FINDING RELATION BETWEEN AGING AND TELOMERE BY APRIORI AND DECISION TREE

Jieun Sung[1], Youngshin Joo, and Taeseon Yoon

Department of National Science, Hankuk Academy of Foreign Studies, Yong-In, Republic of Korea

ABSTRACT

There have been many researches about aging. Our study suggests a new research method using RTEL (Regulator of Telomere Elongation Helicase). Telomere is the very-end part of chromatid to preserve the terminal information which can be lost during DNA replication. We have compared the RTEL DNA sequence among mammals whose lifespan is all different so that we can see how RTEL DNA sequence influence lifespan and aging. Also, all of these studies were proved by apriori and decision tree. Furthermore, we can continue this experiment to Cancer. In fact, telomere is under the spotlight for the clue of cancer because a cancer cell has telomerase so that it can divide for limitlessly. So, our results could be used as background knowledge for the research of Cancer by specifying the relations between telomere and aging in mammals.

KEYWORDS

Apriori, Decision Tree, Telomere, RTEL (Regulator of Telomere Elongation Helicase), Aging, Cancer, mammal

1. INTRODUCTION

1.1. DNA EXPRESSION

DNA is the abbreviation of deoxyribonucleic acid. It is a molecule of heredity. DNA sequence is also called nucleic acid sequence. It is a consecutive letters that indicates the sequence of nucleotides within DNA. This DNA sequence is then translated to specific amino acid which forms protein. DNA sequence is translated into amino acid. 3 bases are called codon and it is translated into 1 amino acid [1][2].

1.2. DIFFERENCE BETWEEN MAMMALS

We chose 4 species of mammal as a delegate to compare the relation between lifespan and DNA sequence. Table 1 shows the different life expectancy and DNA similarities between 4 different kinds of mammal. This would be the background knowledge comparing the lifespan.

Table 1 Comparison among mammals

	Life expectancy	DNA similiarities
Mus Musculus	2-3	0.75[2]
Ovis Avis	10-12	0.77
Gorilla	35-40	0.96
Homo sapiens	66-70	1

1. 3. TELOMERE

Telomere is the very end-part of the chromatid that protects the DNA sequence during the chromosome replication [3]. During the DNA replication of the eukaryotes that have the linear chromosome, the synthesis of Okazaki fragments needs the RNA primers attached to the lagging strands. It cuts the chromosome and their sequences and implicit information might be lost. To avoid the loss, telomere is located at each end of the chromatid and it takes a role of buffer being eliminated instead of DNA. The structure of telomere was revealed by Blackburn and Szostak that the G-rich repeats protect the end of the linear plasmid of yeast. Afterward, it was identified that the telomere DNA sequence is TTAGGG [4] for the most of mammalian vertebrates. As DNA is being replicated, telomere is continually getting short.

Whereas a telomere disappears, there is an enzyme elongating the telomere named telomerase and RTEL (Regulator of Telomere Elongation Helicase). Telomerase continually adds TTAGGG to the terminal region of the telomere. It synthesizes the telomeric DNA in order to produce new telomere. In fact, telomerase is only acting in a few cells for multiple eukaryotic organisms such as stem cell, germ cell or white blood cell. For the rest, the number of cell division is limited as the division of telomere is defined. In addition, there are some theories saying that the limitation of telomere causes the senescence and it might be used for the cancer prevention. With telomere and telomerase, the DNA replication can be stable.

1. 4. CANCER AND TELOMERE

Cancer express telomerase. Telomerase catalytic protein component into telomerase-silent cells length cellular life span and build up telomerase activity. This shows that telomerase-induced telomere length manipulations have benefit for tissue engineering and dividing the molecular mechanisms including cancer [6]. Our research could be used as background knowledge for the research of Cancer as the cancer and telomerase is in great relation.

2. RESEARCH METHOD

2.1. MODELING APPROACH FOR APPLYING APRIORI TO THE DATABASE

Apriori is the algorithm which finds the relation by printing the algorithm that finds the groups of high frequency based on the minimum approval set amount. Apriori Algorithm reveals the relevance using nucleotide sequence. It works by candidate item sets of length k which are generated from length k-1, then the candidates with infrequent sub pattern are removed [7]. By apriori we could find out the high frequency group in amino acid between 4 mammals. We compared the number of each amino acid of each animal by apriori. The x axis shows the kinds of amino acid and y axis shows the number of the amino acid.

2.2. MODELING APPROACH FOR APPLYING DECISION TREE TO THE DATABASE

A decision tree is a part of an artificial intelligence to display algorithm that employs a tree diagram or model of decisions and their possible results [5]. Decision trees are usually used in operational research, specifically in decision reasoning, to help identify a strategy most likely to reach a goal. By decision tree, we found out the high frequency group in amino acid. Then, we compared the number of each amino acid of each animal. The x axis shows the kinds of amino acid and y axis shows the number of the amino acid.

3. RESULTS

In our experiment we classified the RTEL sequence into 9 window, 13 window and 17 window and compared the number of amino acid. The classification was based on the figure of the protein which is 3-dimensional.

3.1 APRIORI

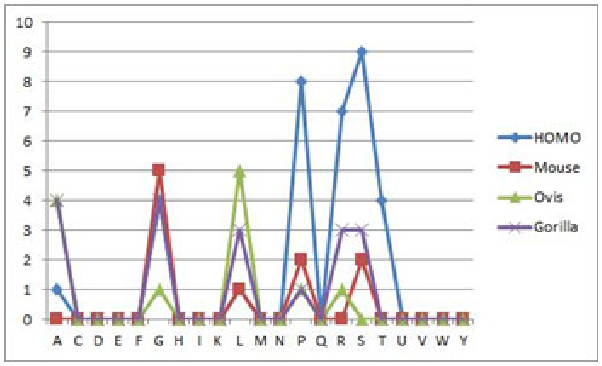

Figure 1. The results of Apriori 9

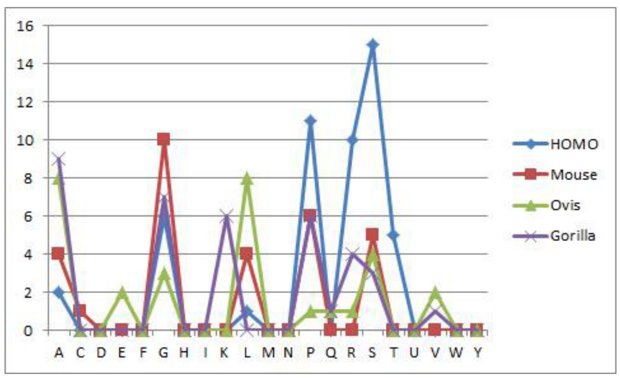

Figure 2. The results of Apriori 13

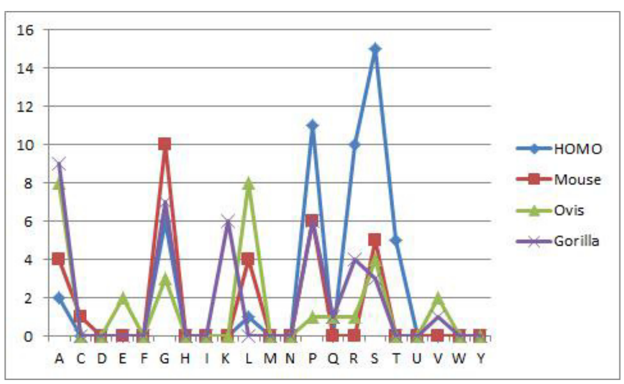

Figure 3. The results of Apriori 17

	Gorilla	Mus Musculus	Ovis Avis	Human
9 window	Alanine Glycine	Glycine	Alanine Leucine	Arginine Proline Serine
13 window	Glycine Proline	Glycine	Alanine Leucine	Proline Serine
17 window	Alanine Glycine	Glycine	Alanine	Arginine Proline Serine

Table 2. High-Frequency amino acid for each species

The tendency of amino acid expression is similar in each species. However, the most frequent amino acid is quite different among mammals. This frequency does not show the direct relation to the lifespan of each species.

3.2 DECISION TREE

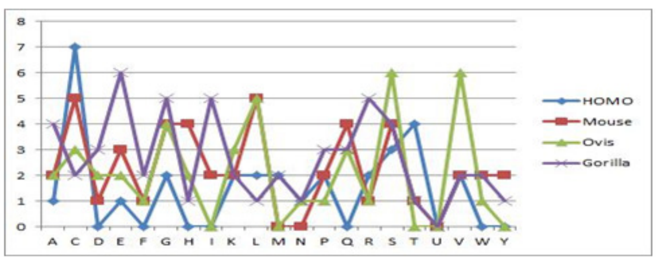

Figure 4. Decision Tree window 9

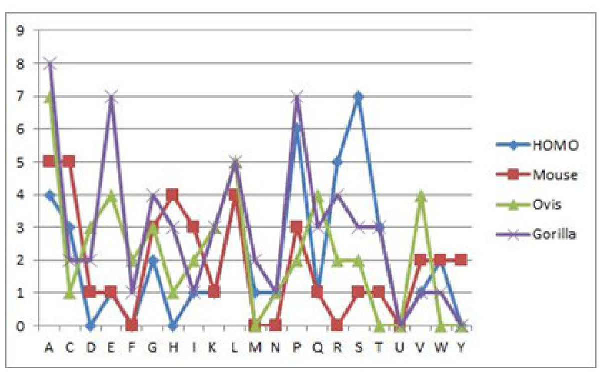

Figure 5. Decision Tree window 13

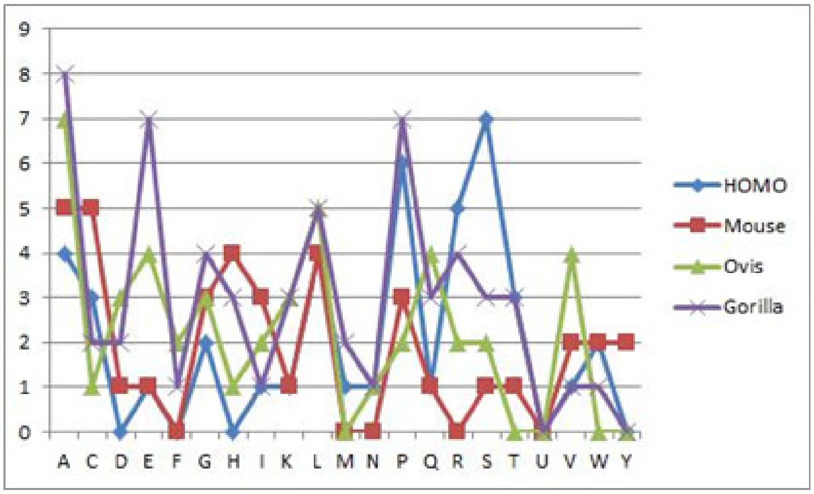

Figure 6. Decision Tree widow 17

	Gorilla	Mus Musculus	Ovis Avis	Human
9 window	Glutamic acid	Cysteine Leucine	Serine Valine	Cysteine
13 window	Glycine	Glycine	Aspartic acid	
17 window	Alanine	Alanine Cysteine	Alanine	Serine Proline

Table 3. High-Frequency amino acid for each species

The tendency of amino acid expression is similar in each species. However, the most frequent amino acid is quite different among mammals as shown in table 3. This frequency does not show the direct relation to the lifespan of each species. The result is similar with apriori analysis.

4. CONCLUSIONS

Telomere is the end-part of the chromatid used as a buffer during the DNA replication. It prevents the information loss and limits the number of cell division. Also, RTEL (Regulator of Telomere elongation helicase) is needed for telomere elongation to preserve its length. In our experiment, we tried to find the relationship between RTEL and lifespan of few species in mammal. Moreover, we found the high-frequency amino acid their correlations. In here, two methods were used: Apriori and Decision Tree.

Apriori and Decision Tree are usually used for finding the most frequent group and expecting the available results. By using them, we could find the rate of concordance in RTEL DNA sequence of 4 mammals. The frequency of amino acid among mammals was quite same, and we could find the molecular biological evidence of the mammals. On the other hand, when we see the common amino acids, we hardly found the correlation between the lifespan and amino acid. We concluded that there is no relation between them because the animals which had the similar amino acid with human differed in each apriori. Furthermore, lifespan and aging is in a deep association. RTEL is managing the elongation of telomere and we can tell that it also manages aging. Our experiment is in value of relating the sequence of the telomerase and the aging. In further experiment we would like to compare the telomerase sequence of the animals which have similar life expectancy. Also, we can support this experiment to Cancer. Telomere is receiving the attention for main factor for finding the solution of cancer. Our results would be used as background knowledge for the research of Cancer by specifying the relations between telomere and aging, lifespan in mammals.

REFERENCES

[1] "LABORATORY OF CELL BIOLOGY AND GENETICS." The Rockefeller University » Laboratory for Cell Biology and Genetics.

[2] "Lineage-Specific Biology Revealed by a Finished Genome Assembly of the Mouse." PLOS Biology.

[3] "How Telomeres Solve the End-Protection Problem." How Telomeres Solve the End-Protection Problem.

[4] “Sequence Comparison of Human and Yeast Telomeres Identifies Structurally Distinct Subtelomeric Domains” Oxford journals, 1997.

[5] "Decision Tree Classifier for Network Intrusion Detection with GA-based Feature Selection." Decision Tree Classifier for Network Intrusion Detection with GA-based Feature Selection.

[6] Shay, J. W., and W. E. Wright. "Ageing and cancer: the telomere and telomerase connection." Novartis Foundation symposium. Vol. 235. 2000.

[7] Eunby Go, Seungmin Lee, and Taeseon Yoon, “Analysis of Ebolavirus with Decision Tree and Apriori algorithm”, International Journal of Machine Learning and Computing, Vol. 4, No. 6, December 2014

AUTHORS

Jieun Sung -She was born in Korea, in 1997. She is a junior, eleventh grade student in the national science program at Hankuk Academy of Foreign studies since 2014.

Youngshin Joo -She was born in Korea, in 1997. She is a junior, eleventh grade student in the national science program at Hankuk Academy of Foreign studies since 2014.

Taeseon Yoon -He was born in Seoul, Korea, in 1972. He was Ph.D. Candidate degree in Computer education from the Korea University, Seoul, Korea, in 2003. From 1998 to 2003, he was with EJB analyst and SCJP. From 2003 to 2004, he joined the Department of Computer Education. Since December 2004, he has been with the Hankuk Academy of Foreign Studies, where he was a Computer Science and Statistics Teacher.