Short Communication

Comparative Genomics of SARS-CoV and SARS-CoV-2 Using Fuzzy Sets: Insights into Viral Therapeutic Implications

Tamil Bharathi Viswanathan[1], Anitha Devi Dinakaran[2] and *Gnanendra Shanmugam[1]

Department of Biotechnology, Vivekanandha College of Arts and Sciences of Women (Autonomous), Tiruchengode, Tamilnadu, India.

Department of Mathematics, Shri Sakthikailassh Women's College, Salem, Tamilnadu, India

*Correspondence: gnani@vicas.org (GS)

Abstract: Severe acute respiratory syndrome coronavirus (SARS-CoV) and the 2019 novel coronavirus (SARS-CoV-2) are highly infectious pathogens that primarily affect the human respiratory system, causing a range of respiratory illnesses from mild to severe. The SARS-CoV outbreak in 2003 resulted in significant global fatalities, while the ongoing COVID-19 pandemic, caused by SARS-CoV-2, has led to millions of deaths worldwide and has had profound impacts on healthcare systems and the global economy. Genomic analysis plays a crucial role in understanding these viruses and developing effective therapies. In this commentary, we discuss the use of genomic data, including high-throughput sequencing and gene expression analysis, in drug development for COVID-19. We highlight the similarities between the genomes of SARS-CoV and SARS-CoV-2, using fuzzy logic to estimate their distance. Our analysis indicates considerable similarity between the two genomes, suggesting potential commonalities in drug targets. The comparison of these genomes demonstrates the importance of genomics in understanding viral pathogenicity and developing targeted therapies. The use of fuzzy logic in genomic analysis provides a valuable tool for comparing genetic sequences and identifying potential drug targets. Continued research in genomics and bioinformatics is essential for combating current and future viral outbreaks..

Keywords: SARS-CoV, SARS-CoV-2, fuzzy logic, drug development, viral genomes.

Citation: Tamil Bharathi V, Anitha D and Gnanendra S. Comparative Genomics of SARS-CoV and SARS-CoV-2 Using Fuzzy Sets: Insights into Viral Therapeutic Implications. Int J Adv Interdis Res 2024, 04, e132.

Received | 02 March 2024
Revised | 18 March 2024
Accepted | 24 March 2024
Published | 25 March 2024

Introduction

Severe acute respiratory syndrome coronavirus (SARS-CoV) and the 2019 novel coronavirus (SARS-CoV-2) are the primary pathogens that predominantly target the human respiratory system. Their infections can cause mild respiratory illness to acute pneumonia and even respiratory failure. SARS-CoV was an extremely fatal virus that went away following extensive public health interventions [1]. The SARS 2003 epidemic ended in June 2003, with a total of 8098 cases recorded worldwide, 774 fatalities, and a case fatality rate of 97%, with the majority of cases acquired nosocomially. In contrast, the new SARS-CoV-2 virus that began in Wuhan, China, and spread over the globe caused a worldwide health emergency [2]. As of February 2022, COVID-19 has been associated to roughly 280 million illnesses and over 5.3 million deaths, with instances continuing to rise due to the lack of a viable therapy[3]. This on-going pandemic impacted negatively on healthcare systems and the

global economy, necessitating the development of efficient medicines to prevent the disease's spread and severity [4]. For the coronavirus that causes COVID19, there are currently no target-specific drugs available; however, novel therapeutic options that target the viral replication cycle are being researched [5].

The first step in discovering a novel medicine is to look for genomes that are closely similar. Recent genome sequencing projects have generated a huge amount of data on the function and structure of biological molecules and sequences [6]. We have access to a vast number of genomes including SARS-CoV and SARS-CoV-2, protein structures, and genes, all of which have their expression levels tested in tests. Genomics has opened up new avenues for drug development, notably high-throughput sequencing and characterisation of expressed genes [7]. Knowing all of the genes and their activities might lead to more effective preventative measures, as well as changes in medication research strategy and development methods. Handling such a large volume of data, which is sometimes inaccurate and ambiguous, necessitates the use of strong integrated bioinformatics systems and novel technologies. In genomic comparisons, fuzzy logic and fuzzy technologies are now commonly employed [8-10]. As a result, in this investigation, we used fuzzy sets to estimate the distance between two genomes, SARS-CoV and SARS-CoV-2, which can indicate considerable similarities between the genomes in this study.

Methods

Data Collection

Genomic sequences of SARS-CoV (NC_004718.3) and SARS-CoV-2 (NC_045512.2) were obtained from the NCBI genome database.

Genome Analysis

The complete genomes of SARS-CoV and SARS-CoV-2 were analyzed to determine their lengths, number of genes, and G+C content using in-house developed python scripts

Codon Analysis

The number of nucleotides at each of a codon's three base positions in the coding sequences of SARS-CoV and SARS-CoV-2 were calculated using R scripts

Fuzzy Set Generation

The corresponding fractions of nucleotides at each of a codon's three base positions in the coding sequences of SARS-CoV and SARS-CoV-2 were calculated and represented as fuzzy sets.

Fuzzy Set Comparison

The distance between the fuzzy sets representing the frequencies of nucleotides of SARS-CoV and SARS-CoV-2 was calculated using a specified equation, likely involving a method for

comparing fuzzy sets such as the Jaccard distance, Hamming distance, or other appropriate metric for fuzzy sets [11,12]. All calculations were executed using R scripts.

Result and Discussion

The genomes of SARS-CoV (NC_004718.3) and SARS-CoV-2 (NC_045512.2) were obtained from the NCBI genome database (https://www.ncbi.nlm.nih.gov/genome/).

The complete genome of SARS-CoV comprises of 29751 bp holding 13 genes and having 40.8% of G+C content. The number of nucleotides at the three base sites of a codon in the coding sequences is shown in Table.1 and we have the corresponding fractions (Table 2) as fuzzy set. This set can be considered as a point in the hypercube I^{12}.

Table 1: The number of nucleotides at each of a codon's three base positions in the coding sequence of SARS-CoV.

	A	T	G	C
First base	8282	8913	6015	5797
Second base	8272	8901	6035	5799
Third base	8297	8271	6038	5790

Table 2: The fractions of nucleotides at each of a codon's three base positions in the coding sequence of SARS-CoV.

	A	T	G	C
First base	0.2855	0.3072	0.2073	0.1998
Second base	0.2851	0.3068	0.2080	0.1999
Third base	0.2921	0.2912	0.2126	0.2039

The fuzzy set of genome frequencies of SARS-CoV is

(0.2855, 0.3073, 0.2074, 0.1998, 0.2852, 0.3069,
0.2081, 0.1999, 0.2922, 0.2913, 0.2126, 0.2039) $\in I^{12}$

The complete genome of SARS-CoV-2 comprises of 29903 bp holding 11 genes and having 38% of G+C content. The number of nucleotides at the three base sites of a codon in the coding sequences is shown in Table.3 and we have the corresponding fractions (Table 4) as fuzzy set.

Table 3: The number of nucleotides at each of a codon's three base positions in the coding sequence of SARS-CoV-2.

	A	T	G	C
First base	8732	9360	5722	5341
Second base	8716	9386	5709	5344
Third base	8714	9367	5388	5364

Table 4: The fractions of nucleotides at each of a codon's three base positions in the coding sequence of SARS-CoV-2.

	A	T	G	C
First base	0.3010	0.3227	0.1973	0.1841
Second base	0.3005	0.3236	0.1968	0.1842
Third base	0.3069	0.3299	0.1897	0.1889

The fuzzy set of genome frequencies of SARS-CoV-2 is

(0.3010, 0.3227, 0.1973, 0.1841, 0.3005, 0.3236,
0.1968, 0.1842, 0.3069, 0.3299, 0.1897, 0.1889) $\in I^{12}$

Using the given below Equation, it is possible to compute the distance between these two fuzzy sets representing the frequencies of the nucleotides of SARS-CoV and SARS-CoV-2.

d(SARS-CoV and SARS-CoV-2) = 0.2069/3.1162 ≈ 0.0663.

Conclusion

In conclusion, the comparison of the genomes of SARS-CoV and SARS-CoV-2 using fuzzy sets suggests a relatively low distance between the two viruses, indicating considerable similarity in their nucleotide frequencies. Despite their similarities, these viruses have distinct characteristics, as evidenced by their different pathogenicity and impact on global health. Understanding the genetic similarities and differences between these viruses is crucial for developing effective therapies and preventive measures against COVID-19. Further research into the genomic sequences and structures of these viruses is essential for combating the ongoing pandemic and preparing for future viral outbreaks.

References

[1] Reperant LA. Osterhaus ADME. AIDS, avian flu, SARS, MERS, Ebola, Zika… what next? Vaccine. 2017;35:4470–4474.

[2] Liu J, Dai S, Wang M, Hu Z, Wang H, Deng F. Virus like particle-based vaccines against emerging infectious disease viruses. Virol Sin. 2016;31:279–287.

[3] Qu G, Li X, Hu L, Jiang G. An imperative need for research on the role of environmental factors in transmission of novel coronavirus (COVID-19) Environ Sci Technol. 2020;54:3730–3732.

[4] Torres A, Nieto JJ. The fuzzy polynucleotide space: basic properties. Bioinformatics. 2003;19(5):587–592.

[5] Tomida S, Hanai T, Honda H, Kobayashi T. Analysis of expression profile using fuzzy adaptive resonance theory.

Bioinformatics. 2002;18(8):1073–1083.

[6] Schlosshauer M, Ohlsson M. A novel approach to local reliability of sequence alignments. Bioinformatics. 2002;18(6):847–854.

[7] Cordon O, Gomide F, Herrera F, Ho ´ ffmann F, Magdalena L. Ten years of genetic fuzzy systems: current framework and new trends. Fuzzy Sets and Systems. 2004;141(1):5–31.

[8] Belacel N, Cuperlovi ˇ c-Culf M, Laflamme M, Ouellette R. ´ Fuzzy J-Means and VNS methods for clustering genes from microarray data. Bioinformatics. 2004;20(11):1690–1701.

[9] Huang Y, Li Y. Prediction of protein subcellular locations using fuzzy k-NN method. Bioinformatics. 2004;20(1):21–28.

[10] Carleos C, Rodriguez F, Lamelas H, Baro JA. Simulating complex traits influenced by genes with fuzzy-valued effects in pedigreed populations. Bioinformatics. 2003;19(1):144–148.

[11] Dembel´ e D, Kastner P. Fuzzy C-means method for clustering microarray data. Bioinformatics. 2003;19(8):973–980.

[12] Heger A, Holm L. Sensitive pattern discovery with 'fuzzy' alignments of distantly related proteins. Bioinformatics. 2003;19(suppl 1):i130–i137.