doi:10.20944/preprints202004.0486.v1

A quantitative comparisons of β-coronavirus genomes and their associated genes

Sk. Sarif Hassan[a,*], Ranjeet Kumar Rout[b]

[a] *Deptartment of Mathematics, Pingla Thana Mahavidyalaya, Paschim Medinipur-721140, India*
[b] *Department of Computer Science & Engineering, National Institute of Technology, Hazratbal, Srinagar, India*

Abstract

A precise understanding of the genes and associated genomes of SARS-CoV2 is important for various reasons such as discovering origin of the virus and virulence and so on. A thorough descriptive understanding of the SARS-CoV2 genomes and other coronavirus of the beta-coronavirus genus is primarily important. In this article, a set of ten genomes of four CoVs and their associated genes are considered for this present study. A spatial representations of nucleotide bases including purine-pyrimidine representations of the different genes of the corresponding genomes are quantified using Hurst exponent, Shannon entropy and density estimation of different nucleotides including GC content, in order to draw a comparison and contrast among the ten genomes of different types of CoVs which include MERS, SARS-CoV, HKU1 (Human Coronavirus) and associated their genes.

Keywords: COVID-19; Hurst exponent; Shannon entropy; Nucleotide Density; Purine-Pyrimidine representation.

---

[*] Corresponding author

Email addresses: `sarimif@gmail.com` (Sk. Sarif Hassan), `ranjeetkumarrout@nitsri.net` (Ranjeet Kumar Rout)

Preprint submitted to Elsevier. *April 27, 2020*

1. Introduction

The coronavirus disease (COVID-19) is a highly transmittable viral infection caused by severe acute respiratory syndrome coronavirus 2 (SARS-CoV2), which is primarily emerged in Wuhan, China and accordingly it spreads over various countries and world has been experiencing the pandemic [1, 2, 3, 4, 5, 3]. The family of coronaviruses (CoVs) can be grouped into four different types such as α, β, γ and δ coronavirus [6, 7]. The SARS-CoV2 belongs to the β-CoV group [8]. The intermediate source of origin and transfer to humans is still unsettled although various claims have been published. However, the rapid human to human transfer has been confirmed widely [9]. It has been observed that the transmission rate of SARS-CoV2 is higher than SRAS-CoV and the reason could be genetic recombination event at S gene/protein in the RBD region of SARS-CoV2 may have enhanced its transmission ability [10]. It is reported earlier that the genome of the SARS-CoV2 has been reported over 76% identical to the previous SARS-CoV [11]. From the sequential similarity of SARS-CoV and SARS-CoV2 genomes, it has been seen that SARS-CoV-2 lies close to the group of SARS-CoVs in the evolutionary tree (phylogeny) [11]. It is also reported recently that notable variations in SARS-CoV and SARS-CoV-2 such as the absence of 8a protein and variations in the number of amino acids in 8b and 3c primary protein sequences in SARS-CoV2 [12].

The genome of coronaviruses contains approximately 26kb and 32kb nucleotides which represent a variable number (from 6 to 11) of open reading frames (ORFs) [13, 14, 15]. The ORF1 represents (67% of the genome) 16 non-structural proteins (nsps) and other ORFs encode accessory proteins and structural proteins [16]. A comprehensive comparison of genomes of coronaviruses must comprise all the ORFs (protein encoding frames). It has been observed by many researchers that the present SARS-CoV2 genomes have string similarities with other CoVs of the β-CoV group such as middle east respiratory syndrome coronavirus (MERS-CoV) and human CoV (HKU1) [17, 18]. It is noticed that the only sequence based similarity does not illuminate actual difference (microscopic) among these genes and genomes. It only presents a qualitative understanding of the evolutionary connections. Therefore, we intend to discover the spatial arrangements of each nucleotide over the RNA genes and genomes using various mathematical methods. In this present study, a comprehensive comparison among the whole genomes and associated genes of SARS-CoV2, MERS, SARS-CoV-1, HKU1 (Human Coronavirus) is made, from the spatial and molecular organizations including different density estimations of nucleotides.

1.1. Dataset and Its Specifications

From the NCBI Virus Database ($https://www.ncbi.nlm.nih.gov/labs/virus/vssi/$) we took ten genomes and their associated genes of which the detailed list is given in the Table-1.

Table 1: List of genomes and their associated information

Name	**Accession**	**Species**	**Length**	**Geo_Location**
S1	NC_045512	SARS	29903	China
S2	MT012098	SARS	29854	India
S3	NC_038294	MERS	30111	UK
S4	NC_006577	Human coronavirus HKU1	29926	Unknown
S5	KP143510	α-CoV-1	28967	UK
S6	KF923891	β-CoV-1	30713	China
S7	MT126808	SARS	29876	Brazil
S8	MT077125	SARS	29785	Italy
S9	MT292571	SARS	29782	Spain
S10	MT322402	SARS	29898	USA

Let $G(i,j)$ denotes the j^{th} gene of the i^{th} genome Si as mentioned in the Table-1.

Our objective is to consider the above data is to determine spatial arrangement of nucleotide bases (A, T, C and G) and in the level of purine (A,G) and pyrimidine (C, T) bases of the genes and their associated whole genomes. We took here six SARS-CoV2 genomes and associated genes from six countries China, India, Brazil, Italy, Spain and USA. Another set of four genomes and their associated genes of MERS, HKU1, α-CoV-1 and β-CoV-1 types are considered in order to find similarities and dissimilarities.

We transform each genome Si and their corresponding genes $G(i,j)$ in different binary forms as demonstrated in the following.

Firstly, we transform each sequence to a binary sequence of $1's$ and $0's$ as per the definition 2. Here purine (A,G) and pyrimidine (T,C) bases are represented as "1" and "0" respectively. This binary representation is named as purine-pyrimidine representation [19, 20, 21].

$$\begin{aligned} A/G &\rightarrow 1 \\ T/C &\rightarrow 0 \end{aligned} \tag{1}$$

Also we transform each sequence to a binary sequence with respect to a nucleotide base B of $1's$ and $0's$ as per the following definition **??**.

$$\begin{aligned} X &\rightarrow 1 \quad if\ X = B \\ X &\rightarrow 0 \quad if\ X \neq B \end{aligned} \tag{2}$$

Hence four binary representations for each nucleotide $B \in \{A, T, C, G\}$ would be obtained for a given nucleotide sequence. These binary representations are actually the spatial template of each nucleotides. Each of these spatial templates are to be analysed using various means as mentioned in the following section.

doi:10.20944/preprints202004.0486.v1

2. Methods

Here some popular spatial features such as the amount of uncertainty, conservation entropy and auto-correlation of the spatial arrangements (binary representations) for every gene and genome, have been found using Shannon entropy and Hurst exponent which are described in brief in the following subsections.

2.1. Shannon entropy

There are two kinds of Shannon entropy measures are considersin the present study as follows.

- **Binary Shannon Entropy**: The Shannon entropy (SE) measures information entropy of a Bernoulli process with probability p of the two outcomes (0/1) [22, 23]. It is defined as

  $$SE = -\sum_{i=1}^{2} p_i log_2(p_i)$$

  where $p_1 = \frac{k}{2^l}$ and $p_2 = \frac{l-k}{2^l}$; here l is the length of the binary sequence and k is the number of 1's in the binary sequence of length l [**?**]. The binary Shannon entropy is a measure of the uncertainty in a binary sequence. Whenever the probability $p = 0$, the event is certain never to occur, and so there is no uncertainty, leading to an entropy of 0. Similarly, if the probability $p = 1$, the result is certain, so the entropy must be 0. When $p = 0.5$, the uncertainty is at a maximum and consequently the SE is 1.

- **Nucleotide Conservation Shannon Entropy (Con_S):** Conservation of each of the four nucleotides has been determined using Shannon entropy [22]. For a given nucleotide sequence, the conservation SE is calculated as follows:

  $$SE = -\sum_{i=1}^{4} p_{N_i} log_2(p_{N_i})$$

doi:10.20944/preprints202004.0486.v1

where $p_{N_i} = \frac{k}{m}$ where 'k' represents the occurrence frequency of a nucleotide N_i in a RNA sequence of length 'm'.

2.2. Hurst Exponent

The Hurst Exponent (HE) is used for time series analysis to interpret the trend which could be positive or negative [24, 25]. The HE lies in between 0 to 1. The HE value $0 < HE < 0.5$ and $0.5 < HE < 1$ designate negative and positive auto-correlation (trend) of a time series/binary sequence respectively and $HE = 0.5$ denotes a absolute randomness of a time series/binary sequence.

The HE of a binary sequence s_n is defined as

$$\left(\frac{n}{2}\right)^{HE} = \frac{X(n)}{Y(n)} \tag{3}$$

where

$$Y(n) = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(s_i - m)}$$

and $X(n) = maxT(i,n) - minT(i,n)$, where

$$T(i) = \sum_{j=1}^{n}(s_i - t)$$

and

$$t = \sqrt{\frac{1}{n}\sum_{i=1}^{n}s_i}$$

The auto-correlation/trend of the binary representations as mentioned for all the ten genomes Si and their corresponding genes is obtained through the Hurst exponent.

In addition to these two spatial, length independent features, some basic derivative features such as GC content, purine-pyrimidine density etc. are obtained by using the percentage of density of each nucleotide in a given nucleotide sequence [26, 27, 28, 29, 30, 31, 32, 33].

3. Results and Illustrations

In the following subsections firstly genes of all types of CoVs are classified based on their length and accordingly based their spatial features genes are genomes are compared.

3.1. Length based Gene-Classification

Before proceeding further to investigate the spatial features of the genes and associated genomes, we present the list of genes with their respective length and location in the respective genome a given in Table-2. This information would classify the genes over the genomes based on their length.

Table 2: Genes across all the ten genomes with their location and respective length

Gene	**Location**	**Length**	Gene	**Location**	**Length**
G(1,11)	location: 29558..29674	117	**G(7,5)**	location: 26523..27191	669
G(2,10)	location: 29542..29658	117	**G(8,5)**	location: 26467..27135	669
G(7,10)	location: 29558..29674	117	**G(9,5)**	location: 26469..27137	669
G(8,11)	location: 29502..29618	117	**G(10,5)**	location: 26518..27186	669
G(9,11)	location: 29504..29620	117	**G(4,6)**	location: 27633..28304	672
G(10,11)	location: 29553..29669	117	**G(3,6)**	location: 26839..27513	675
G(1,8)	location: 27756..27887	132	**G(5,5)**	location: 24954..25667	714
G(8,8)	location: 27700..27831	132	**G(3,5)**	location: 26092..26832	741
G(9,8)	location: 27702..27833	132	**G(5,7)**	location: 25913..26707	795
G(10,8)	location: 27751..27882	132	**G(1,3)**	location: 25393..26220	828
G(1,6)	location: 27202..27387	186	**G(2,3)**	location: 25377..26204	828
G(2,6)	location: 27186..27371	186	**G(7,3)**	location: 25393..26220	828
G(7,6)	location: 27202..27387	186	**G(8,3)**	location: 25337..26164	828
G(8,6)	location: 27146..27331	186	**G(9,3)**	location: 25339..26166	828
G(9,6)	location: 27148..27333	186	**G(10,3)**	location: 25388..26215	828
G(10,6)	location: 27197..27382	186	**G(6,2)**	location: 21506..22342	837
G(5,3)	location: 24576..24788	213	**G(5,8)**	location: 26720..27853	1134
G(5,4)	location: 24736..24957	222	**G(4,2)**	location: 21773..22933	1161
G(1,4)	location: 26245..26472	228	**G(3,9)**	location: 28565..29800	1236
G(2,4)	location: 26229..26456	228	**G(1,10)**	location: 28274..29533	1260
G(7,4)	location: 26245..26472	228	**G(2,9)**	location: 28258..29517	1260
G(8,4)	location: 26189..26416	228	**G(7,9)**	location: 28274..29533	1260
G(9,4)	location: 26191..26418	228	**G(8,10)**	location: 28218..29477	1260
G(10,4)	location: 26240..26467	228	**G(9,10)**	location: 28220..29479	1260
G(3,7)	location: 27589..27837	249	**G(10,10)**	location: 28269..29528	1260
G(4,5)	location: 27373..27621	249	**G(6,3)**	location: 22354..23625	1272
G(5,6)	location: 25654..25902	249	**G(4,7)**	location: 28320..29645	1326
G(5,9)	location: 27858..28163	306	**G(6,6)**	location: 29079..30425	1347
G(3,3)	location: 25531..25842	312	**G(2,2)**	location: 21550..25368	3819
G(3,4)	location: 25851..26180	330	**G(1,2)**	location: 21563..25384	3822
G(4,4)	location: 27051..27380	330	**G(7,2)**	location: 21563..25384	3822
G(6,5)	location: 27792..28121	330	**G(8,2)**	location: 21507..25328	3822
G(1,7)	location: 27394..27759	366	**G(9,2)**	location: 21509..25330	3822
G(1,9)	location: 27894..28259	366	**G(10,2)**	location: 21558..25379	3822
G(2,7)	location: 27378..27743	366	**G(3,2)**	location: 21455..25516	4062
G(2,8)	location: 27878..28243	366	**G(4,3)**	location: 22942..27012	4071
G(7,7)	location: 27394..27759	366	**G(6,4)**	location: 23640..27716	4077
G(7,8)	location: 27894..28259	366	**G(5,2)**	location: 20164..24564	4401
G(8,7)	location: 27338..27703	366	**G(5,1)**	location: 114..20167	20054
G(8,9)	location: 27838..28203	366	**G(3,1)**	location: 278..21513	21236
G(9,7)	location: 27340..27705	366	**G(6,1)**	location: 210..21496	21287
G(9,9)	location: 27840..28205	366	**G(1,1)**	location: 266..21555	21290
G(10,7)	location: 27389..27754	366	**G(2,1)**	location: 253..21542	21290
G(10,9)	location: 27889..28254	366	**G(7,1)**	location: 266..21555	21290
G(4,8)	location: 28342..28959	618	**G(8,1)**	location: 210..21499	21290
G(5,10)	location: 28168..28788	621	**G(9,1)**	location: 212..21501	21290
G(3,8)	location: 27852..28511	660	**G(10,1)**	location: 261..21550	21290
G(1,5)	location: 26523..27191	669	**G(4,1)**	location: 206..21753	21548
G(2,5)	location: 26507..27175	669			

The observations and immediate consequences can be drawn as follows from Table-2.

- The SARS-CoV2 genomes S1, S8, S9 and S10 contain eleven genes whereas the genomes S2 and S10 contain ten genes. It has been reported that the genomes S2 (from India) and S1 (from China) is similar with approximately 99.98% [**?**].
- The length of each of the genes G(1,11), G(2,10), G(7,10), G(8,11), G(9,11) and G(10,11) is 117. Note that the last most gene from each SARS-CoV2 genome is the smallest gene of length 117 among all the genes present

across the ten genomes. The loci of the genes is contained within the frame (29502..29674). It is worth noting that the genes of length 117 present in the genome S1 (from China) and S7 (from Brazil) at the fixed location: 29558..29674 though the length of the genomes (length of S1: 29903 and the length of the genome S7: 29876) are different.

- The genes G(i,6) of length 186 are present over all the SARS-CoV2 genomes Si where i=1,2,7,8,9 & 10. The locus of each genes G(i,6) presented in all the six SARS-CoV2 genomes contained in the frame 27146..27387. It is noted that the location of the genes G(1,6) and G(7,6) is at a fixed location: 27202..27387.
- The genes G(i,4) of length 228 are present over all the SARS-CoV2 genomes Si where i=1,2,7,8,9 & 10. The locus of each genes G(i,6) presented in all the six SARS-CoV2 genomes contained in the frame 26189..26472. It is noted that the location of the genes G(1,4) and G(7,4) are at a fixed location: 26245..26472.
- There is a pair of genes of length 366 present all over the six SARS-CoV2 genomes. The genes are G(1,7) [location=27394..27759], G(1,9) [location=27894..28259], G(2,7) [location=27378..27743], G(2,8) [location=27878..28243], G(7,7) [location=27394..27759], G(7,8) [location=27894..28259], G(8,7) [location=27338..27703], G(8,9) [location=27838..28203], G(9,7) [location=27340..27705], G(9,9) [location=27840..28205], G(10,7) [location=27389..27754] and G(10,9) [location=27889..28254]. Note that the pair of genes of length 366 present at a fixed gap of length 135 across all the six SARS-CoV2 genomes. As previously pointed out, the loci of the pair of genes presented over the genome S1 and S7 are exactly same. It is worth noting that the pair of genes of length 366 present consecutively over the genomes S2 and S7 (containing ten genes) whereas in others four genomes (containing 11 genes)

contain such pair of genes with a gap of another genes G(1,8), G(8,8), G(9,8) and G(10,8) of length 132 though surprisingly the gap of length 135 is strictly maintained across all the six SARS-CoV2 genomes.

- The genes G(i,5) of length 669 are present over all the SARS-CoV2 genomes Si where i=1,2,7,8,9 & 10. The locus of each genes G(i,5) presented in all the six CoV genomes contained in the frame 26467..27191. It is noted that the location of the genes G(1,5) and G(7,5) is at a fixed location: 26523..27191.
- The genes G(i,3) of length 828 are present over all the SARS-CoV2 genomes Si where i=1,2,7,8,9 & 10. The locus of each genes G(i,3) presented in all the six SARS-CoV2 genomes contained in the frame 25337..26220. It is noted that the location of the genes G(1,3) and G(7,3) is at a fixed location: 25393..26220.
- The genes G(1,10), G(2,9), G(7,9), G(8,10), G(9,10) and G(10,10) of length 1260 are present over all the six CoV genomes. The locus of each of these genes is contained in the frame 28218..29533. It is noted that the location of the genes G(1,5) and G(7,5) is at a fixed location: 28274..29533.
- The genes G(1,2), G(7,2), G(8,2), G(9,2) and G(10,2) of length 3822 are present over all the five SARS-CoV2 genomes. The locus of each of these genes is contained in the frame 21507..25384. It is noted that the location of the genes G(1,2) and G(7,2) is at a fixed location: 21563..25384. It is noted that there is no gene of length of length 3822 in the genome S2.
- The largest lengthy genes G(1,1), G(2,1), G(7,1), G(8,1), G(9,1) and G(10,1) of length 21290 are present over all the six SARS-CoV2 genomes. The locus of each of these genes is contained in the frame 210..21555. It

is noted that the location of the genes G(1,1) and G(7,1) is at a fixed location: 266..21555.

- There are three genes G(3,7), G(4,5) and G(5,6) of length 249 present among three CoV genomes S3, S4 and S5 respectively. There are other three genes G(3,4), G(4,4) and G(6,5) of length 330 present over the genomes S3, S4 and S6.

3.2. Compare and Contrast among Genes and Genomes of SARS-CoV2

Based on the spatial features (Table-4 and Table-5) such as SE and HE of different data type of the genes and associated genomes of six SARS-CoV2, some salient findings are adumbrated as follows:

- It is observed that the six genomes S1, S2, S7, S8, S9 and S10 do follow nearly same nucleotide densities including spatial distributions in terms of positive auto-correlation (HE) as well as the certainty of presence (SE) of the purine-pyrimidine bases. Also the spatial distributions of four nucleotide bases over the genomes S2 and S8 are exactly same. The occurrence of purine and pyrimidine bases over these six genomes are equally likely and consequently, the SE (0.98) is very close to 1 implying the uncertainty of occurrence is at maximum.
- The minimum and maximum value attained by each parameters is given in the Table 3. From the Table 3, it is observed that the genes G(8,8), G(9,8) and G(10,8) (ORF7b) is pyrimidine-rich sequence with 63.64%. It is noted that the gene M of the genome S1 is also pyrimidine-rich with 63.64% whereas the ORF7b contains very less amount of pyrimidine bases with percentage 53.66%. The density of A, T, C and G are all same over all four aforementioned ORF7b gene including M of genome S1. Accordingly the other parameters are remained same. So the gene M of the genome

S1 does follow the same spatial configurations as the ORF7b do in the genomes S8, S9 and S10.

- It is observed that the gene E in the genome S1 is purine-rich (53.97%) whereas other the gene E over the others five genomes S2, S7, S8, S9 and S10 are pyrimidine rich (60.09). It is noted that the spatial distribution of the gene E of S1 is identical with that of gene N of other five genomes S2, S7, S8, S9 and S10. The gene N of the five aforesaid genomes including gene E G(1,4) contain highest GC content (47.222%).

- The spatial distributions as well as nucleotide-densities over the gene ORF1ab (longest gene among all genes present in SARS-CoV2 genomes) over the six genomes do same. It is worth mentioning that the occurrence of presence of purine bases over ORF1ab is absolutely random/uncertain consequently the SE is turned out to be 1. The GC content of all these gene E in six genomes is 37.45%.

- All the spike gene S are pyrimidine-rich (above 52.1%) over the six genomes S1, S2, S7, S8, S9 and S10. The GC content of the spike gene of S1 is 35.79% whereas that in the spike gene of other five genomes is fixed at 37.3%.

- The gene M over the five genomes S2, S7, S8, S9 and S10 do follows same spatial distribution as well as nucleotide densities. This gene M are all pyrimidine rich sequence especially the gene M over the genome S1 highly pyrimidine-rich with 63.64% which coincides with that of ORF7b of the genomes S8, S9 and S10.

- The density of purine over the gene N across the five genomes S2, S7, S8, S9 and S10 is rich with 53.97% where as the gene N of the genome S1 is pyrimidine-rich with 54.35%. The GC content of the gene N of the five aforesaid genomes is centred at 47.2% whereas the GC content of gene N of the genome S1 is 39.49.

- The gene ORF3a over the six SARS-CoV2 genomes is pyrimidine-rich with 54%. The GC content of the gene in the genome S1 is 34% whereas GC content of the gene ORF3a in other five genomes is centred at 54.47%.
- The amount of purine and pyrimidine in the gene ORF6 is almost same (50%) over the five genomes except S1. The gene G(1, 6) ORF6 is pyrimidine-rich sequence. The GC content of the gene G(1,6) is 38.25% whereas that of the gene ORF6 in other five genomes is fixed at 27.956%.
- The gene G(2,7), G(7,7), G(8,7), G(9,7) and G(10,7) contains more pyrimidine bases (53.83%) than purine bases. But in the case of the gene G(1,7)-ORF7a is purine-rich with 50.54%. The GC content of the gene ORF7a over the five genomes is 38.252% but GC content of the gene G(1,7) is 27. 96%.
- The genes G(1,5)-M, G(8,8)-ORF7b, G(9,8)-ORF7b, G(10,8)-ORF7b follow exactly same density of nucleotides and pyrimidine-rich. The purine-pyrimidine representations of all these four genes are positively trending with HE exactly 0.72. It is worth noting that the representations of four nucleotide bases are exactly same.

By the above observations made and corresponding numerical features as presented in the Table 4, it is evident that the density of four nucleotide bases as well as the purine-pyrimidine bases spatial distributions are significantly different from that of the genes over the five genomes of SARS-CoV2.

The feature vectors of each gene and genome are of length 19. Based on euclidean distance among the SARS-CoV2 genomes and their associated genes are calculated and based on the distance matrix a phylogenetic tree is made which is presented in the Fig. 1.

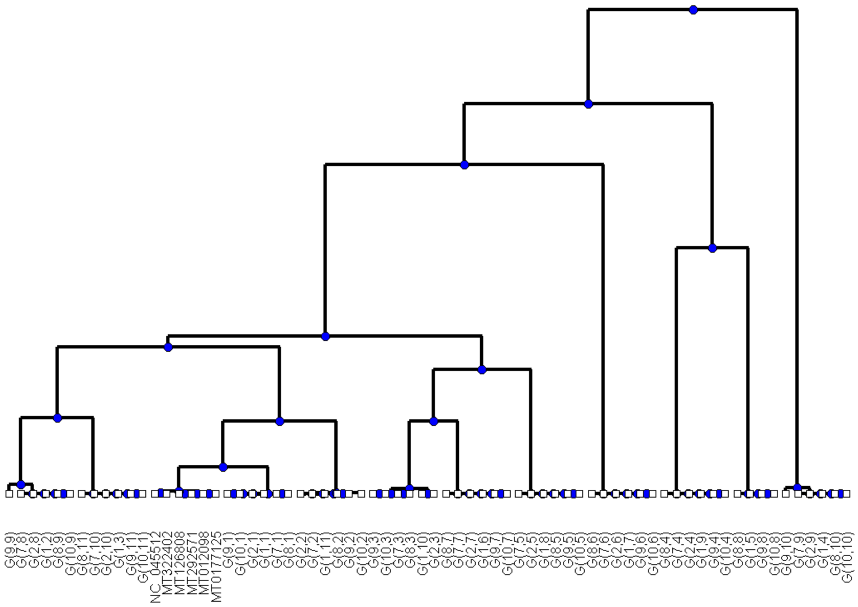

Figure 1: Phylogenetic tree of the six SARS genomes and their associated genes

The above phylogeny depicts the closeness among the six genomes and respective genes based on the 19 compound features. It is observed that the genome S1 is significantly distant from the other SARS-CoV2 genomes. It is noticed that the genes G(10,10) and g(7,10) are significantly away from each other although they both belong to the same genome S10. The pair of genes G(2,9), G(7,9) and G(8,9), G(10,9) are very close to each other but the genes G(2,9) and G(8,9) are far away from each other in the phylogenetic tree. Clearly, the above item-wise findings are reflected in the phylogenetic tree.

doi:10.20944/preprints202004.0486.v1

3.3. Compare and Contrast among Genes and Genomes of MERS, HKU1, α-CoV-1 and β-CoV-1

Based on the spatial features (Table-6) such as SE and HE of different data type of the genes and associated genomes of four genomes MERS, HKU1, α-CoV-1 and β-CoV-1 some salient findings are presented:

- The pyrimidine density for gene E over the genomes S3, S4 and S5 are 55.02%, 59.84% and 53.82% respectively, i.e. the density of pyrimidine bases is significantly richer than that of purine in the gene E of the genomes S3, S4 and S5. It is noted that β-CoV-1 does not possess any E gene.
- The three genomes S3, S4 and S6 contain more pyrimidine bases than purine bases whereas the genome S5 is purine-rich sequence. The density of pyrimidine over the genomes of MERS (S3) and HKU1 (S4) is 53% whereas the density of purine and pyrimidine dominates each other with full of uncertainty of presence and absence ($SE \simeq 1$). The GC content of the genomes S3, S4, S5 and S6 are 41.18%, 32.06%, 38.36% and 36.62% respectively.
- The gene HE is one of the important structural gene which is present in the β-CoV-1 such as in S4 and S6 genomes. The amount of pyrimidine is significantly higher than purine bases in these two HE genes. The GC content of the gene G(6,3) is significantly higher than that of the HE gene G(4,2).
- The purine-pyrimidine representations of the gene ORF1ab of the genomes S4(HKU1) and S6 (β-CoV-1) and genes NS3b, ORF7b of the genomes S3 and S5 respectively are positively trending with HE: 0.62.
- The pyrimidine density of the gene ORF1ab in the genome S3 and S4 is 52% whereas the pyrimidine density of the same gene ORF1ab of the

genomes S5 and S6 is 50%. It is noticed that individual nucleotides density over the gene ORF1ab is different, respective to the genomes S3 and S4 but pyrimidine density are kept balanced in both S3 and S4.

- The spike gene S contains more pyrimidine bases than purine bases in the genome S3, S4 and S6 whereas the spike gene in the genome S6 contains almost equal number of purine and pyrimidine bases.
- The gene N of S6 (β-CoV-1) and ORF3b of the α-CoV-1 genome S5 contain more number of purine bases (54%) than pyrimidine bases. The GC content for the gene N of S6 is 46.4% whereas the GC content of the gene ORF3b in S5 is 31.5% which is significantly lesser that the former one.
- The gene G(3,1)-ORF1ab and G(10,2)- S are pyrimidine rich with percentage 52% and the purine-pyrimidine representation for both the genes possess HE: 0.59; that is the representation is positively trending.
- The gene G(3,6)-NS3d of the MERS genome S3 is pyrimidine rich and the purine-pyrimidine representations is slightly negatively trending. Note that the representation of the nucleotide base G is also negatively trending. This is an unique feature of the gene NS3d in the MERS genome over all the genes and genomes consider for the present study.

Distance among the other four genomes and their associated genes based on euclidean metric are calculated and based on the distance matrix a phylogenetic tree is made which is presented in the Fig. 2.

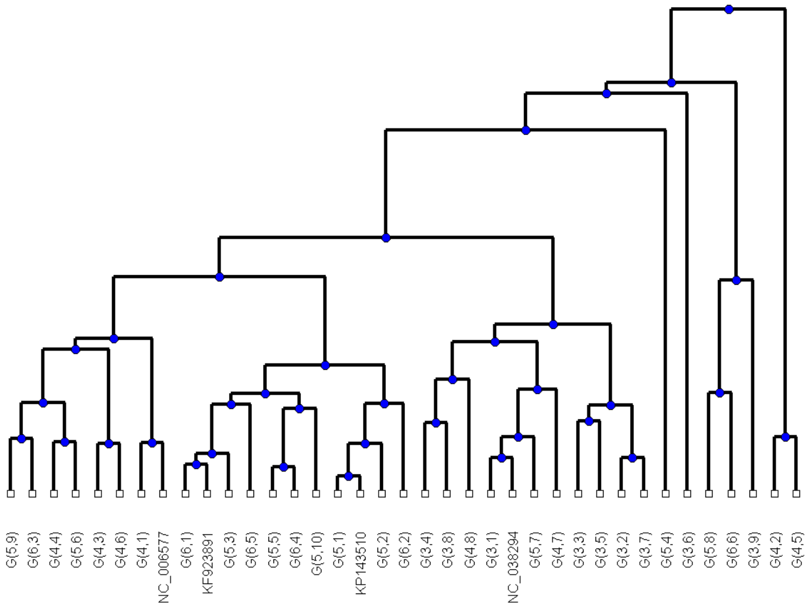

Figure 2: Phylogenetic tree of the four genomes and their associated genes

Based on the phylogenetic analysis as figured in the Fig. 2, it is observed that the genes G(4,2) and G(4,5) are very close to each other. But the gene G(4,4) is quite far away from the gene G(4,4). Also the four genomes are away from each others. The gene G(5,1) and the genome S5 are under the same leaf nodes. Similarly the genome S3 and the gene G(3,1) belong to the same leaf node implying the features vectors are close to each other. Similar observation are made for the gene G(4,1) and the genome S4. Likewise, the gene G(6,1) and the genome S6 are very close to each other.

Although the SARS-CoV2 and SARS-CoV-1, MERS are typically different, the genes and genomes share many spatial features and also differ many features among inter-intra genes and genomes. Some features are reported below based on the result tabulated in the Tables 4, 5 and 6:

- The gene N in genome S5 (α-CoV-1 and S9 (SARS-CoV2) is purine rich with 54% approximately. It is noticed that the GC content is significantly high in the gene N over the genomes S5 and S9 with respective percentage 43.39% and 47.06%.
- The densities of purine and pyrimidine bases over the gene G(10,6)- ORF6 and gene G(5,1)- ORF1a/1b are almost same. That is both the purine and pyrimidine bases are equally likely to be present over the genes and consequently the binary SE is turned out to be 0.9999.
- The genes G(3,4)-NS3b and G(1,8)-ORF7b possess similar purine and pyrimidine density over the sequences. In both the genes, the binary representation of the purine and pyrimidine bases are positively trending. More precisely the purine-pyrimidine representation of ORF7b is much trendier than that of the gene NS3b in the genome S3.
- The presence of purine bases over the genomes S1, S10, S3 (MERS), S9, S7 ans S8 is almost uncertain (SE: 0.999) but the purine-pyrimidine representations are positively trending (HE: 0.65).

4. Conclusions and Summary

This study reports that the genes and genomes sequences of SARS-CoV2 are very much stable (variations are less) in terms of spatial organizations as well as frequency distribution of nucleotides whereas the other genes and genomes of β-CoVs are typically different as mentioned in the results. It is one of the noticeable observations that the genome sequence S1 (NC_045512) and its respective genes are significantly differently spatially organized than other CoV genomes.

Table 3: SARS Genes and their associated genomes with quantitative features

Seq	% (A)	% (T)	% (C)	% (G)	(%) Pu	(%) Py	Pu/Py	% (GC)	B_S	Con_S	S_A	S_C	S_G	S_T	B_H	H_A	H_C	H_G	H_T
G(1,1)	30.18	32.37	17.59	19.87	50.05	49.95	1.00	37.450	1.0000	0.9763	0.88	0.67	0.72	0.91	0.64	0.62	0.57	0.59	0.67
G(1,2)	27.60	36.61	17.49	18.31	45.90	54.10	0.85	35.790	0.9951	0.9658	0.85	0.67	0.69	0.95	0.59	0.57	0.59	0.59	0.63
G(1,3)	29.92	35.90	17.95	16.24	46.15	53.85	0.86	34.190	0.9957	0.9610	0.88	0.68	0.64	0.94	0.65	0.67	0.53	0.52	0.62
G(1,4)	31.75	21.03	25.00	22.22	53.97	46.03	1.17	47.220	0.9955	0.9904	0.90	0.81	0.76	0.74	0.57	0.62	0.51	0.56	0.53
G(1,5)	23.49	45.46	18.18	12.88	36.36	63.64	0.57	31.060	0.9457	0.9180	0.79	0.68	0.55	0.99	0.72	0.71	0.54	0.53	0.72
G(1,6)	29.51	32.24	21.59	16.67	46.18	53.83	0.86	38.250	0.9958	0.9772	0.88	0.75	0.65	0.91	0.60	0.59	0.58	0.52	0.57
G(1,7)	36.56	35.48	13.98	13.98	50.54	49.46	1.02	27.960	0.9999	0.9274	0.95	0.58	0.58	0.94	0.60	0.64	0.51	0.65	0.62
G(1,8)	25.56	31.84	21.82	20.78	46.34	53.66	0.86	42.600	0.9961	0.9895	0.82	0.76	0.74	0.90	0.63	0.64	0.62	0.57	0.66
G(1,9)	21.49	40.35	19.74	18.42	39.91	60.09	0.66	38.160	0.9704	0.9583	0.75	0.72	0.69	0.97	0.62	0.64	0.55	0.57	0.56
G(1,10)	27.17	33.33	21.01	18.48	45.65	54.35	0.84	39.490	0.9945	0.9811	0.84	0.74	0.69	0.92	0.66	0.66	0.60	0.54	0.64
G(1,11)	29.44	33.26	18.92	18.39	47.83	52.17	0.92	37.310	0.9986	0.9756	0.87	0.70	0.69	0.92	0.58	0.60	0.59	0.55	0.62
NC_045512	29.94	32.08	18.37	19.61	49.55	50.45	0.98	37.97	0.9999	0.9785	0.88	0.69	0.71	0.91	0.65	0.61	0.65	0.61	0.68
G(2,1)	30.18	32.38	17.58	19.87	50.05	49.95	1.00	37.444	1.0000	0.9762	0.88	0.67	0.72	0.91	0.64	0.62	0.57	0.59	0.67
G(2,2)	29.43	33.26	18.93	18.38	47.81	52.19	0.92	37.314	0.9986	0.9757	0.87	0.70	0.69	0.92	0.58	0.60	0.59	0.55	0.62
G(2,3)	27.17	33.33	21.01	18.48	45.65	54.35	0.84	39.492	0.9945	0.9811	0.84	0.74	0.69	0.92	0.66	0.66	0.60	0.54	0.64
G(2,4)	21.49	40.35	19.74	18.42	39.91	60.09	0.66	38.158	0.9704	0.9583	0.75	0.72	0.69	0.97	0.62	0.64	0.55	0.57	0.56
G(2,5)	25.56	31.84	21.82	20.78	46.34	53.66	0.86	42.601	0.9961	0.9895	0.82	0.76	0.74	0.90	0.63	0.64	0.62	0.57	0.66
G(2,6)	36.56	35.48	13.98	13.98	50.54	49.46	1.02	27.956	0.9999	0.9274	0.95	0.58	0.58	0.94	0.60	0.64	0.51	0.65	0.62
G(2,7)	29.51	32.24	21.59	16.67	46.18	53.83	0.86	38.252	0.9958	0.9772	0.88	0.75	0.65	0.91	0.60	0.59	0.58	0.52	0.57
G(2,8)	27.60	36.61	17.49	18.31	45.90	54.10	0.85	35.792	0.9951	0.9658	0.85	0.67	0.69	0.95	0.59	0.57	0.59	0.59	0.63
G(2,9)	31.75	21.03	25.00	22.22	53.97	46.03	1.17	47.222	0.9955	0.9904	0.90	0.81	0.76	0.74	0.57	0.62	0.51	0.56	0.53
G(2,10)	29.92	35.90	17.95	16.24	46.15	53.85	0.86	34.188	0.9957	0.9610	0.88	0.68	0.64	0.94	0.65	0.67	0.53	0.52	0.62
MT012098	29.86	32.12	18.39	19.63	49.49	50.51	0.98	38.018	0.9999	0.9786	0.88	0.69	0.71	0.91	0.64	0.61	0.66	0.62	0.68
G(7,1)	30.18	32.37	17.59	19.86	50.04	49.96	1.00	37.45	1.0000	0.9763	0.88	0.67	0.72	0.91	0.64	0.62	0.57	0.59	0.67
G(7,2)	29.44	33.26	18.92	18.39	47.83	52.17	0.92	37.311	0.9986	0.9756	0.87	0.70	0.69	0.92	0.58	0.60	0.59	0.55	0.62
G(7,3)	27.17	33.45	21.01	18.36	45.53	54.47	0.84	39.371	0.9942	0.9806	0.84	0.74	0.69	0.92	0.65	0.66	0.60	0.54	0.64
G(7,4)	21.49	40.35	19.74	18.42	39.91	60.09	0.66	38.158	0.9704	0.9583	0.75	0.72	0.69	0.97	0.62	0.64	0.55	0.57	0.56
G(7,5)	25.56	31.84	21.82	20.78	46.34	53.66	0.86	42.601	0.9961	0.9895	0.82	0.76	0.74	0.90	0.63	0.64	0.62	0.57	0.66
G(7,6)	36.56	35.48	13.98	13.98	50.54	49.46	1.02	27.956	0.9999	0.9274	0.95	0.58	0.58	0.94	0.60	0.64	0.51	0.65	0.62
G(7,7)	29.51	32.24	21.59	16.67	46.18	53.83	0.86	38.252	0.9958	0.9772	0.88	0.75	0.65	0.91	0.60	0.59	0.58	0.52	0.57
G(7,8)	27.60	36.61	17.49	18.31	45.90	54.10	0.85	35.792	0.9951	0.9658	0.85	0.67	0.69	0.95	0.59	0.57	0.59	0.59	0.63
G(7,9)	31.75	21.03	25.00	22.22	53.97	46.03	1.17	47.222	0.9955	0.9904	0.90	0.81	0.76	0.74	0.57	0.62	0.51	0.56	0.53
G(7,10)	29.92	35.90	17.95	16.24	46.15	53.85	0.86	34.188	0.9957	0.9610	0.88	0.68	0.64	0.94	0.65	0.67	0.53	0.52	0.62
MT126808	29.88	32.12	18.38	19.62	49.50	50.50	0.98	38.001	0.9999	0.9786	0.88	0.69	0.71	0.91	0.65	0.61	0.66	0.62	0.68

Table 4: SARS Genes and their associated genomes with quantitative features

Seq	**% (A)**	**% (T)**	**% (C)**	**% (G)**	**(%) Pu**	**(%) Py**	**Pu/Py**	**% (GC)**	**B_S**	**Con_S**	**S_A**	**S_C**	**S_G**	**S_T**	**B_E**	**H_A**	**H_C**	**H_G**	**H_T**
G(8,1)	30.18	32.37	17.59	19.86	50.04	49.96	1.00	37.45	1.0000	0.9763	0.88	0.67	0.72	0.91	0.64	0.62	0.57	0.59	0.67
G(8,2)	29.44	33.26	18.92	18.39	47.83	52.17	0.92	37.311	0.9986	0.9756	0.87	0.70	0.69	0.92	0.58	0.60	0.59	0.55	0.62
G(8,3)	27.17	33.45	21.01	18.36	45.53	54.47	0.84	39.371	0.9942	0.9806	0.84	0.74	0.69	0.92	0.65	0.66	0.60	0.54	0.64
G(8,4)	21.49	40.35	19.74	18.42	39.91	60.09	0.66	38.158	0.9704	0.9583	0.75	0.72	0.69	0.97	0.62	0.64	0.55	0.57	0.56
G(8,5)	25.56	31.84	21.82	20.78	46.34	53.66	0.86	42.601	0.9961	0.9895	0.82	0.76	0.74	0.90	0.63	0.64	0.62	0.57	0.66
G(8,6)	36.56	35.48	13.98	13.98	50.54	49.46	1.02	27.956	0.9999	0.9274	0.95	0.58	0.58	0.94	0.60	0.64	0.51	0.65	0.62
G(8,7)	29.51	32.24	21.59	16.67	46.18	53.83	0.86	38.252	0.9958	0.9772	0.88	0.75	0.65	0.91	0.60	0.59	0.58	0.52	0.57
G(8,8)	23.49	45.46	18.18	12.88	36.36	63.64	0.57	31.061	0.9457	0.9180	0.79	0.68	0.55	0.99	0.72	0.71	0.54	0.53	0.72
G(8,9)	27.60	36.61	17.49	18.31	45.90	54.10	0.85	35.792	0.9951	0.9658	0.85	0.67	0.69	0.95	0.59	0.57	0.59	0.59	0.63
G(8,10)	31.75	21.03	25.00	22.22	53.97	46.03	1.17	47.222	0.9955	0.9904	0.90	0.81	0.76	0.74	0.57	0.62	0.51	0.56	0.53
G(8,11)	29.92	35.90	17.95	16.24	46.15	53.85	0.86	34.188	0.9957	0.9610	0.88	0.68	0.64	0.94	0.65	0.67	0.53	0.52	0.62
MT0177125	29.86	32.12	18.38	19.63	49.50	50.51	0.98	38.016	0.9999	0.9786	0.88	0.69	0.71	0.91	0.65	0.61	0.66	0.62	0.68
G(9,1)	30.18	32.37	17.58	19.87	50.05	49.95	1.00	37.444	1.0000	0.9762	0.88	0.67	0.72	0.91	0.64	0.62	0.57	0.59	0.67
G(9,2)	29.44	33.26	18.92	18.39	47.83	52.17	0.92	37.311	0.9986	0.9756	0.87	0.70	0.69	0.92	0.58	0.60	0.59	0.55	0.62
G(9,3)	27.17	33.45	21.01	18.36	45.53	54.47	0.84	39.371	0.9942	0.9806	0.84	0.74	0.69	0.92	0.65	0.66	0.60	0.55	0.64
G(9,4)	21.49	40.35	19.74	18.42	39.91	60.09	0.66	38.158	0.9704	0.9583	0.75	0.72	0.69	0.97	0.62	0.64	0.55	0.57	0.56
G(9,5)	25.56	31.84	21.82	20.78	46.34	53.66	0.86	42.601	0.9961	0.9895	0.82	0.76	0.74	0.90	0.63	0.64	0.62	0.57	0.66
G(9,6)	36.56	35.48	13.98	13.98	50.54	49.46	1.02	27.956	0.9999	0.9274	0.95	0.58	0.58	0.94	0.60	0.64	0.51	0.65	0.62
G(9,7)	29.51	32.24	21.59	16.67	46.18	53.83	0.86	38.252	0.9958	0.9772	0.88	0.75	0.65	0.91	0.60	0.59	0.58	0.52	0.57
G(9,8)	23.49	45.46	18.18	12.88	36.36	63.64	0.57	31.061	0.9457	0.9180	0.79	0.68	0.55	0.99	0.72	0.71	0.54	0.53	0.72
G(9,9)	27.60	36.34	17.76	18.31	45.90	54.10	0.85	36.066	0.9951	0.9673	0.85	0.67	0.69	0.95	0.59	0.57	0.59	0.59	0.62
G(9,10)	31.75	21.19	24.84	22.22	53.97	46.03	1.17	47.063	0.9955	0.9906	0.90	0.81	0.76	0.75	0.57	0.62	0.51	0.56	0.54
G(9,11)	29.92	35.90	17.95	16.24	46.15	53.85	0.86	34.188	0.9957	0.9610	0.88	0.68	0.64	0.94	0.65	0.67	0.53	0.52	0.62
MT292571	29.87	32.13	18.37	19.64	49.51	50.49	0.98	38.006	0.9999	0.9786	0.88	0.69	0.71	0.91	0.65	0.61	0.66	0.62	0.68
G(10,1)	30.18	32.38	17.57	19.87	50.05	49.95	1.00	37.44	1.0000	0.9762	0.88	0.67	0.72	0.91	0.64	0.62	0.57	0.59	0.67
G(10,2)	29.41	33.26	18.92	18.42	47.83	52.17	0.92	37.337	0.9986	0.9757	0.87	0.70	0.69	0.92	0.58	0.60	0.59	0.55	0.62
G(10,3)	27.17	33.45	21.01	18.36	45.53	54.47	0.84	39.371	0.9942	0.9806	0.84	0.74	0.69	0.92	0.66	0.66	0.60	0.54	0.64
G(10,4)	21.49	40.35	19.74	18.42	39.91	60.09	0.66	38.158	0.9704	0.9583	0.75	0.72	0.69	0.97	0.62	0.64	0.55	0.57	0.56
G(10,5)	25.56	31.84	21.82	20.78	46.34	53.66	0.86	42.601	0.9961	0.9895	0.82	0.76	0.74	0.90	0.63	0.64	0.62	0.57	0.66
G(10,6)	36.56	35.48	13.98	13.98	50.54	49.46	1.02	27.956	0.9999	0.9274	0.95	0.58	0.58	0.94	0.60	0.64	0.51	0.65	0.62
G(10,7)	29.51	32.24	21.59	16.67	46.18	53.83	0.86	38.252	0.9958	0.9772	0.88	0.75	0.65	0.91	0.60	0.59	0.58	0.52	0.57
G(10,8)	23.49	45.46	18.18	12.88	36.36	63.64	0.57	31.061	0.9457	0.9180	0.79	0.68	0.55	0.99	0.72	0.71	0.54	0.53	0.72
G(10,9)	27.60	36.61	17.49	18.31	45.90	54.10	0.85	35.792	0.9951	0.9658	0.85	0.67	0.69	0.95	0.59	0.57	0.59	0.59	0.63
G(10,10)	31.75	21.03	25.00	22.22	53.97	46.03	1.17	47.222	0.9955	0.9904	0.90	0.81	0.76	0.74	0.57	0.62	0.51	0.56	0.53
G(10,11)	29.92	35.90	17.95	16.24	46.15	53.85	0.86	34.188	0.9957	0.9610	0.88	0.68	0.64	0.94	0.65	0.67	0.53	0.52	0.62
MT322402	29.93	32.10	18.35	19.60	49.53	50.45	0.98	37.952	0.9999	0.9798	0.88	0.69	0.71	0.91	0.65	0.61	0.65	0.61	0.68

Table 5: MERS, HKU1, α-CoV-1 and β-CoV-1] Genes and their associated genomes with quantitative features

Seq	% (A)	% (T)	% (C)	% (G)	(%) Pu	(%) Py	Pu/Py	% (GC)	B_S	Con_S	S_A	S_C	S_G	S_T	B_E	H_A	H_C	H_G	H_T
G(3,1)	26.45	32.65	19.35	21.56	48.00	52.00	0.92	40.9	0.9988	0.9852	0.83	0.71	0.75	0.91	0.59	0.58	0.54	0.54	0.58
G(3,2)	25.14	34.12	21.42	19.33	44.46	55.54	0.80	40.74	0.9911	0.9822	0.81	0.75	0.71	0.93	0.60	0.57	0.56	0.60	0.61
G(3,3)	25.00	34.94	22.44	17.63	42.63	57.37	0.74	40.06	0.9843	0.9776	0.81	0.77	0.67	0.93	0.63	0.60	0.59	0.55	0.58
G(3,4)	26.06	29.70	23.94	20.30	46.36	53.64	0.86	44.24	0.9962	0.9933	0.83	0.79	0.73	0.88	0.62	0.59	0.54	0.64	0.64
G(3,5)	25.78	32.66	23.89	17.68	43.46	56.55	0.77	41.57	0.9876	0.9834	0.82	0.79	0.67	0.91	0.54	0.51	0.52	0.57	0.57
G(3,6)	21.04	38.67	23.85	16.44	37.48	62.52	0.60	40.3	0.9543	0.9623	0.74	0.79	0.64	0.96	0.49	0.55	0.60	0.48	0.54
G(3,7)	25.70	34.54	20.48	19.28	44.98	55.02	0.82	39.76	0.9927	0.9799	0.82	0.73	0.71	0.93	0.55	0.57	0.61	0.55	0.61
G(3,8)	24.70	31.67	22.12	21.52	46.21	53.79	0.86	43.64	0.9959	0.9910	0.81	0.76	0.75	0.90	0.63	0.60	0.53	0.59	0.65
G(3,9)	27.88	23.30	25.73	20.01	47.89	49.03	0.98	47.25	0.9997	0.9950	0.87	0.82	0.75	0.78	0.56	0.60	0.54	0.56	0.53
NC_038294	26.23	32.60	20.25	20.93	47.16	52.84	0.89	41.18	0.9977	0.9862	0.83	0.73	0.74	0.91	0.65	0.58	0.66	0.63	0.64
G(4,1)	28.28	39.74	12.05	19.94	48.22	51.78	0.93	31.99	0.9991	0.9380	0.86	0.53	0.72	0.97	0.62	0.62	0.63	0.59	0.66
G(4,2)	25.24	47.46	13.61	13.70	38.93	61.07	0.64	27.3	0.9644	0.8980	0.82	0.57	0.58	1.00	0.61	0.56	0.63	0.58	0.64
G(4,3)	26.21	42.62	14.49	16.68	42.89	57.11	0.75	31.17	0.9854	0.9328	0.83	0.60	0.65	0.98	0.55	0.58	0.60	0.58	0.54
G(4,4)	25.46	39.70	14.55	20.30	45.76	54.24	0.84	34.85	0.9948	0.9516	0.82	0.60	0.73	0.97	0.58	0.56	0.58	0.62	0.60
G(4,5)	24.90	46.99	12.85	15.26	40.16	59.84	0.67	28.11	0.9719	0.9029	0.81	0.55	0.62	1.00	0.69	0.67	0.61	0.54	0.72
G(4,6)	25.30	42.26	14.29	18.16	43.45	56.55	0.77	32.44	0.9876	0.9374	0.82	0.59	0.68	0.98	0.64	0.56	0.58	0.62	0.66
G(4,7)	29.64	31.07	20.74	18.55	48.19	51.81	0.93	39.29	0.9991	0.9828	0.88	0.74	0.69	0.89	0.55	0.60	0.64	0.57	0.65
G(4,8)	28.16	27.51	24.43	19.90	48.06	51.94	0.93	44.34	0.9989	0.9937	0.86	0.80	0.72	0.85	0.59	0.65	0.57	0.58	0.54
NC_006577	27.84	40.10	13.02	19.04	46.88	53.12	0.88	32.06	0.9972	0.9404	0.85	0.56	0.70	0.97	0.67	0.62	0.68	0.65	0.65
G(5,1)	28.88	32.91	16.66	21.55	50.43	49.57	1.02	38.212	0.9999	0.9765	0.87	0.65	0.75	0.91	0.61	0.60	0.57	0.53	0.64
G(5,2)	29.77	31.97	18.25	20.02	49.78	50.22	0.99	38.264	1.0000	0.9794	0.88	0.69	0.72	0.90	0.57	0.58	0.52	0.56	0.59
G(5,3)	26.76	36.15	15.49	21.60	48.36	51.64	0.94	37.089	0.9992	0.9670	0.84	0.62	0.75	0.94	0.61	0.62	0.61	0.59	0.68
G(5,4)	36.94	31.53	14.87	16.67	53.60	46.40	1.16	31.532	0.9962	0.9477	0.95	0.61	0.65	0.90	0.61	0.68	0.63	0.57	0.67
G(5,5)	28.43	36.56	16.11	18.91	47.34	52.66	0.90	35.014	0.9980	0.9626	0.86	0.64	0.70	0.95	0.55	0.58	0.55	0.62	0.58
G(5,6)	26.51	38.15	15.66	19.68	46.19	53.82	0.86	35.342	0.9958	0.9593	0.83	0.63	0.72	0.96	0.55	0.60	0.69	0.52	0.65
G(5,7)	27.55	32.45	18.74	21.26	48.81	51.20	0.95	40	0.9996	0.9835	0.85	0.70	0.75	0.91	0.58	0.65	0.60	0.56	0.60
G(5,8)	32.45	24.16	21.16	22.22	54.67	45.33	1.21	43.386	0.9937	0.9892	0.91	0.74	0.76	0.80	0.53	0.60	0.57	0.58	0.59
G(5,9)	25.49	37.91	17.97	18.63	44.12	55.88	0.79	36.601	0.9900	0.9649	0.82	0.68	0.69	0.96	0.63	0.63	0.53	0.59	0.61
G(5,10)	29.79	33.49	18.04	18.68	48.47	51.53	0.94	36.715	0.9993	0.9734	0.88	0.68	0.69	0.92	0.62	0.59	0.57	0.56	0.60
KP143510	29.06	32.57	17.20	21.16	50.23	49.77	1.01	38.364	1.0000	0.9781	0.87	0.66	0.74	0.91	0.57	0.60	0.62	0.59	0.63
G(6,1)	27.27	36.36	14.31	22.07	49.33	50.67	0.97	36.379	0.9999	0.9622	0.85	0.59	0.76	0.95	0.62	0.60	0.62	0.58	0.65
G(6,2)	31.30	32.50	15.77	20.43	51.73	48.27	1.07	36.201	0.9991	0.9699	0.90	0.63	0.73	0.91	0.59	0.57	0.61	0.55	0.61
G(6,3)	25.39	39.47	16.59	18.55	43.95	56.05	0.78	35.141	0.9894	0.9562	0.82	0.65	0.69	0.97	0.57	0.63	0.56	0.60	0.56
G(6,4)	28.70	35.76	16.41	19.13	47.83	52.17	0.92	35.541	0.9986	0.9659	0.86	0.64	0.70	0.94	0.58	0.55	0.56	0.56	0.59
G(6,5)	29.70	36.06	14.24	20.00	49.70	50.30	0.99	34.242	1.0000	0.9578	0.88	0.59	0.72	0.94	0.58	0.56	0.55	0.60	0.57
G(6,6)	29.84	23.76	22.35	24.05	53.90	46.10	1.17	46.399	0.9956	0.9954	0.88	0.77	0.80	0.79	0.59	0.63	0.63	0.52	0.56
KF923891	27.69	35.69	15.16	21.47	49.15	50.85	0.97	36.626	0.9998	0.9663	0.85	0.61	0.75	0.94	0.61	0.61	0.67	0.63	0.66

A vivid spatial organizations based comparisons are made for a sample of ten genomes and their associated genes which reveal their inter and intra linked relationships among the SARS-CoV2 and other CoVs of the same genus. In near future, we wish to explore the whole set of genome data of SARS-CoV2, MERS and SARS-like other coronavirus in different strains such as Bat, pangolin and etc.

Authors Contributions and Conflicts of Interest:

The author SH has formulated and carried out the study with RKR. The authors SH and RKR analyse the data results and written the manuscript and finally both the authors checked and approved the manuscript. The authors declare that there is no conflicts of interest.

References

[1] L. van der Hoek, K. Pyrc, M. F. Jebbink, W. Vermeulen-Oost, R. J. Berkhout, K. C. Wolthers, P. M. Wertheim-van Dillen, J. Kaandorp, J. Spaargaren, B. Berkhout, Identification of a new human coronavirus, Nature medicine 10 (4) (2004) 368–373.

[2] K. V. Holmes, Sars-associated coronavirus, New England Journal of Medicine 348 (20) (2003) 1948–1951.

[3] C. Sohrabi, Z. Alsafi, N. O'Neill, M. Khan, A. Kerwan, A. Al-Jabir, C. Iosifidis, R. Agha, World health organization declares global emergency: A review of the 2019 novel coronavirus (covid-19), International Journal of Surgery (2020).

[4] P. Sun, X. Lu, C. Xu, W. Sun, B. Pan, Understanding of covid-19 based on current evidence, Journal of Medical Virology (2020).

[5] M. Lipsitch, D. L. Swerdlow, L. Finelli, Defining the epidemiology of covid-19—studies needed, New England Journal of Medicine (2020).

[6] A. S. Fauci, H. C. Lane, R. R. Redfield, Covid-19—navigating the uncharted (2020).

[7] F. Jiang, L. Deng, L. Zhang, Y. Cai, C. W. Cheung, Z. Xia, Review of the clinical characteristics of coronavirus disease 2019 (covid-19), Journal of General Internal Medicine (2020) 1–5.

[8] J. Stebbing, A. Phelan, I. Griffin, C. Tucker, O. Oechsle, D. Smith, P. Richardson, Covid-19: combining antiviral and anti-inflammatory treatments, The Lancet Infectious Diseases (2020).

[9] J. F.-W. Chan, C. C.-Y. Yip, K. K.-W. To, T. H.-C. Tang, S. C.-Y. Wong, K.-H. Leung, A. Y.-F. Fung, A. C.-K. Ng, Z. Zou, H.-W. Tsoi, et al., Improved molecular diagnosis of covid-19 by the novel, highly sensitive and specific covid-19-rdrp/hel real-time reverse transcription-polymerase chain reaction assay validated in vitro and with clinical specimens, Journal of Clinical Microbiology (2020).

[10] M. A. Shereen, S. Khan, A. Kazmi, N. Bashir, R. Siddique, Covid-19 infection: origin, transmission, and characteristics of human coronaviruses, Journal of Advanced Research (2020).

[11] J. Zheng, Sars-cov-2: An emerging coronavirus that causes a global threat, International journal of biological sciences 16 (10) (2020) 1678.

[12] S. Khan, R. Siddique, M. A. Shereen, A. Ali, J. Liu, Q. Bai, N. Bashir, M. Xue, The emergence of a novel coronavirus (sars-cov-2), their biology and therapeutic options, J Clin Microbiol (2020) 00187–20.

[13] S. Zhang, M. Y. Diao, L. Duan, Z. Lin, D. Chen, The novel coronavirus (sars-cov-2) infections in china: prevention, control and challenges, Intensive Care Medicine (2020) 1–3.

[14] C. Ceraolo, F. M. Giorgi, Genomic variance of the 2019-ncov coronavirus, Journal of Medical Virology (2020).

[15] M. A. Marra, S. J. Jones, C. R. Astell, R. A. Holt, A. Brooks-Wilson, Y. S. Butterfield, J. Khattra, J. K. Asano, S. A. Barber, S. Y. Chan, et al., The genome sequence of the sars-associated coronavirus, Science 300 (5624) (2003) 1399–1404.

[16] A. Wu, Y. Peng, B. Huang, X. Ding, X. Wang, P. Niu, J. Meng, Z. Zhu, Z. Zhang, J. Wang, et al., Genome composition and divergence of the novel coronavirus (2019-ncov) originating in china, Cell host & microbe (2020).

[17] J. Xu, S. Zhao, T. Teng, A. E. Abdalla, W. Zhu, L. Xie, Y. Wang, X. Guo, Systematic comparison of two animal-to-human transmitted human coronaviruses: Sars-cov-2 and sars-cov, Viruses 12 (2) (2020) 244.

[18] W.-B. Yu, G.-D. Tang, L. Zhang, R. T. Corlett, Decoding the evolution and transmissions of the novel pneumonia coronavirus (sars-cov-2) using whole genomic data, ChinaXiv 202002 (2020) v2.

[19] C. Cattani, G. Pierro, On the fractal geometry of dna by the binary image analysis, Bulletin of Mathematical Biology 75 (9) (2013) 1544–1570.

[20] C. Cattani, Fractals and hidden symmetries in dna, Mathematical problems in engineering 2010 (2010).

[21] S. S. Hassan, P. P. Choudhury, B. Daya Sagar, S. Chakraborty, R. Guha, A. Goswami, Quantitative description of genomic evolution of olfactory receptors, Asian-European Journal of Mathematics 8 (03) (2015) 1550043.

[22] C. E. Shannon, Prediction and entropy of printed english, Bell system technical journal 30 (1) (1951) 50–64.

[23] S. Noorizadeh, E. Shakerzadeh, Shannon entropy as a new measure of aromaticity, shannon aromaticity, Physical Chemistry Chemical Physics 12 (18) (2010) 4742–4749.

[24] A. Carbone, G. Castelli, H. E. Stanley, Time-dependent hurst exponent in financial time series, Physica A: Statistical Mechanics and its Applications 344 (1-2) (2004) 267–271.

[25] J. Mielniczuk, P. Wojdyłło, Estimation of hurst exponent revisited, Computational Statistics & Data Analysis 51 (9) (2007) 4510–4525.

[26] Y. Benjamini, T. Speed, Estimation and correction for gc-content bias in high throughput sequencing, Nucleic Acids Res 40 (10) (2011) e72.

[27] D. Risso, K. Schwartz, G. Sherlock, S. Dudoit, Gc-content normalization for rna-seq data, BMC bioinformatics 12 (1) (2011) 480.

[28] N. Galtier, G. Piganeau, D. Mouchiroud, L. Duret, Gc-content evolution in mammalian genomes: the biased gene conversion hypothesis, Genetics 159 (2) (2001) 907–911.

[29] F. Hildebrand, A. Meyer, A. Eyre-Walker, Evidence of selection upon genomic gc-content in bacteria, PLoS genetics 6 (9) (2010).

[30] J. K. Das, P. P. Choudhury, A. Chaudhuri, S. S. Hassan, P. Basu, Analysis of purines and pyrimidines distribution over mirnas of human, gorilla, chimpanzee, mouse and rat, Scientific reports 8 (1) (2018) 1–19.

[31] R. K. Rout, S. S. Hassan, S. SINDHWANI, H. M. PANDEY, S. Umer, Intelligent classification and analysis of essential genes species using quantitative

methods, ACM Transactions on Mulitmedia Computing, Communications, and Applications (TOMM) (2019) TOMM–2019.

[32] J. P. Banerjee, J. K. Das, P. P. Choudhury, S. Mukherjee, S. S. Hassan, P. Basu, The variations of human mirnas and ising like base pairing models, BioRxiv (2018) 319301.

[33] R. K. Rout, P. Pal Choudhury, S. P. Maity, B. Daya Sagar, S. S. Hassan, Fractal and mathematical morphology in intricate comparison between tertiary protein structures, Computer Methods in Biomechanics and Biomedical Engineering: Imaging & Visualization 6 (2) (2018) 192–203.