OPEN ACCESS
Fractals, Vol. 29, No. 2 (2021) 2150123 (11 pages)
© The Author(s)
DOI: 10.1142/S0218348X21501231

World Scientific
www.worldscientific.com

Fractals 2021.29. Downloaded from www.worldscientific.com
by UNIVERSITI TEKNOLOGI MALAYSIA (UTM) on 03/08/22. Re-use and distribution is strictly not permitted, except for Open Access articles.

COMPLEXITY-BASED ANALYSIS OF THE ALTERATIONS IN THE STRUCTURE OF CORONAVIRUSES

HAMIDREZA NAMAZI,[*,†,¶] ALI SELAMAT[†,‡,§] and ONDREJ KREJCAR[†,‡]
[*]*College of Engineering and Science*
Victoria University, Melbourne, Australia
[†]*Center for Basic and Applied Research*
Faculty of Informatics and Management
University of Hradec Kralove, Hradec Kralove, Czechia
[‡]*Malaysia-Japan International Institute of Technology (MJIIT), Universiti Teknologi Malaysia Kuala Lumpur*
Jalan Sultan Yahya Petra, Kuala Lumpur 54100, Malaysia
[§]*Media and Games Center of Excellence (MagicX)*
Universiti Teknologi Malaysia and School of Computing
Faculty of Engineering, Universiti Teknologi
Malaysia, Johor, Malaysia
[¶]*hamidreza.namazi@vu.edu.au*

Received November 30, 2020
Accepted February 3, 2021
Published March 20, 2021

Abstract

The coronavirus has influenced the lives of many people since its identification in 1960. In general, there are seven types of coronavirus. Although some types of this virus, including 229E, NL63, OC43, and HKU1, cause mild to moderate illness, SARS-CoV, MERS-CoV, and SARS-CoV-2 have shown to have severer effects on the human body. Specifically, the recent

[¶]Corresponding author.

This is an Open Access article published by World Scientific Publishing Company. It is distributed under the terms of the Creative Commons Attribution 4.0 (CC BY-NC-ND) License which permits use, distribution and reproduction in any medium, provided the original work is properly cited.

known type of coronavirus, SARS-CoV-2, has affected the lives of many people around the world since late 2019 with the disease named COVID-19. In this paper, for the first time, we investigated the variations among the complex structures of coronaviruses. We employed the fractal dimension, approximate entropy, and sample entropy as the measures of complexity. Based on the obtained results, SARS-CoV-2 has a significantly different complex structure than SARS-CoV and MERS-CoV. To study the high mutation rate of SARS-CoV-2, we also analyzed the long-term memory of genome walks for different coronaviruses using the Hurst exponent. The results demonstrated that the SARS-CoV-2 shows the lowest memory in its genome walk, explaining the errors in copying the sequences along the genome that results in the virus mutation.

Keywords: Coronavirus; Genome Walk; Complexity; Fractal Dimension; Approximate Entropy; Sample Entropy; Memory; Hurst Exponent.

Fractals 2021.29. Downloaded from www.worldscientific.com
by UNIVERSITI TEKNOLOGI MALAYSIA (UTM) on 03/08/22. Re-use and distribution is strictly not permitted, except for Open Access articles.

1. INTRODUCTION

COVID-19 is the most dangerous disease since 2019, which has affected the lives of many people around the world. This disease which is caused by an RNA virus (SARS-CoV-2) has similar symptoms like flu.[1,2] Due to the high infection and fatality rate, investigating this virus has become one of the most important research areas in science. In fact, analysis of the virus genome is a crucial step for developing a vaccine for the disease.

An important category of works in investigating a virus is to study its evolution. Therefore, researchers look at other versions of a virus in the past and analyze how it has been changed. It is known that the coronavirus is a family of seven viruses that for the first time identified in 1960. For years, the structure of this virus has been changed. SARS-CoV-2, as the last type of coronavirus, is the most dangerous member of this family. Between different identified coronaviruses, the analysis of SARS-CoV-2, SARS-CoV, and MERS-CoV has aroused many scientists' attention due to the highest rate of infection and fatality. Reviewing the literature shows great works that focused on the analysis of SARS-CoV,[3–5] and MERS-CoV.[6–8] Specifically, we can identify some categories of works that investigated the SARS-CoV-2 genome. The studies on the analysis of the virus genomic variance,[9] tracking of virus' movements,[10] phylogenetic analysis of the virus,[11] and characterization of its genome[12] are worthy of being mentioned.

We can also find some limited studies that compared SARS-CoV-2 with other types of coronavirus. This is specifically important in comparing the SARS-CoV-2 genome with the genomes of other members of this family. The reported studies that analyzed genome composition, nucleotide, codon usage indices, relative synonymous codons usage, and the effective number of codons (ENc) between SARS-CoV-2, SARS-CoV, and MERS-CoV[13–17] are worthy of being mentioned. The common point between all these studies is that they ignored the complex structure of genome walks. To address this issue and make a more precise comparison between different coronaviruses, we analyzed the differences in these coronaviruses' complexity.

Since the genome walks have chaotic patterns, complexity-based techniques can be utilized to investigate the genome alterations among the different coronaviruses.

The fractal theory quantifies the complexity of self-affine and self-similar objects. In fact, a self-affine object does not follow the same scaling exponent in various directions, which causes its complexity. Therefore, the fractal dimension is utilized to quantify complexity.[18] Genome walks that map genome sequences into chaotic fluctuations are self-affine fractals, and therefore, we can use fractal analysis to decode their complexity.

Many works have evaluated the various kinds of time series[19–22] and images[23–25] using the fractal theory. The review of the literature shows the application of fractal theory in the analysis of genomic sequences. The studies that investigated lung cancer DNA sequences,[26] evaluated the fractal shape of DNA walks,[27] compared the DNA walks between cancerous and normal cells,[28] analyzed binary images of DNA,[29] and studied the 3D fractal assembly of DNA,[30] are worthy of being mentioned. In a recent work, we showed that the SARS-CoV-2 genome walk has larger complexity than the ones for HIV and dengue virus genomes.[31] In another investigation, we evaluated how the complexity of the SARS-CoV-2 genome walks

Fractals 2021.29. Downloaded from www.worldscientific.com
by UNIVERSITI TEKNOLOGI MALAYSIA (UTM) on 03/08/22. Re-use and distribution is strictly not permitted, except for Open Access articles.

changes among various countries[32] and cities in the USA.[33]

We also employed approximate entropy to investigate the genome walks' complexity. Approximate entropy is an index to identify the amount of complexity in time series. In fact, a time series with greater approximate entropy is more complex. Researchers have widely used approximate entropy to evaluate the complexity of different data.[34–37] However, there have a few studies that evaluated genomic data using approximate entropy. The works that identified the similarity among different DNA sequences[38] and analyzed cervical neoplasia gene-expression signatures[39] can be mentioned. In a recent work,[31] we proved that the genome walks of SARS-CoV-2 has greater approximate entropy than the ones for HIV and dengue virus.

As another index, we also employed sample entropy to analyze the alterations of the complexity of genome walks among various kinds of coronavirus. Sample entropy is independent of the length of data, and since the genome walks of various coronaviruses have various lengths, it is used in this study to verify the findings of fractal analysis and approximate entropy. Based on our search, only two studies worked on the analysis of genome sequences using sample entropy. In Ref. 40, researchers used sample entropy to predict enhancer regions from DNA walks. In Ref. 31, we showed that the genome walks of SARS-CoV-2 have larger sample entropy than the ones for HIV and dengue virus.

This work, for the first time, not only checked the difference in the genomic sequences of different coronaviruses but also decoded the hidden complexity in the genomic sequences to distinguish between these coronaviruses. Besides, by analyzing the memory of genome walks, we discussed the genomic mutation.

We detail the method of analysis in the following section. Then we present the database and the conducted analysis. The results of the analysis will be presented thereafter. The discussion and some concluding remarks will be brought in the last section.

2. METHOD

In this study, we examined the alterations among the complex structures of three dangerous types of coronavirus, namely, SARS-CoV-2, SARS-CoV, and MERS-CoV. For this purpose, first, we extracted the genome walks of these viruses. To do this job, we employed the technique that

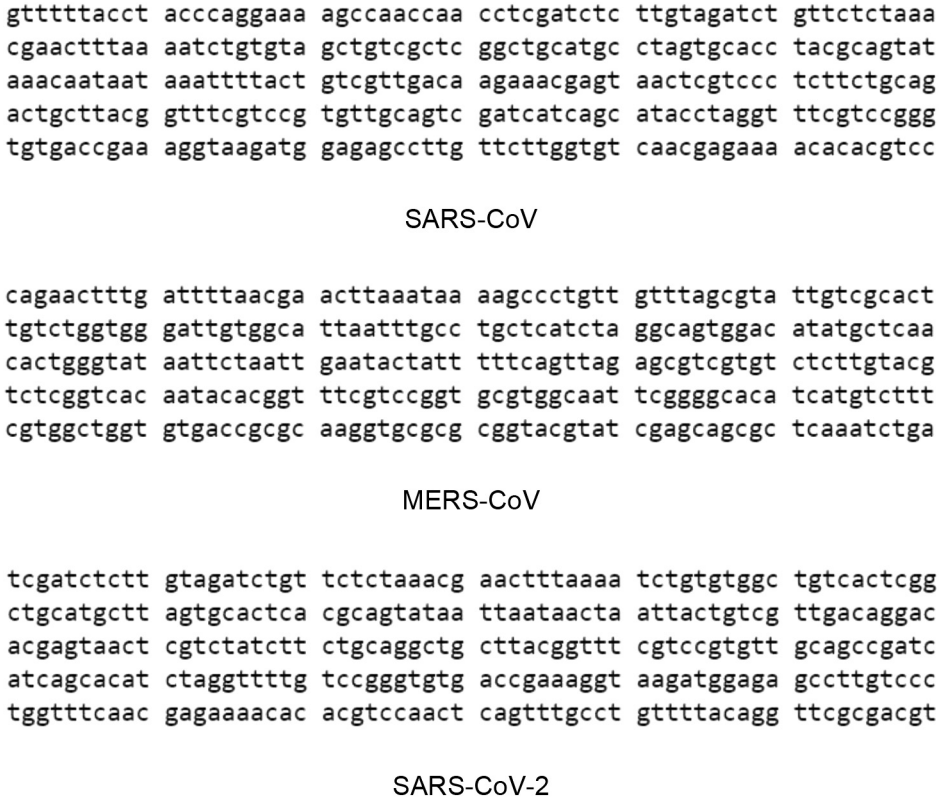

Fig. 1 Partial genome sequences for different coronaviruses.

Fractals 2021.29. Downloaded from www.worldscientific.com
by UNIVERSITI TEKNOLOGI MALAYSIA (UTM) on 03/08/22. Re-use and distribution is strictly not permitted, except for Open Access articles.

was developed by Peng *et al.*,[41] which maps the nucleotide sequences onto a walk. Peng *et al.*[41] called this map as a DNA walk. In this research, since coronaviruses are RNA-type viruses, we call the generated random walk as a "genome walk", which indicates the correlation between nucleotides along the genome chain.

Parts of different coronaviruses' genome sequences are illustrated in Fig. 1. The genome sequences in the case of each virus mainly consist of four characters A, G, C, and T, which respectively indicate adenine, guanine, cytosine, and thymine bases. In Peng *et al.*'s method, purines (A/G), and pyrimidines (C/T) are considered for the generation of genome walks.

According to this method, we mapped each purine (A/G) and pyrimidine (C/T) to -1 and $+1$, respectively, and then defined the displacement (dimensionless) using Eq. (1). As shown in Eq. (1), displacement (W) is defined as a combination of up ($c(j) = +1$) and down ($c(j) = -1$) fluctuations after N steps. By plotting displacements in different nucleotide distances, we obtained the coronavirus genome walk. As illustrated in Fig. 2, the genome walk is a random walk, and therefore, we can employ complexity techniques to analyze it.

$$W(l) = \sum_{j=1}^{N} c(j). \tag{1}$$

We employed the fractal analysis to investigate the alterations in the complex structure of different coronaviruses. We quantified the complexity of genome walks using the fractal dimension.

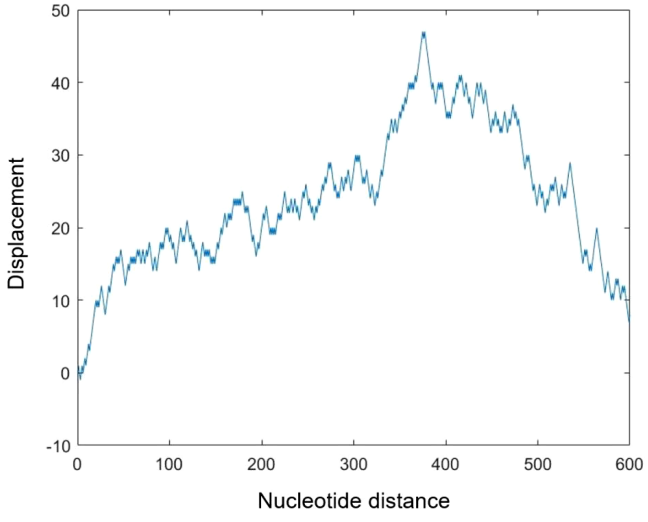

Fig. 2 Partial genome walks for the SARS-CoV-2.

The fractal dimension can be calculated using different techniques. In this work, we utilized the box-counting algorithm. In several steps, this algorithm segments the genome walk using boxes of the same size (ϕ) and count their number (N). Finally, the fractal dimension is computed using the following series of ϕ and N[42]:

$$F = \lim_{\phi \to 0} \frac{\log N(\phi)}{\log 1/\phi}. \tag{2}$$

The general fractal dimension is formulated as follows:

$$F_t = \lim_{\phi \to 0} \frac{1}{t-1} \frac{\log \sum_{i=1}^{N} p_i^t}{\log \phi}, \tag{3}$$

where t is the order of F, and p_i^t demonstrates the probability:

$$p_i^t = \lim_{l \to \infty} \frac{r_i}{l}, \tag{4}$$

where r_i is the number of occurrences in the i'th segment, and l denotes the whole nucleotide distance.

Besides, we also computed the genome walks' approximate entropy. Non-stationary biological data are usually split into short time intervals in which the physiological system can be assumed in (quasi) stationary conditions.[43] It is known that the estimation of the regularity of these data is not easy since the system dynamics cannot be explored fully. To overcome this issue, Pincus[44] introduced the approximate entropy as a rough but quite stable indicator of complexity even from short and noisy epochs.

Approximate entropy indicates the probability of none-repetition of similar patterns in the length of data.[45] Therefore, a greater value of approximate entropy indicates greater complexity in the data. Here, we define the mathematical concept of approximate entropy.[46]

If we consider a genome walk that includes "n" sample points in the form of "$(w(1), w(2), w(3), \ldots, w(n))$", we can formulate a vector in m-dimensional space in the form of

$$V(i) = [w(i), w(i+1), \ldots, w(i+m-1)]. \tag{5}$$

Considering "r" as the tolerance (filtering level), and $d[V, V] = \max_k |w(k) - w^*(k)|$, in which, $w(k)$ are the m scalar components of V, we define

Fractals 2021.29. Downloaded from www.worldscientific.com
by UNIVERSITI TEKNOLOGI MALAYSIA (UTM) on 03/08/22. Re-use and distribution is strictly not permitted, except for Open Access articles.

$P_i^m(r)$ as number of $V(j)$ such that

$$P_i^m(r) = \frac{\times d[V(i), V(j) \leq r]}{n - m + 1}. \quad (6)$$

We define $L^m(r)$ as the average of $\log(P_i^m(r))$ over $n - m + 1$:

$$L^m(r) = \frac{\sum_{i=1}^{n-m+1} \log(P_i^m(r))}{n - m + 1}. \quad (7)$$

Then approximate entropy is defined using the following equation:

$$\text{AppEntropy} = L^m(r) - L^{m+1}(r). \quad (8)$$

Since the lengths of genome walks for various types of coronavirus are different, we verified the findings of fractal analysis and approximate entropy by calculating the sample entropy of genome walks for various kinds of coronavirus. Sample entropy as the indicator of complexity is independent of data length and has been used by researchers to verify the result of other non-linear analysis techniques. Considering a signal with "n" sample points "$(w(1), w(2), w(3), \ldots, w(n))$", the embedding dimension of m and the tolerance of r, sample entropy (SamEntropy) is defined as[47]

$$\text{SamEntropy}(m, r, n) = -\log\frac{B}{C}. \quad (9)$$

In Eq. (9), B and C are defined as the number of vector pairs that $d[v_{m+1}(i), v_{m+1}(j)] < r$ and $d[v_m(i), v_m(j)] < r$ respectively, in which, $v_m(i)$ is defined using the following equation:

$$v(i) = [w(i), w(i+1), \ldots, w(i+m-1)]. \quad (10)$$

As was mentioned previously, we are also interested in studying the high mutation rate of SARS-CoV-2. As was indicated by Loewe,[48] one way to think of DNA and RNA is that they are substances that carry the long-term memory of the information required for an organism's reproduction. On the other hand, it is known that the mutation of a virus is due to mistakes in copying some characters (A, G, C, T) along its sequence.[49] Therefore, we hypothesize that these mistakes could be related to the memory of the genome. Since SARS-CoV-2 has a high mutation rate,[50] therefore, we hypothesize that its genome should have less memory than SARS-CoV and MERS-CoV. Therefore, in this study, we evaluated the memory of genome walks for various types of coronavirus using the Hurst exponent. Based on the results, we discussed the link between the variations of memory of coronavirus RNA and the memory of its genome walk. Although Hurst exponent has been used in some studies for the analysis of DNA walks, however, linking the concept of virus memory to the genome walk memory is novel.

Hurst exponent is the index of long-term memory of time series. For a time series, Hurst exponent (H) has a relationship with the fractal dimension (F):

$$H = 2 - F. \quad (11)$$

Hurst exponent can have any value between 0 and 1. The value of 0.5 for the Hurst exponent indicates a completely random process. If the value of the Hurst exponent deviates from 0.5 toward 0 or 1, it indicates the higher memory of the process.

Therefore, we computed the Hurst exponent of genome walks for different coronaviruses to evaluate which one has the lowest memory.

We computed the fractal exponent, approximate entropy, and sample entropy for the genome walks of different samples of coronaviruses to evaluate the differences in these coronaviruses' complexity. Then, we computed the Hurst exponent of genome walks to relate the virus mutation to the long-term memory of the genome in copying itself.

We also checked the non-linearities of genome walks using surrogate analysis. This technique specifies a null hypothesis, which describes a linear process and generates several surrogate data sets accordingly. A discriminating statistic (fractal dimension, sample entropy, and approximate entropy in this research) is then computed for the original (genome walks) and all the surrogate data, and if its value is significantly different between the original and the surrogate data, the null hypothesis

Table 1 Used Genome Sequences.

SARS-CoV Genome Accession	MERS-CoV Genome Accession	SARS-CoV-2 Genome Accession
AY291451	KF600620	MT079843
AY502923	KF600627	MT079846
AY502924	KF600628	MT079847
AY502925	KF600630	MT079849
AY502927	KF600634	MT079853
AY502928	KF600644	MT079854
AY502929	KF600645	MT428551
AY502932	KF600652	MT435080
FJ882945	KP719929	MT435084
FJ882948	KP719930	MT435085
FJ882951	KP719931	MT435086
FJ882952	KP719932	MT483553
FJ882957	KP719933	MT483557
FJ882958	MK357908	MT483558
FJ882961	MK357909	MT483560

Fractals 2021.29. Downloaded from www.worldscientific.com
by UNIVERSITI TEKNOLOGI MALAYSIA (UTM) on 03/08/22. Re-use and distribution is strictly not permitted, except for Open Access articles.

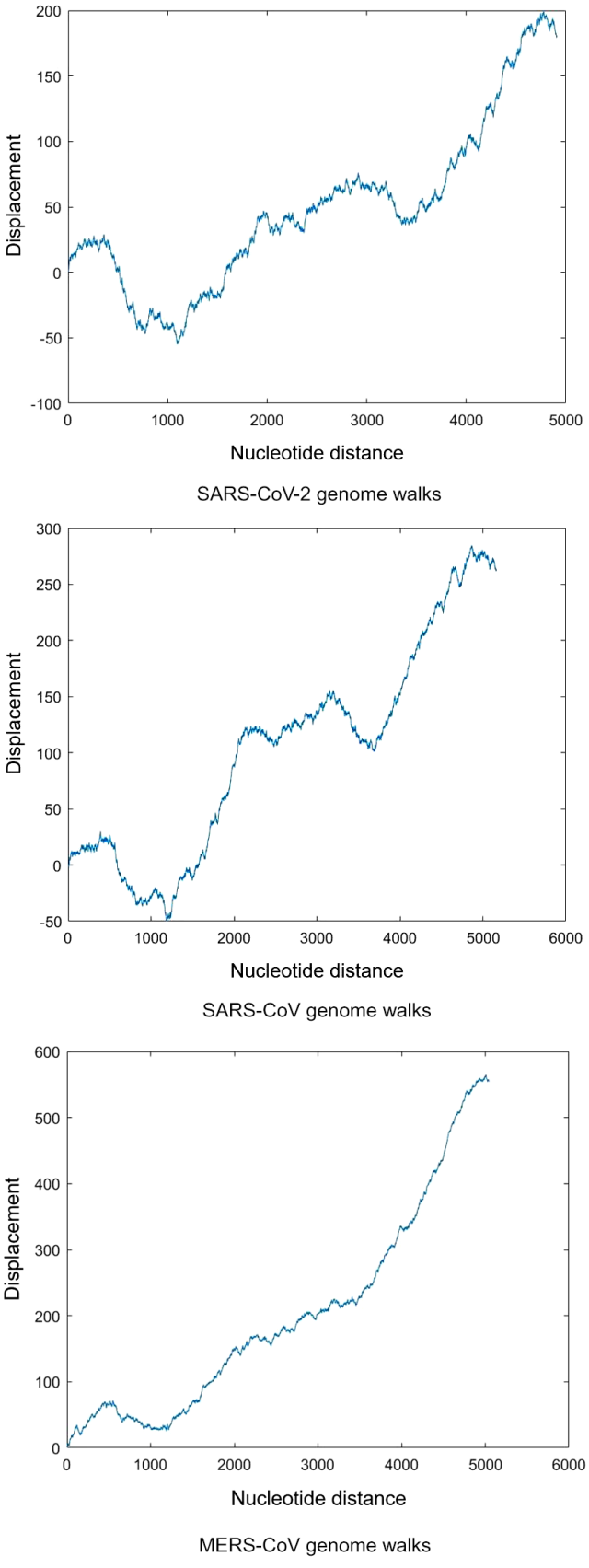

Fig. 3 Complete genome walks for different coronaviruses.

is rejected, and therefore the non-linearity is concluded.

2.1. Database and Analysis

We extracted the coronaviruses genomes from the open-access nucleotide database that is available in Ref. 51. In the case of each type of coronavirus, we selected fifteen samples from this database. These samples represent complete genome sequences. The genome accession number for each used sample in this study is listed in Table 1.

We mapped the genome sequences for each sample in the form of genome walks. Figure 3 illustrates the samples of genome walks for various viruses. As can be seen, all plots show similar patterns. Therefore, complexity theory was used to decode their differences.

We computed the fractal dimension, approximate entropy, sample entropy, and Hurst exponent for different samples of coronaviruses genome walks using a set of codes that have been written in MATLAB, according to the discussed methodology.

We also ran ANOVA tests on the calculated fractal dimension, approximate entropy, sample entropy, and Hurst exponent to check the significance of their changes between different coronaviruses. Besides, the posthoc Tukey test was chosen to perform pairwise comparisons between different types of coronavirus. All statistical analyses were performed using a significance level of 0.01.

Besides, to check the behavior of non-linearities of genome walks, we ran the surrogate analysis. For this purpose, we generated amplitude adjusted phase shuffled surrogate genome walks in MATLAB. To obtain the significance level of $2/(39+1) = 5\%$ for each genome walk, we generated 39 different sets of surrogate data for each set of original data belong to SARS-CoV-2, SARS-CoV, and MERS-CoV.

3. RESULTS

Figure 4 shows the fractal exponents of different coronaviruses' genome walks. The standard deviation of calculated values for fifteen samples of each type of coronavirus is shown using the related error bar.

The genome walk for SARS-CoV-2 has the highest fractal dimension compared to other types of

Fractals 2021.29. Downloaded from www.worldscientific.com
by UNIVERSITI TEKNOLOGI MALAYSIA (UTM) on 03/08/22. Re-use and distribution is strictly not permitted, except for Open Access articles.

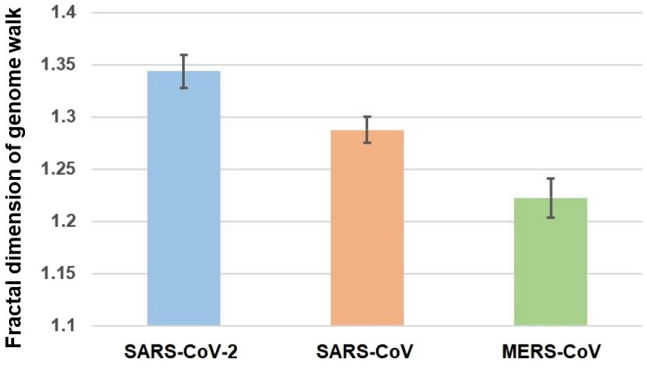

Fig. 4 The fractal exponent of the genome walks.

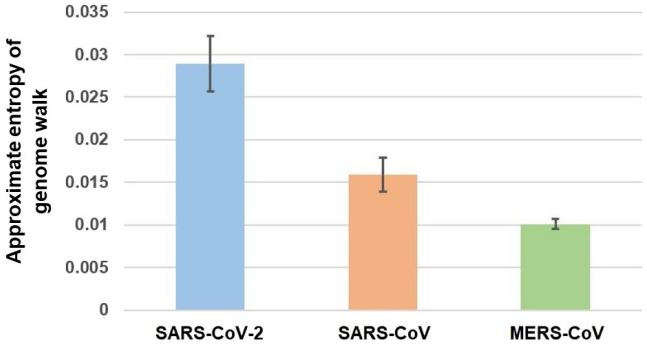

Fig. 5 The approximate entropy of the genome walks.

coronavirus. By moving to SARS-CoV and MERS-CoV, the genome walks' fractal exponent decreases. Therefore, the genome walk of SARS-CoV-2 has a higher complexity than SARS-CoV, and MERS-CoV genome walks. The alterations of the genome walks' fractal exponent among various types of coronavirus reflect their different genomic structures.

The ANOVA test's result (p-value = 0.0000, F-value = 215.8735) on the computed fractal exponents indicates that these viruses have significantly different structures. On the other hand, the Tukey test results in Table 2 demonstrate significant differences in pairwise comparisons of these viruses' genomic structures. In other words, these viruses have significantly different structures.

The results of the calculation of approximate entropy of different genome walks are illustrated in Fig. 5. The error bars represent standard deviations.

As it is clear, the trend of alterations of the genome walk's approximate entropy is similar to the trend of the alterations of the fractal exponent for these genome walks. Since greater approximate entropy stands for the greater complexity, therefore, we can state that the genome walks of SARS-CoV-2 has the highest complexity compared to other types of coronavirus. This result indicates differences between the genomic structures of different types of coronavirus.

The ANOVA test's result (p-value = 0.0000, F-value = 279.1928) on the computed values of the approximate entropy indicates that these viruses have significantly different complex structures. On the other hand, the Tukey test results in Table 3 demonstrate significant differences in pairwise comparisons of these viruses' genomic structures. In other words, these viruses have significantly different structures. Therefore, the results of the approximate entropy validate the fractal analysis results.

Since the lengths of genome walks for various types of coronavirus are different, we verified the results of fractal analysis and approximate entropy by calculating the sample entropy of genome walks for different types of coronavirus. The results of this analysis are presented in Fig. 6. The error bars represent the standard deviations.

As can be seen in Fig. 6, the result of the analysis of sample entropy is similar to the findings of fractal analysis and approximate entropy, where the genome walks for SARS-CoV-2 and MERS-CoV respectively has the largest and smallest sample entropy. In other words, the complexity of

Table 2 Comparison of Genome Walks Based on the Fractal Dimension.

Comparison	*P*-Value
SARS-CoV-2 genome walk versus SARS-CoV genome walk	0.0000
SARS-CoV-2 genome walk versus MERS genome walk	0.0012
SARS-CoV genome walk versus MERS genome walk	0.0000

Table 3 Comparison of Genome Walks Based on the Approximate Entropies.

Comparison	*P*-Value
SARS-CoV-2 genome walk versus SARS-CoV genome walk	0.0000
SARS-CoV-2 genome walk versus MERS genome walk	0.0265
SARS-CoV genome walk versus MERS genome walk	0.0000

Fractals 2021.29. Downloaded from www.worldscientific.com
by UNIVERSITI TEKNOLOGI MALAYSIA (UTM) on 03/08/22. Re-use and distribution is strictly not permitted, except for Open Access articles.

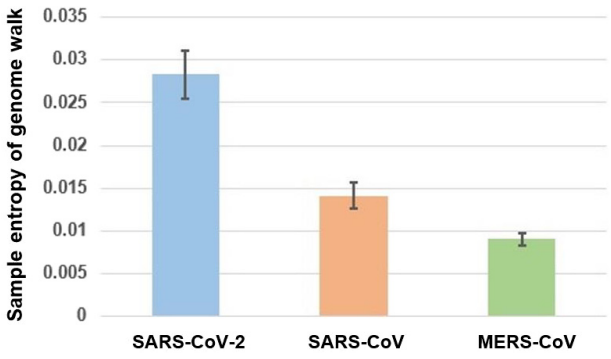

Fig. 6 The sample entropy of the genome walks.

genome walks between SARS-CoV-2, SARS-CoV, and MERS-CoV reduces. Besides, the ANOVA test's result (p-value = 0.0000, F-value = 415.4311) on the computed values of the sample entropy indicates that these viruses have significantly different complex structures. Therefore, the findings of the sample entropy analysis verify the obtained findings in Figs. 4 and 5.

As was mentioned previously, to investigate the genetic mutation, we computed the Hurst exponent of genome walks. These findings are illustrated in Fig. 7.

Based on the results, the genome walks of SARS-CoV-2 has the lowest value of the Hurst exponent compared to other coronaviruses. Since the bigger deviation of the Hurst exponent from 0.5 (completely random process) indicates less irregularity and higher memory in the random walks, therefore, we can state that genome walks of SARS-CoV-2 is more irregular and contains less memory than the genome walks of SARS-CoV, which itself is more irregular and contain less memory than the genome walks of MERS-CoV. Therefore, the variations of the Hurst exponent between different types of coronavirus also prove different structures of these viruses.

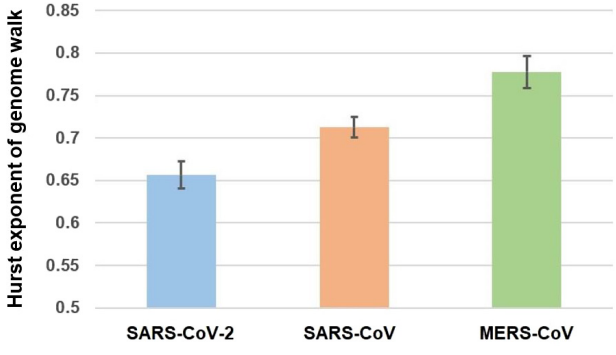

Fig. 7 The Hurst exponent of the genome walks.

Table 4 Comparison of the Hurst Exponent of Genome Walks Among Various Coronaviruses.

Comparison	*P*-Value
SARS-CoV-2 genome walk versus SARS-CoV genome walk	0.0000
SARS-CoV-2 genome walk versus MERS genome walk	0.0012
SARS-CoV genome walk versus MERS genome walk	0.0000

Table 5 Comparison (p-Value) Between Original and Surrogate Data.

	SARS-CoV-2	SARS-CoV	MERS-CoV
Fractal dimension	0.00001	0.00001	0.00001
Approximate entropy	0.00001	0.00001	0.00001
Sample entropy	0.00001	0.00001	0.00001

The ANOVA test's result (p-value = 0.0000, F-value = 215.8735) on the computed values of the Hurst exponent indicates that these viruses have significantly different irregular structures and contain different memories. The results of the Tukey test in Table 4 also indicate significant differences in pairwise comparisons of genomic structures of these viruses.

As was mentioned previously, we checked the nonlinearity of genome walks for various coronaviruses using surrogate analysis. Table 5 indicates significant differences in the fractal dimension, approximate entropy, and sample entropy of original versus surrogate data. Therefore, the null hypothesis is rejected, and we can conclude on the nonlinearity of genome walks for various coronaviruses.

Overall, based on the conducted analysis, we can state that three dangerous types of coronavirus have significantly different structures. Since all these viruses have the same root as the coronavirus, the variations between their structures are due to the virus's genetic mutations.

4. DISCUSSION AND CONCLUSION

In this work, we evaluated the differences among the genomic structures of different coronaviruses.

Fractals 2021.29. Downloaded from www.worldscientific.com
by UNIVERSITI TEKNOLOGI MALAYSIA (UTM) on 03/08/22. Re-use and distribution is strictly not permitted, except for Open Access articles.

We extracted the genome walks of the coronaviruses and then analyzed their complexity using fractal exponent, approximate entropy, and sample entropy. Although mapping of genomes in the form of genome walks showed that three types of coronavirus have similar patterns of genome walks, however, our analysis demonstrated that the SARS-CoV-2 genome walk has a more complex structure than other coronaviruses. The ANOVA test's result demonstrated a significant difference in the structures of different types of coronavirus. Besides, pairwise comparisons also indicated that each type of virus has a significantly different genomic structure than other types of coronavirus. Besides, we also verified the nonlinear structure of genome walks using surrogate analysis, where the results showed the non-linearity of genome walks in the case of different coronaviruses.

In Ref. 31, we showed that the higher complexity of genome walks could be related to the greater danger of the virus. Since, based on the results, SARS-CoV-2 has higher complexity than other coronaviruses, we can state that SARS-CoV-2 is more dangerous than SARS-CoV and MERS-CoV. Therefore, this finding is in line with the reported studies,[52,53] which state SARS-COV-2 is more dangerous than SARS-CoV and MERS-CoV. Here, we should note that we cannot exactly prove that such a coupling may always exist. However, we can state that based on our obtained results, there is an association between the complexity of genome walks for a virus and its danger for humans. This investigation should be further continued in case of other types of viruses.

To investigate the virus mutation, which is caused by mistakes in copying the virus sequence, we analyzed the long-term memory of genome walks for different types of coronavirus using the Hurst exponent. The obtained results demonstrated that the genome walk of SARS-CoV-2 has lower memory than SARS-CoV and MERS-CoV genome walks. The statistical analyses also proved the significant differences in the long-term memory of genome walks among these viruses. Therefore, we can state that the higher mutation of SARS-CoV-2 than SARS-CoV and MERS-CoV[54] can be understood by analyzing the long-term memory of genome walks using the Hurst exponent. In fact, in this study, for the first time, we showed a coupling between the memory of the virus and the memory of its RNA walk. The proposed methodology for the memory of coronavirus and its genome walk has been initiated from our recent study[55] in which we reported a link between the fractal dimension of EEG signals (as the feature of brain activity) and the correctness of answers that subjects gave after watching videos in normal and 3D modes, which reflects their memory. In that study, we found that while watching a 3D video, the fractal dimension of EEG signals is larger compared to watching a video in normal conditions. Considering the link between fractal dimension and Hurst exponent ($H = 2 - F$), the value of Hurst exponent for EEG signals in 3D mode was more deviated from 0.5. Therefore, EEG signals contained higher memory in 3D mode than the normal watching condition. On the other hand, subjects gave more correct answers when they watched a video in 3D mode compared to the normal watching mode that reflects their higher memory. We concluded that the memory of the system (brain) is reflected on the memory of EEG signals (as the indicator of this system). Similarly, in this study, we followed the same strategy in which coronavirus replaces the brain, and RNA walks replace EEG signals. We found that there is a link between the memory of coronavirus and the memory of its RNA walk. Besides, we can also refer to our other study[56] in which we showed that by presenting a stimulus to a subject, the Hurst exponent of his/her EEG signals increases. As it is known, each stimulus that we receive increases our memory, and therefore, the observed increment in the Hurst exponent (memory) of EEG signals can be a good indicator of memory of the brain. Here, we should note that our work in this paper can be examined in the case of other systems to study the link between the memory of a system and the memory of its random walk.

Therefore, we can conclude that analyzing the complexity and memory content of genome walks is a robust tool to not only decode the differences between different types of coronavirus but also to explain the virus mutation.

ACKNOWLEDGMENTS

This work was supported in part by the Project (2021/2204), Grant Agency of Excellence, the University of Hradec Kralove, the Faculty of Informatics and Management, Czech Republic.

REFERENCES

1. S. Azekawa *et al.*, Co-infection with SARS-CoV-2 and influenza A virus, *IDCases* **20** (2020) e00775.

Fractals 2021.29. Downloaded from www.worldscientific.com
by UNIVERSITI TEKNOLOGI MALAYSIA (UTM) on 03/08/22. Re-use and distribution is strictly not permitted, except for Open Access articles.

2. C. Jiang *et al.*, Comparative review of respiratory diseases caused by coronaviruses and influenza A viruses during epidemic season, *Microbes Infect.* **22**(6) (2020) 236–244.
3. J. Hu *et al.*, Evolution and variation of the SARS-CoV genome, *Genom. Proteom. Bioinf.* **1**(3) (2016) 216–225.
4. G. M. Pavlović-Lažetić *et al.*, SARS-CoV genome polymorphism: A bioinformatics study, *Genom. Proteom. Bioinf.* **3**(1) (2005) 18–35.
5. Y. Ma *et al.*, Structural basis and functional analysis of the SARS coronavirus nsp14-nsp10 complex, *Proc. Natl. Acad. Sci. USA* **112**(30) (2015) 9436–9441.
6. T. Briese *et al.*, Middle East respiratory syndrome coronavirus quasispecies that include homologues of human isolates revealed through whole-genome analysis and virus cultured from dromedary camels in Saudi Arabia, *mBio* **5**(3) (2014) e01146-14.
7. Z. Zhang, L. Shen and X. Gu, Evolutionary dynamics of MERS-CoV: Potential recombination, positive selection and transmission, *Sci. Rep.* **6** (2016) 25049.
8. Q. Xie *et al.*, Genomic sequencing and analysis of the first imported Middle East Respiratory Syndrome Coronavirus (MERS CoV) in China, *Sci. China Life Sci.* **58**(8) (2015) 818–820.
9. O. M. Ugurel, O. Ata and D. Turgut-Balik, An updated analysis of variations in SARS-CoV-2 genome, *Turk. J. Biol.* **44**(3) (2020) 157–167.
10. L. Ferretti *et al.*, Quantifying SARS-CoV-2 transmission suggests epidemic control with digital contact tracing, *Science* **368**(6491) (2020) eabb6936.
11. P. Stefanelli *et al.*, Whole genome and phylogenetic analysis of two SARS-CoV-2 strains isolated in Italy in January and February 2020: Additional clues on multiple introductions and further circulation in Europe, *Euro Surveill.* **25**(13) (2020) 2000305.
12. X. Ou *et al.*, Characterization of spike glycoprotein of SARS-CoV-2 on virus entry and its immune cross-reactivity with SARS-CoV, *Nat. Commun.* **11** (2020) 1620.
13. M. Kandeel *et al.*, From SARS and MERS CoVs to SARS-CoV-2: Moving toward more biased codon usage in viral structural and nonstructural genes, *J. Med. Virol.* **92**(6) (2020) 660–666.
14. A. Wu *et al.*, Genome composition and divergence of the novel coronavirus (2019-nCoV) originating in China, *Cell Host Microbe.* **27**(3) (2020) 325–328.
15. Y. Wan *et al.*, Receptor recognition by the novel coronavirus from Wuhan: An analysis based on decade-long structural studies of SARS coronavirus, *J. Virol.* **94** (2020) e00127-20.
16. J. Chon *et al.*, Genomic characterization of the 2019 novel human-pathogenic coronavirus isolated from a patient with atypical pneumonia after visiting Wuhan, *Emerg. Microbes Infect.* **9**(1) (2020) 221–236.
17. L. R. Lopes, G. de Mattos Cardillo and P. B. M. Paiva, Molecular evolution and phylogenetic analysis of SARS-CoV-2 and hosts ACE2 protein suggest Malayan pangolin as intermediary host, *Braz. J. Microbiol.*, doi:10.1007/s42770-020-00321-1 (2020).
18. H. Namazi, E. Aghasian and T. Seifi Ala, Complexity-based classification of EEG signal in normal subjects and patients with epilepsy, *Technol. Health Care.* **28**(1) (2020) 57–66.
19. H. Namazi and O. Krejcar, Analysis of pregnancy development by complexity and information-based analysis of fetal phonocardiogram (PCG) signals, *Fluct. Noise Lett.* (2021), doi:10.1142/S0219477521500280.
20. M. Soundirarajan *et al.*, Analysis of brain-facial muscle connection in the static fractal visual stimulation, *Int. J. Imaging Syst. Technol.* (2020) 1–7, doi:10.1002/ima.22480.
21. S. M. Kamal, M. H. Babini, O. Krejcar and H. Namazi, Complexity-based decoding of the coupling among heart rate variability (HRV) and walking path, *Front. Physiol.* **11** (2020) 602027.
22. S. Omam *et al.*, Complexity-based decoding of brain-skin relation in response to olfactory stimuli, *Comput. Methods Programs Biomed.* **184** (2020) 105293.
23. M. A. Elblbesy and M. Attia, Optimization of fractal dimension and shape analysis as discriminators of erythrocyte abnormalities: A new approach to a reproducible diagnostic tool, *MBE* **17**(5) (2020) 4706–4717.
24. H. Namazi *et al.*, The fractal based analysis of human face and DNA variations during aging, *Biosci. Trends.* **10**(6) (2017) 477–481.
25. H. Namazi and V. V. Kulish, Complexity-based classification of the coronavirus disease (COVID-19), *Fractals* **28**(5) (2020) 2050114.
26. M. Babič, J. Mihelič and M. Calì, Complex network characterization using graph theory and fractal geometry: The case study of lung cancer DNA sequences, *Appl. Sci.* **10** (2020) 3037.
27. C. Cattani, Fractals and hidden symmetries in DNA, *Mathematical Problems in Engineering* **2010** (2010) 507056, pp. 1–31.
28. H. Namazi and M. Kiminezhadmalaie, Diagnosis of lung cancer by fractal analysis of damaged DNA, *Computational and Mathematical Methods in Medicine* **2015** (2015), p. 242695.
29. C. Cattani and G. Pierro, On the fractal geometry of DNA by the binary image analysis, *Bull. Math. Biol.* **75**(9) (2013) 1544–1570.
30. A. Carbone *et al.*, 3D fractal DNA assembly from coding, geometry and protection, *Natural Comput.* **3** (2004) 235–252.

Fractals 2021.29. Downloaded from www.worldscientific.com
by UNIVERSITI TEKNOLOGI MALAYSIA (UTM) on 03/08/22. Re-use and distribution is strictly not permitted, except for Open Access articles.

31. H. Namazi, Complexity-based classification of the coronavirus genome versus genomes of the human immunodeficiency virus (HIV) and dengue virus, *Fractals* (2020), doi:10.1142/S0218348X20501297.
32. H. Namazi and N. B. M. Dawi, Information and complexity-based analysis of the variations of the coronavirus genome between different countries, *Fractals* (2020), doi:10.1142/S0218348X20501340.
33. H. Namazi, O. Krejcar and A. Subasi, Complexity and information-based analysis of the variations of the SARS-CoV-2 genome in the United States of America (USA), *Fractals* **28**(7) (2020) 2150023.
34. D. C. Nascimento *et al.*, Entropy analysis of high-definition transcranial electric stimulation effects on EEG dynamics, *Brain Sci.* **9**(8) (2019) 208.
35. L. C. Giannasi *et al.*, Intra-day and inter-day reliability of measurements of the electromyographic signal on masseter and temporal muscles in patients with down syndrome, *Scientific Rep.* **10** (2020) 7477.
36. J. Li, D. Liu, J. Zhu and T. Liu, Application of approximate entropy in evaluation of motion state based on wearable devices, 18*th Int. Symp. Distributed Computing and Applications for Business Engineering and Science* (DCABES), Wuhan, China, 2019, pp. 211–213, doi:10.1109/DCABES48411.2019.00059.
37. K. Harezlak and P. Kasprowski, Application of time-scale decomposition of entropy for eye movement analysis, *Entropy* **22** (2020) 168.
38. X. Zhang, X. A. Zhou and Y. H. Yu, Similarity analysis of DNA using improved approximate entropy, *Int. Conf. Biomedical Engineering and Biotechnology*, Macao, 2012, pp. 511–514, doi:10.1109/iCBEB.2012.353.
39. S. K. Botting, J. P. Trzeciakowski, M. F. Benoit *et al.*, Sample entropy analysis of cervical neoplasia gene-expression signatures, *BMC Bioinf.* **10** (2009) 66.
40. A. P. Singh, S. Mishra and S. Jabin, Sequence based prediction of enhancer regions from DNA random walk, *Sci. Rep.* **8** (2018) 15912.
41. C. K. Peng *et al.*, Long-range correlations in nucleotide sequences, *Nature* **356**(6365) (1992) 168–170.
42. M. R. A. Ahamed, H. Babini and H. Namazi, Complexity-based analysis of brains' synchronization in human-human interaction, *Fractals*, doi:10.1142/S0218348X20501029.
43. L. Mesin, Estimation of complexity of sampled biomedical continuous time signals using approximate entropy, *Front. Physiol.* **9** (2018) 710.
44. S. M. Pincus, Approximate entropy as a measure of system complexity, *Proc. Natl. Acad. Sci. USA* **88**(6) (1991) 2297–2301.
45. K. K. Ho *et al.*, Predicting survival in heart failure case and control subjects by use of fully automated methods for deriving nonlinear and conventional indices of heart rate dynamics, *Circulation* **96**(3) (1997) 842–848.
46. A. Delgado-Bonal and A. Marshak, Approximate entropy and sample entropy: A comprehensive tutorial, *Entropy* **21**(6) (2019) 541.
47. H. Namazi, E. Herrera-Viedma and O. Krejcar, Complexity-based detection of similarity between animal coronaviruses and SARS-CoV-2 in humans, *Fractals* **28**(7) (2020) 2150031.
48. L. Loewe, Genetic mutation, *Nature Education* **1**(1) (2008) 113.
49. B. L. Smith and C. O. Wilke, Virus evolution: A new twist in measuring mutation rates, *eLife Sci.*, doi:10.7554/eLife.29586 (2017).
50. J. Chen, R. Wang, M. Wang and G. W. Wei, Mutations strengthened SARS-CoV-2 infectivity, *J. Mol. Biol.* **432**(19) (2020) 5212–5226.
51. Nucleotide database, https://www.ncbi.nlm.nih.gov/nuccore/.
52. B. Vellingiri *et al.*, COVID-19: A promising cure for the global panic, *Sci. Total Environ.* **725** (2020) 138277.
53. C. Yang *et al.*, Coronavirus disease 2019: Reassembly attack of coronavirus, *Int. J. Environ. Health Res.* (2020) 1–9, doi:10.1080/09603123.2020.1747602.
54. C. Wang *et al.*, The establishment of reference sequence for SARS-CoV-2 and variation analysis, *J. Med. Virol.* **92**(6) (2020) 667–674.
55. M. H. Babini, V. V. Kulish and H. Namazi, Physiological state and learning ability of students in normal and virtual reality conditions: Complexity-based analysis, *J. Med. Internet Res.* **22**(6) (2020) e17945.
56. H. Namazi and V. V. Kulish, Fractional diffusion based modelling and prediction of human brain response to external stimuli, *Comput. Math. Methods Med.* **2015** (2015) 148534, doi:10.1155/2015/148534.