MATEC Web of Conferences **69**, 01005 (2016) DOI: 10.1051/matecconf/20166901005
ICCPE 2016

Deeper understanding of Flaviviruses including Zika virus by using Apriori Algorithm and Decision Tree

Youjin Yang[1], Bokyung Gu[2] and Taeseon Yoon[3]

[1]*Nature Science Course, Hankuk Academy of Foreign Studies, Yongin, South Korea*
[2]*Nature Science Course, Hankuk Academy of Foreign Studies, Yongin, South Korea*
[3]*Hankuk Academy of Foreign Studies, Yongin, South Korea*

Abstract. Zika virus is spreaded by mosquito. There is high probability of Microcephaly. In 1947, the virus was first found from Uganda, but it has broken out all around world, specially North and south America. So, apriori algorithm and decision tree were used to compare polyprotein sequences of zika virus among other flavivirus; Yellow fever, West Nile virus, Dengue virus, Tick borne encephalitis. By this, dissimilarity and similarity about them were found.

1 Introduction

Today Zika virus is being a very big problem. In Brazil, more than 1.5 million people have become infected since April, 2015. The Zika virus was first found in Uganda in 1947 [1, 2]. As time goes by, zika virus arose in many countries like Malaysia, Micronesia, and Tahiti island in the South Pacific. Now, in recent 2 months, Zika virus has been found in about 39 countries. And, because of zika virus, few people died [3]. Not only Zika virus, but also other viruses (like west nile virus, yellow fever, tick borne encephalitis, and dengue virus) are fatal to people. And all of them belong to flavivirus. Unfortunately, there is no specific treatment of the viruses. Actually, vaccines exist for yellow fever and dengue virus [4, 5] but other viruses (west nile virus, tick borne encephalitis, and zika virus) don't have vaccine. And treatments are only relief from symptoms. This is why we are doing this experiment. An investigation on figuring out similarities and differences between the viruses and yellow fever may lead us to have a deeper understanding about the viruses. By using apriori algorithm and decision tree, we are able to expect to find out the possibility of a treatment that is effective to them.

2 Materials and methods

Materials we used in the experiment are flaviviruses ; Zika Virus, West Nile Virus, Yellow Fever Virus, Tick Borne Encephalitis, and Dengue Virus. Their genome sequences were gathered from the National Center for Biotechnology Information (NCBI). And we use apriori algorithm and decision tree to progress the experiment.

2.1 Zika Virus

In 1947, zika virus was first identified in Uganda rhesus monkey. [1, 2] People can find that the virus is shared to homo-sapiens from mosquito. Conducted three to seven days, people who are infected can notice only a minimal symptoms such as rash, acute fever, conjunctivitis and muscle pain appear. But, rougly 80% cases are an inapparent infection [6]. This virus has high probability of Microcephaly. So many countries warn to woman who can become pregnant [3].

2.2 West Nile Virus

Mainly west nile virus is infected by mosquito, but people can catch it from horse, crow and sparrow. Symptoms like headache, seizure, feeling stiff appear for 2 to 14 days after being infected. The virus can disturb the state of the central nervous system of the brain [7, 8].

2.3 Tick Borne Encephalitis (TBE)

Tick borne encephalitis(TBE) is infected by mite who comes down with the virus. The incubation period for the virus is about 7~14 days. It has symptoms like lack of appetite, vomiting in the incipient stage and the central nervous system like nuchal rigidity or lethargy later [7, 9, 10].

2.4 Yellow Fever

Yellow Fever is like hemorrhagic fever that is fashionable disease in Africa and North America. The arbovirus that causes yellow fever is transmitted by mosquitoes [11]. After the incubation period of three to six days, people can experience symptoms like fever, cold fit, headache, and vomiting [12].

2.5 Dengue Virus

Dengue virus exists in saliva of Aedes albopictus. When the mosquito the blood of people or animals the virus is injected into a person and then the disease goes to the person. The disease occurs 5~7 days after being bitten by mosquitoes. Fever and skin rash are the major symptoms [13-15].

2.6 Apriori algorithm

The Apriori Algorithm is the one of Algorithms used for data mining [14, 16]. In this research, We used apriori algorithm to find frequency of polyprotein sequence of Flavivurses. We divided 3 window rules; 9-window, 13-window and 17-window. The number of window represents is the number of sequences of virus which we have disposed before applying algorithm. We can compare with Zika virus and other Flaviruses to use frequency of polyprotein sequence. In other words, we can find similarity or dissimilarity about Zika virus and the viruses.; Dengue virus, West nile virus, Yellow fever, and Tick borne encephalitis.

2.7 Decision tree

One of the common data mining methods is decision tree. It uses branching method by drawing leaves and branches to generate model which expects target variable based on several input variables [6, 16]. Decision tree is a graph that every internal node corresponds to input variables and each branch corresponds to possible outcome of the input variables. And leaf node is a value of target variable when each input node has a level of route from root node to leaf node. Decision tree stretches continuously by adding the input variables. It expands until node of subset equals to target variable or the new predictive value cannot be added caused by division. If we use decision tree, it is effective to find out differences between data [13, 17]. So in our experiment, decision tree method is applied in order to compare and contrast the viruses.

3 Result

3.1 Apriori algorithm

Using apirori algorithm, we found out results like these.

Zika Virus
1. pos3 is G 47
2. pos8 is L 47
3. pos1 is G 41
4. pos5 is G 41
5. pos7 is L 40
6. pos2 is L 38
7. pos9 is L 38

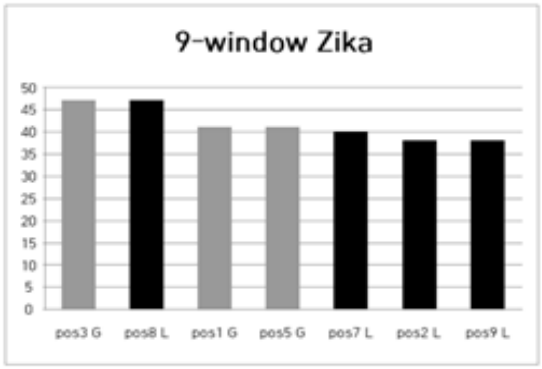

Figure 1. 9-window Zika polyprotein sequence.

Fig. 1 is the result we found out by analyzing Zika virus using apriori algorithm window-9. Results from other viruses are appeared in similar patterns. Averagely, about 40 rules were found in each virus (including 9, 13 and 17 windows), so we got total 201 rules. We analyze rules of polyprotein sequences by the appearance frequency. And then, the most symbolic rule of each virus was chosen and was used for finding similarity between 5 viruses. Tables below are the results.

Table 1. Apriori 9-window rules

Type	rule
Dengue virus	pos13=L 29
TBE	pos17=G 27
Yellow fever	pos10=L 27
West nile virus	pos1=L 26 pos13=A 26 pos15=V 26
Zika virus	pos1=G 30

Table 2. Apriori 13-window rules

Type	rule
Dengue virus	pos1=L 42
TBE	pos2=G 44 pos3=L 44
Yellow fever	pos5=L 43
West nile virus	pos7=L 46
Zika virus	pos3=G 47 pos8=L 47

Table 3. Apriori 17-window rules

class (virus)	Dengue virus	TBE	West Nile	Yellow Fever	Zika Virus
Dengue virus	86	78	92	67	54
TBE	81	64	89	82	64
West Nile	96	68	86	70	62
Yellow Fever	111	76	76	70	46
Zika virus	99	87	87	68	40

Table 1, Table 2, Table 3 show an analysis of amino acid sequences of 5 flaviviruses in 9, 13, 17 window. 9-window: Dengue virus amino1 Leucine, Tick borne encephalitis amino2 Glutamine and amino3 Leucine, Yellow Fever animo5 Leucine, West Nile amino7 Leucine, Zika Virus amino3 Glutamine and amino8 Leucine. 13-window: Dangue Virus amino1 Leucine and amino7 Valine, Tick borne encephalitis amino8 Leucine, Yellow Fever amino12 Glutamine, West Nile amino2 Glutamine and amino9 Leucine, Zika Virus amino7 and 13 Leucine. 17-window: Dengue Virus13 Leucine, Tick borne encephalitis amino17 Glutamine, Yellow Fever amino10 Leucine, West Nile amino1 Leucine and amino13 Alanine and amino15 Valine, Zika Virus amino1 Glutamine.

3.2 Decision tree

Table 4. Decision tree 9-window rules

Type	**rule**
Dengue virus	pos13=L 29
TBE	pos17=G 27
Yellow fever	pos10=L 27
West nile virus	pos1=L 26 pos13=A 26 pos15=L 26
Zika virus	pos1=G 30

Analyzing Table 4, we found out that Dengue virus is similar with other virus. On the other hand, Zika virus is different with other viruses. Dengue virus, tick borne encephalitis, west nile virus, and yellow fever virus have their own characteristics. But, lack characteristics was found at Zika virus.

Especially in Table 6, west nile virus is similar with dengue virus. Except this, similarity of other 4 viruses is high. And none of their own characteristics of all 5 flaviviruses was found.

Compared with Table 4, the number of rules of zika virus got higher at Table 5. Zika virus is similar with other viruses and has its own characteristics like dengue virus and tick borne encephalitis. In contrast, viruses such as west nile and yellow fever are different with other viruses and have less characteristics of their own.

Table 5. Decision tree 13-window rules

class (virus)	Dengue virus	TBE	West Nile	Yellow Fever	Zika Virus
Dengue virus	59	53	42	42	65
TBE	52	67	43	35	66
West Nile	62	58	39	41	65
Yellow Fever	60	51	49	36	67
Zika virus	54	50	39	50	71

Table 6. Decision tree 17-window rules

class (virus)	Dengue virus	TBE	West Nile	Yellow Fever	Zika Virus
Dengue virus	38	37	60	35	30
TBE	37	35	42	45	42
West Nile	42	41	36	40	43
Yellow Fever	30	45	45	36	45
Zika virus	36	50	40	35	41

4 Discussion and conclusion

We measured the frequency of 5 flaviviruses′ polyprotein sequences using apriori algorithm. Experiments were separated into 3(9-window, 13-window, 17-window) and were done on each flavivirus. We can find out that G(glutamine) and L(leucine) showed higher frequency of all viruses. Exceptively V(valine) was frequently found in 13-window of dengue virus, V(valine) and A(alanine) were frequently found in 17-window of West Nile virus. By comparing and contrasting zika virus with other 4 viruses, we expected that we can find out treatment of zika virus.

To know exact correlation between zika virus and other 4 viruses, we used decision tree set to 10-fold cross validation. Type of proteins that showed higher frequency is almost same at apriori. However, we can find out each virus has its own characteristics. Namely, there may be

some difficulties at combining 4 flaviviruses to develop treatment of zika virus.

So we think that further research have to be progressed.

References

1. Dick, G.w.s, S.f Kitchen, and A.j Haddow. Transactions of the Royal Society of Tropical Medicine and Hygiene 46.5 **209-20** (1952)
2. Dick, G.w.a. Transactions of the Royal Society of Tropical Medicine and Hygiene 46.5 **521-34** (1952)
3. M. Robert W., J. Homan, M. V. Callahan, J. Glasspool-Malone, L. Damodaran, A. D. B. Schneider, R. Zimler, J. Talton, R. R. Cobb, I. Ruzic, J. Smith-Gagen, D. Janies, and J. Wilson. PLoS Negl Trop Dis PLOS Neglected Tropical Diseases 10.3 (2016)
4. Mcarthur, Monica A., Marcelo B. Sztein, and Robert Edelman. "Dengue Vaccines: Recent Developments, Ongoing Challenges and Current Candidates." Expert Review of Vaccines 12.8 **933-53** (2013)
5. M. Milton, F. Da Silva Pereira Cruz, M. T. Cordeiro, M. A. Da Motta, K. M. De Melo Cassemiro, R. De Cassia Carvalho Maia, R. C. Bressan Queiroz De Figueiredo, R. Galler, M. Da Silva Freire, J. T. August, Ernesto T. A. Marques, and Rafael Dhalia. PLoS Negl Trop Dis PLOS Neglected Tropical Diseases 9.4 (2015)
6. Hayes, Edward B. Emerg. Infect. Dis. Emerging Infectious Diseases 15.9 **1347-350** (2009)
7. B. Cécile, M. Jimenez-Clavero, A. Leblond, B. Durand, N. Nowotny, I. Leparc-Goffart, S. Zientara, E. Jourdain, and S. Lecollinet. International Journal of Environmental Research and Public Health IJERPH 10.11 **6039-083** (2013)
8. Winkelmann, R. Evandro, H. Luo, and T. Wang. F1000Research F1000Res (2016)
9. M. Ecker, S. L. Allison, T. Meixner, and F. X. Heinz. Journal of General Virology 80.1 **179-85** (1999)
10. P. Vasiliki, S. Geroy, E. Diza, A. Antoniadis, and A. Papa. Vector-Borne and Zoonotic Diseases 7.4 **611-16** (2007)
11. B. Matthias, D.A. Groneberg, D. Klingelhoefer, and A. Gerver. Parasites & Vector Parasit Vectors 6.1 **331** (2013)
12. Gardner, L. Christina, and K. D. Ryman. Clinics in Laboratory Medicine 30.1 **237-60**(2010)
13. E.B. Go, S.M. Lee, and T.S. Yoon. International Journal of Machine Learning and Computing IJMLC **543-46** (2014)
14. H.S. Kim, J.Y. Yoo, and T.S. Yoon. International Journal of Computer and Communication Engineering IJCCE 294-301(2015)
15. R. Rodriguez-Roche, and Ernest A. Gould. BioMed Research International 2013 **1-20** (2013)
16. Y. S. Kim, S. M. Kim, J. W. Lee, J. Ann , J. Lim, and T. S. Yoon. Intelligent Computing Theories and Methodologies Lecture Notes in Computer Science **426-35** (2015)
17. Jang, S. P., K. H. Park, Y. L. Kim, H. N. Cho, and T. S. Yoon. Journal of Biosciences and Medicines JBM 03.06 **49-53**(2015)