doi.org/10.1128/mbio.00270-11>