ERICSSON

Ericsson White Paper

A Novel Method of Spatial and Temporal Genome Sequence Tracking Applied to SARS-CoV-2

Background

The Coronavirus genome is characterized as a large non-segmented, single-stranded, positive-sense RNA virus [11]. Beginning in the mid-1960s, a series of four Coronavirus variants (229E, NL63, OC43, and HKU1) have been identified; these Coronaviruses generally cause mild to moderate "common cold" type of infections of the upper respiratory tract [19, 20].

In the past 20 years, however, three novel human coronaviruses have emerged that, unlike the initial four identified Coronaviruses, have caused significant diseases and mortality [13].

1. In December 2019, a set of cases that appeared to present as viral pneumonia emerged in Wuhan, Hubei, China. This represented the beginning of the COVID-19 outbreak, with the first case being confirmed on December 1, 2019 [18]. As of April 15, 2020, the SARS-CoV-2 virus has infected 2,016,020 individuals and killed 129,045. Due to the rapid spread of this novel pathogen, studying its transmission and epidemiology is imperative to combat the current pandemic and prepare for future pandemic events [13]. The Reproduction Number (R0), which defines the number of people that each person with a virus is likely to infect is estimated to be between 1.5 and 3.5 according to The Imperial College Group [21].

2. MERS-CoV (2012): MERS-CoV infected 1831 and killed 787 (35.67%) [12]. The R0 according to the WHO for MERS is < 1 [22].

3. SARS-CoV (2003): SARS-CoV infected 8422 individuals and killed 916 (11%) [16]. For the SARS pandemic in 2003, scientists estimated the original R0 to be around 2.75 but as the pandemic continued it reduced to < 1 [21].

Introduction

In this paper we identify temporal and spatial variations of the genome of the SARS-COV-2 virus, which causes the illness known as COVID-19 over different geographical locations during a set time period. We classify each sequence in groups and subgroups, depending on the level of similarities amongst the sequences. It is hoped that this knowledge will help to identify temporal and spatial variations of the SARS-CoV-2 virus genome such that these considerations may be taken into account in the efforts to develop effective treatments, deploy vaccines and create associated policies to control COVID-19 transmission.

To achieve this, we first obtain the publicly available genome sequences from real time databases such as the NIH genetic sequence database, GenBank [17] and GISAID [17]. We then convert the genome sequences to waveforms and determine a metric that measures the degree of difference or similarity that exists amongst the sequences.

We use waveforms, complex random values which can be analyzed using digital signal processing (DSP) techniques, to reduce time operation complexities and to perform rapid computations. Also, waveforms can be easily analyzed via statistical resampling methods, resulting in higher accuracy of metrics of interest. Finally, for sequence analysis waveforms may be easier to visualize than sequence character form.

This metric, based on the confidence interval of the autocorrelation and cross-correlation functions of the waveforms, is finally compared to various threshold levels, so that each of the sequences can be classified into a tree data structure.

This technique enables a hierarchal classification with different branch levels to represent and to depict distinctive degrees of similarities and allows for the tracing of their evolution paths. Despite the SARS-CoV-2 focus of this study, this method can be applied in general to any other type of genome sequence.

Approach

The complexity of previous methods developed for comparing DNA sequences [1][2] have time complexities of $O(m \cdot n)$, where m and n are the lengths of each sequence in the pair. The method in [3] uses FFT (Fast Fourier Transform) which has a complexity of $O(n \cdot log(n))$ and uses nine FFT operations. Furthermore, the method in [4] has a similar complexity as that of [3] but only uses three FFT operations. The mapping is done from the sequence characters to four fixed numerical values. This method was used to compare two different DNA sequences.

Our approach is an extension to the work in [4] but instead of mapping to four fixed numerical values, we map to random values. One advantage of this approach is that it enables the mapping of patterns instead of just single elements of the sequences since there is no limitation in the random values that can be used. This proves useful in the study of, for example, specific features of the genome sequence where correspondingly, there is no restriction in the number of patterns that can be used. Another advantage of this approach is robust estimation. By taking this approach and then incorporating resampling methods, it is possible to reduce the noise level of the correlation function shown in [4]. This is important, for example, when studying genome sequence virus evolution, where even small variations in the genome sequences may indicate possible virus mutations. Finally, we also propose a metric that measures the degree of correlation of these sequences by comparing them with different threshold levels and classifying the sequences into different groups in a hierarchical structure.

The method we developed is described in the below section:

1. **Input sequences**

   Let the pair of input sequences that we want to compare be denoted by:
   - $\boldsymbol{g_1}$: $(1 \times l_1)$ vector for the first genome sequence and
   - $\boldsymbol{g_2}$: $(1 \times l_2)$ vector for the second genome sequence,

   where $l_1 \geq l_2$ denote the sequence lengths for the first and second sequences respectively.

2. **Conversion from the input sequences to random signal waveforms**

   Let $\boldsymbol{z} = [z(1), \dots, z(q)]$; be a Gaussian complex random vector with a circular symmetric distribution given by $\sim \mathcal{CN}(\boldsymbol{0}, \boldsymbol{I})$, with $\boldsymbol{0}$ $(1 \times q)$ mean vector, and $\boldsymbol{I}$ $(q \times q)$ correlation matrix, where q is length of the alphabet of the input sequences, i.e., the number of unique characters of the input sequence.

There may be different options for mapping the characters of the input sequences to the random vectors. It could be one to one, many to one, or one to many, depending on the sequence features and weighting factors one wants to capture.

In this paper we focus on the one to one mapping, where each element of the input sequence is mapped to an element of the random vector z, as shown below.

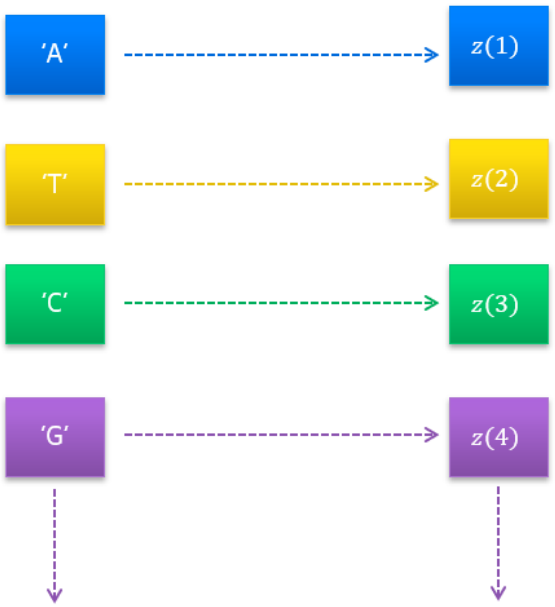

Figure 1: Sequence alphabet to random vector mapping

For each of the input sequences $\boldsymbol{g_1}, \boldsymbol{g_2}$, we obtain the corresponding signal waveforms $\tilde{\boldsymbol{s}}_1, \tilde{\boldsymbol{s}}_2$.

Figure 2: Genome sequence to signal waveform mapping

We finally pad with zeros the shortest output signal waveform, to ensure that both output signal waveforms have an identical length l, where $l = \max(l_1, l_2)$ as indicated by the $(1 \times l)$ vectors below:

$$\boldsymbol{s_1} = [\tilde{s}_1(1), \ldots, \tilde{s}_1(l)]; \boldsymbol{s_2} = [\tilde{s}_2(1), \ldots, \tilde{s}_2(l_2), \ldots, 0]$$

Figure 3 shows the conversion of the pair of input sequence into their corresponding signal waveforms.

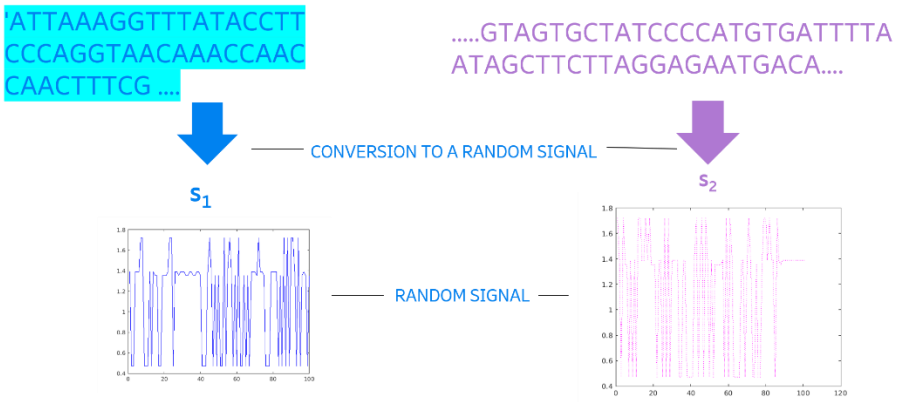

Figure 3: Conversion of the input sequences to their corresponding signal waveforms

3. Metric definition to measure the similarities of a pair of signal waveforms

Let the circular autocorrelation and cross-correlation of the input complex valued signals, as defined in [10], be given by:

$$r_{11}(m) = \sum_{n=0}^{l-1} s_1(n) s_1^*\big((n-m)\big)_l$$

$$r_{12}(m) = \sum_{n=0}^{l-1} s_1(n) s_2^*((n-m))_l$$

where $s_i^*(\quad) = conj(s_i(\quad))$ and $()_l$ denotes the sequence module, i.e., it cyclically repeats every l samples.

And let the correlation $(1 \times l)$ vectors be given by:

$$\boldsymbol{r_{11}} = [r_{11}(1), \dots, r_{11}(l)]; \boldsymbol{r_{12}} = [r_{12}(1), \dots, r_{12}(l)];$$

Furthermore, let's define the following coefficients c_{11}, c_{12}:

$$c_{11} = \max_m(\boldsymbol{r_{11}}(m)), \quad c_{12} = \max_m(\boldsymbol{r_{12}}(m))$$

We propose the following metric:

$$m_{12} = \left(\frac{c_{11}}{c_{12}} - 1\right)$$

This is, we extract the maximum peak of the correlation over all the lags m , to capture the point of maximum agreement between the pair of sequences. As the metric value approaches to as the metric value approaches to zero, no significant variations occur in the input signal waveforms. When the metric value diverges from zero, variations occur in the pair of input waveform signals.

4. Evaluation of correlation coefficients

Since the genome sequences are considerably long (e.g., SARS-CoV-2 genome sequences are ~ 30,000 samples long), we calculate the correlation function in the frequency domain, which simplifies this computation considerably. This is because the correlation of the signal vectors in the time domain becomes an element by element multiplication of the signal vectors in the frequency domain.

Let the n_{fft} point DFT (Discrete Fourier Transform) of the input signals be given by:

$$S_1(k) = \sum_{n=0}^{n_{fft}-1} s_1(n) e^{-j2\pi nk/n_{fft}} \; ; \; k = 0, \dots, n_{fft} - 1$$

$$S_2^*(k) = \sum_{n=0}^{n_{fft}-1} s_2^*(n) e^{-j2\pi nk/n_{fft}} \; ; \; k = 0, \dots, n_{fft} - 1$$

Then we calculate the following:

$$R_{11}(k) = S_1(k) S_1^*(k)$$

$$R_{12}(k) = S_1(k) S_2^*(k)$$

And by applying IDFT (Inverse Discrete Fourier Transform), we obtain:

$$r_{11}(k) = \frac{1}{n_{fft}} \sum_{n=0}^{n_{fft}-1} R_{11}^*(k) e^{j2\pi nk/n_{fft}} \; ; \; n = 0, \dots, n_{fft} - 1$$

$$r_{12}(k) = \frac{1}{n_{fft}} \sum_{n=0}^{n_{fft}-1} R_{12}^*(k) e^{j2\pi nk/n_{fft}} \; ; \; n = 0, \dots, n_{fft} - 1$$

Furthermore, the DFT operation can be performed in segments as described in [5], which allows one to be able to not only perform analysis of long sequences but also analyze the sequence in separated regions.

5. Confidence interval

Furthermore, since the mapping performed in Figure 2 is random, i.e., there is a different mapping at each realization of this process, the evaluated metric is also random. Therefore, it is necessary to focus on the confidence interval of the metric instead of focusing on the value of a specific instantiation of the metric. In this analysis, we choose to evaluate the $\alpha = 99\%$ confidence interval [*] of the metric, as shown below:

$$[m_{12}]_{\alpha}$$

To find this confidence interval, we use bootstrap methods as described by Efron in[6], where we find the empirical distribution of this metric by resampling and extracting the 99% confidence interval.

6. Metric evaluation of all combinations of input pair sequences

First, we calculate all possible metrics for all the combinations of pairs of input sequences. More specifically, for a given number, N, input sequences, we find all possible combination pairs, as follows:

$$\frac{N(N-1)}{2}$$

and calculate their confidence interval metrics as shown in Figure 4.

[*] We have used a confidence interval of α=99% as it effectively captures the most typical variations, however this parameter can be set to other number is appropriate for the analysis.

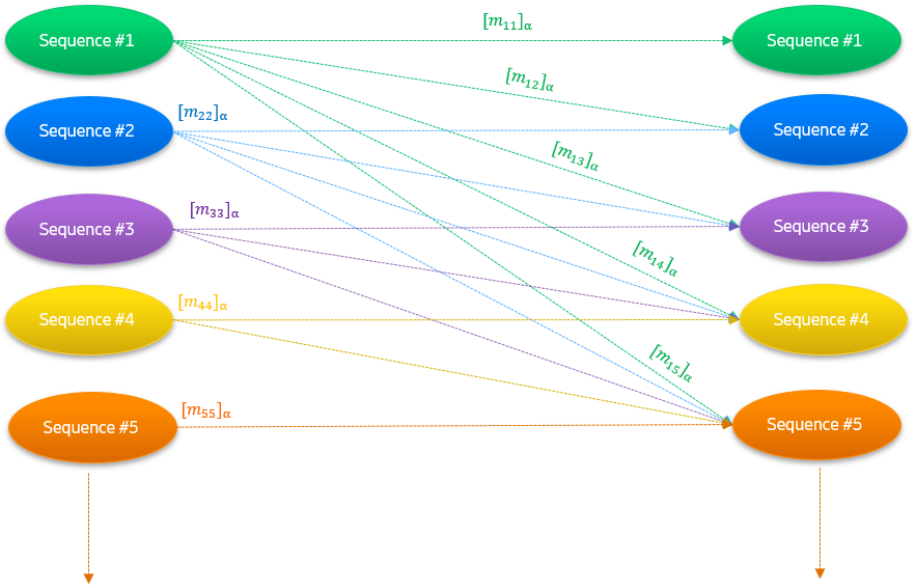

Figure 4: Combinations of input pair sequences

7. Input sequence classification in groups in a tree structure

In this step we classify the input sequences into groups in a tree structure.

Let $\boldsymbol{\Gamma_i^j}$ be a given group where $\boldsymbol{j}$ denotes the group parent and $\boldsymbol{i}$ denotes the group level.

Let $[\boldsymbol{m_j^i}]_{\boldsymbol{\alpha}}$ be the confidence interval of the metric between the sequence of the parent group $\boldsymbol{i}$ and the sequence of group level $\boldsymbol{j}$ (i.e., threshold level T_j)

And, let $[\boldsymbol{m_{jk}^i}]_{\boldsymbol{\alpha}}$ be the confidence interval of the metric of sequences at level $\boldsymbol{i}$, (i.e., threshold level T_i) between sequence of group $\boldsymbol{j}$ and a sequence of group $\boldsymbol{k}$.

We start by classifying an input sequence, as shown in Figure 5, by calculating the metric of the sequence and the groups at threshold level T_1. If the confidence interval of the metric of the sequence and all possible groups at level T_1 are all greater than the threshold T_1, then the sequence forms a new group at level T_1. Otherwise, the sequence falls into the group at level T_1 with the lowest metric value.

To further find, to which group at level T_2 the sequence belongs to, then, for the given parent group, all possible metrics of the sequence with the groups at level T_2 , are calculated. If all metric values are all greater than the threshold value T_2, then a new group is formed at level T_2. Otherwise, the sequence falls into the group at level T_2 with the lowest metric value.

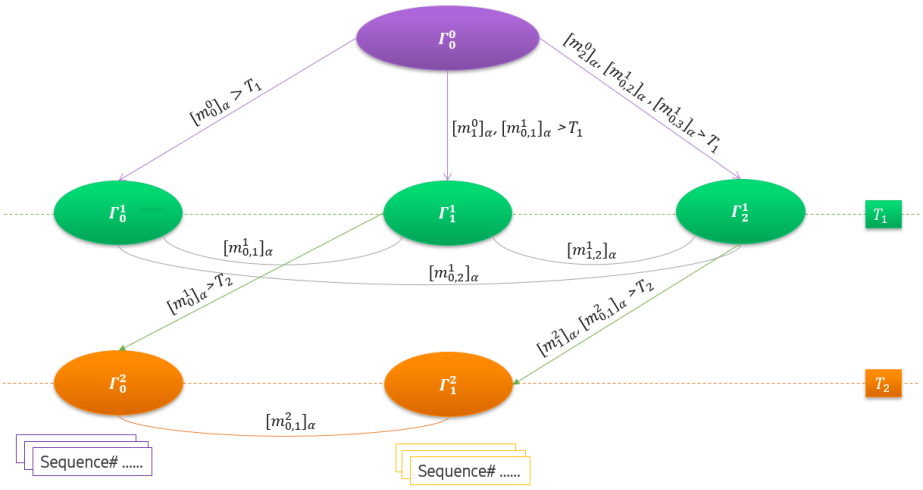

Figure 5: Classification of input sequences in groups in a tree structure

Findings from this study

Spatio-temporal Variations

Table 1 below displays the time stamps of the input sequences and Figure 6 & Figure 7 illustrate their corresponding tree structure and geographic locations.

Sample Name	Sample Location	Date
MN908947	Wuhan, China	1-12-2020
MN994468	California, United States	1-22-2020
MT066156	Lombardy, Italy	1-30-2020
MT093571	Sweden (city unspecified)	2-07-2020
LC528233	Japan (city unspecified)	2-10-2020
MT233519	Spain (city unspecified)	2-27-2020
MT126808	Brazil (city unspecified)	2-28-2020

Table 1: Sample Information – 7 countries

To analyze these sequences, we evaluate the confidence interval of the metric for all possible combinations of pairs of input sequences. We obtain 2 values as shown in Figure 6. We have used the following threshold values: $T_1 = 3,\ T_2 = 1$ to sort these sequences in a hierarchical structure as shown in Figure 7.

It can be seen from these illustrations, that the sequence from Lombardy, Italy, has higher similarity to the sequence from the United States versus the sequence from China.

Using Wuhan-China as a point of reference, Figure 7 shows two groups with the largest variations, i.e., at threshold level T_1, Wuhan-China and Spain.

The reason the sequence from Spain greatly differs from the sequence from Wuhan, China may be due to the incomplete state (two samples missing) of the Spain genomic sequence at the time it was gathered by this team.

It is noted that there are four subgroups under the Wuhan-China group, i.e., at threshold level T_2,: The United States, Italy, Japan and Sweden. Furthermore, Brazil is very similar to Italy and the United States. However, we remark that the Italy genome sequence is the only other sequence with an incomplete state in this study (one sample missing). Therefore, to do a valid comparison, Brazil is classified under the United States Group.

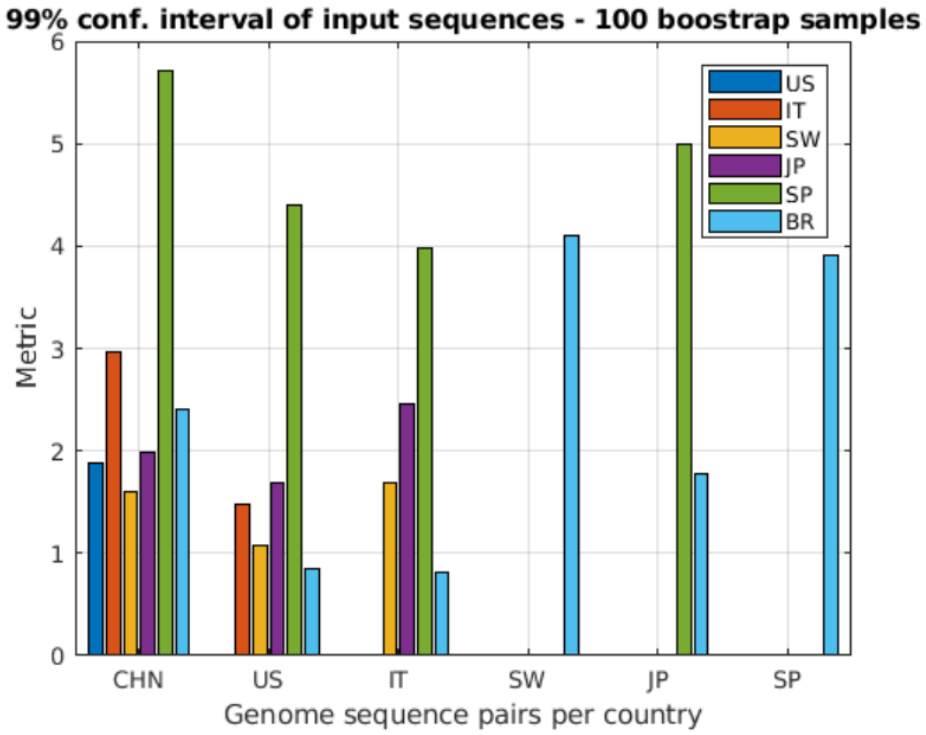

Figure 6: 99% Confidence Interval of the Measured Metric, using 100 Bootstrap Samples

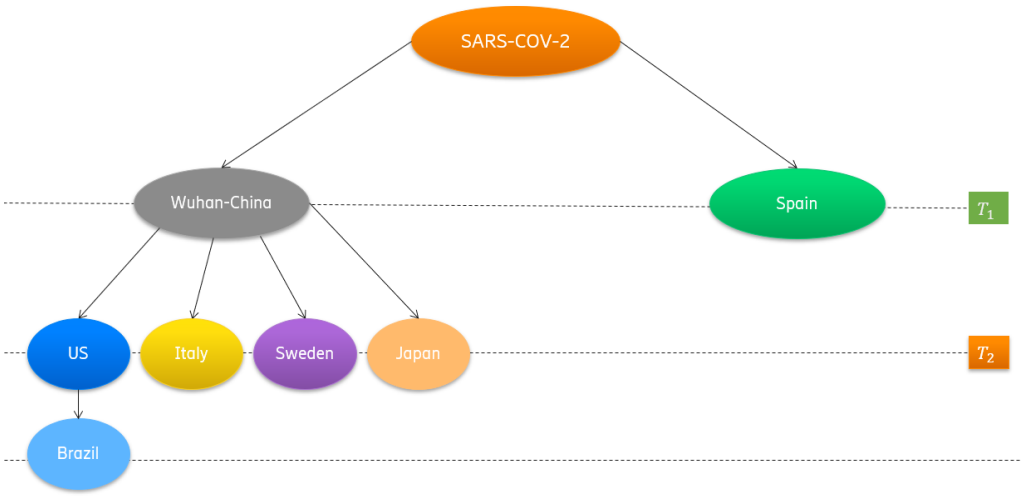

Figure 7: Tree Classification of Genome Sequences from Seven Countries

Figure 8 shows the geographic locations of the seven countries and SARS-CoV-2 migration connections. It is the assertion of this study that one reason why the Brazil sequence is so closely correlated to the United states sequence could be related to the degree to which the United States is a popular transit point for many flying on commercial airlines from China to Brazil.

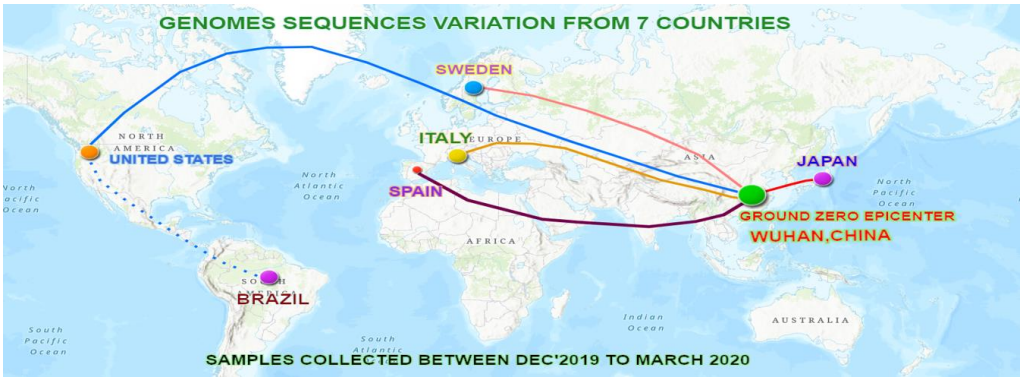

Figure 8: Geographic Genome Sequence Variation from Seven Countries

Temporal Variation

Table 2 below shows the input genome sequence information from China. In this section we are interested only in analyzing the time variations within this particular geographical region.

Sample Name	Sample Location	Date
MN908947	China	20-12-2019
MN996527	China	30-12-2019
MN975262	China	1-11-2020
MT253696	China	1-23-2020
MT135042	China	1-28-2020
MT121215	China	2-02-2020
MN019529	China	2- XX-2020

Table 2: Sample Information – China

Figure 9 demonstrates the metric evaluated in two consecutive samples., i.e., first value of the metric corresponds first and second sequences, second value of the metric corresponds second and third sequences, and so on. First value in the figure shows the metric between 1st and 2nd sequences, while last value in the graph shows the metric between the 6th and 7th samples as indicated in Table 2).

We can observe strong variations for the first five consecutive sequences, indicated by the first four metric values shown in the figure. We note that each of these sequences was taken about 10 days apart. We also note a stronger variation between the fourth and fifth sequences, which were taken five days apart. Finally, there is little variations in the last three sequences. These results indicate that the sampled sequences belong to five different groups, with significant temporal

variations within the region of China that were captured during a two-month period.

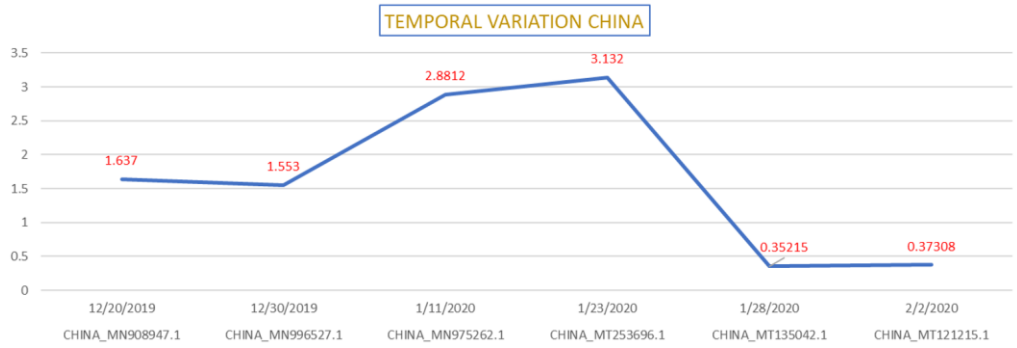

Figure 9: Temporal Variation of Genome Sequence in China

Discussion

We proposed an approach to analyze the variations of genome sequences from different geographic locations over a set period. The method presented in this paper is more robust than some other proposed models as it is able to detect large and small variations within genome sequences of the same virus, by using resampling methods. Also, it is more flexible since it enables for the analysis of different features of the sequence, with no restriction in the number of mapped elements.

This method converts the genome sequences into waveforms and measures their differences with a metric we defined. In this approach, we used digital signal processing tools which allowed us to calculate this metric efficiently. Also, the waveform representation allows an easy visualization of the different patterns of these sequences.

This method is very flexible since it can be easily adapted to measure different features of the sequence. This method is very flexible as it does not require alignment in the input sequences; length segments, tree depths and thresholds can also be configured according the desired behavior. In our study, we analyzed samples from seven countries and found variations for the samples from China, Italy, Japan, the United States, Sweden and Spain. We also found that the sample from Brazil sequence are very similar to the one from the United States. In analyzing these samples from seven countries, we have been able to definitively identify different variations and track their evolution paths.

Conclusion

This approach provides the ability to take the publicly available genome sequences from multiple geographic locations and determine not only the variations in SARS-CoV-2 strains that are circulating, but also the level of correlation between any two strains. This knowledge provides a critical view into whether treatment identification and vaccine production may need to account for these variations in strains and indicates that safety and efficacy trials may need to be conducted against multiple divergent strains.

This method was able to capture genome variations both within a country and between countries and as additional sequences are made available, this analysis can increase its sample size and add to the evidence for the advent of multiple strains of COVID-19 amidst this pandemic outbreak.

The impact of this work is that it tracks spatial and temporal variations of the SARS-CoV-2 virus, allowing us to classify the sequences in different groups based on these variations. This knowledge is crucial for determining approaches to potential treatments and for developing vaccines, since trials may be necessary to be conducted across all the identified groups to ensure vaccine efficiency and to make the determination whether single or multiple vaccines must be tested and put into production to eventually cover a nearly all of the global population.

References

[1] J Mol Biol. 1970 Mar;48(3):443-53.A general method applicable to the search for similarities in the amino acid sequence of two proteins. Needleman SB, Wunsch CD.

[2] Proc Natl Acad Sci U S A. 1983 Feb;80(3):726-30. Rapid similarity searches of nucleic acid and protein data banks. Wilbur WJ, Lipman DJ. Nucleic Acids Res. 2002 Jul 15; 30(14): 3059–3066. doi: 10.1093/nar/gkf436.

[3] PMCID: PMC135756, PMID: 12136088. MAFFT: A novel method for rapid multiple sequence alignment based on fast Fourier transform. Kazutaka Katoh, Kazuharu Misawa,1 Kei-ichi Kuma, and Takashi Miyataa.

[4] E.A. Cheever ; D.B. Searls ; W. Karunaratne ; G.C. Overton, Using signal processing techniques for DNA sequence comparison, Proceedings of the Fifteenth Annual Northeast Bioengineering Conference, Boston, March 1989.

[5] TEM Journal. Volume 8, Issue 1, Pages 56-64, ISSN 2217-8309, DOI: 10.18421/TEM81-07, February 2019. 56 TEM Journal – Volume 8 / Number 1 / 2019. On the Development of STFT-analysis and ISTFT-synthesis Routines and their Practical Implementation Hristo ZHIVOMIROV.

[6] Bradley Efron, R.J. Tibshirani, An Introduction to the Bootstrap (Monographs on Statistics and Applied Probability), with R. Tibshirani. Chapman and Hall, NY, 1993

[7] Janja Paliska Soldo and Ana Sovic Krzic and Damir Sersic, "Fast low-level pattern matching algorithm", ArXiv, 2016, volume=abs/1611.06115.

[8] Jayanta Pal, Soumen Ghosh, Bansibadan Maji, Dilip Kumar Bhattacharya, "Use of FFT in Protein Sequence Comparison under Their Binary Representations", Computational Molecular Bioscience, 2016, 6, 33-40, Published Online June 2016 in SciRes. http://www.scirp.org/journal/cmb

[9] C. Barton, C. Liu, S.P. Pissis (2016) On-Line Pattern Matching on Uncertain Sequences and Applications. In: Chan TH., Li M., Wang L. (eds) Combinatorial Optimization and Applications. COCOA 2016. Lecture Notes in Computer Science, vol 10043. Springer, Cham.

[10] John G. Proakis, D. Manolakis, Digital Signal Processing, Prentice Hall, 1996.

[11] https://onlinelibrary.wiley.com/doi/full/10.1002/path.4454

[12] Middle East Respiratory Syndrome Coronavirus (MERS-CoV) infection: epidemiology, pathogenesis and clinical characteristics. https://www.ncbi.nlm.nih.gov/pubmed/30070331

[13] https://coronavirus.jhu.edu/data

[14] First Case of 2019 Novel Coronavirus in the United States. https://www.nejm.org/doi/full/10.1056/NEJMoa2001191

[15] Genomic characterisation and epidemiology of 2019 novel coronavirus: implications for virus origins and receptor binding. https://www.cdc.gov/coronavirus/2019-ncov/downloads/genomic-characterization-of-2019-nCoV-Lancet-1-29-2020.pdf

[16] Consensus document on the epidemiology of severe acute respiratory syndrome (SARS), World Health Organization 2003. https://www.who.int/csr/sars/en/WHOconsensus.pdf

[17] SARS-COV-2 (Severe acute respiratory syndrome) coronavirus Sequences https://www.ncbi.nlm.nih.gov/genbank/sars-cov-2-seqs/

File	Collection Date	Location	Data Source	Host sex	Virus Identifier (Isolate)	Title	Taxonomy ID
CHINA_WUHAN_LR757998.1	2019-12-26	China: Wuhan	NCBI	Male	BetaCov/Wuhan/WH01/2019	2019-nCoV_WH01 Complete Genome	2697049
CHINA_WUHAN_MN996531.1	30-Dec-2019	China: Wuhan	NBCI				2697049
CHINA_WUHAN_LR757996.1	2020-01-01	China: Wuhan	NBCI	Female	BetaCov/Wuhan/WH03/2020	2019-nCoV_WH03 Complete Genome	2697049
USA_WA_MN985325.1	19-Jan-2020	USA: WA	NBCI		2019-nCoV/USA-WA1/2020		2697049
italy_MT066156.1	30-Jan-2020	Italy	NBCI		SARS-CoV-2/human/ITA/INMI1/2020		2697049
Sweden_MT093571.1	2020-02-07	Sweden	NBCI		SARS-CoV-2/human/SWE/01/2020		2697049
Japan_LC528233.1	2020-02-10	Japan	NBCI		SARS-CoV-2/Hu/DP/Kng/19-027		2697049

Brazil_MT126808.1	28-Feb-2020 l	Brazil (patient traveled from Switzerland and Italy (Milan) to Brazil)	NBCI		SARS-CoV-2/human/BRA/SP02/2020		2697049
Spain_MT233519.1	27-Feb-2020	Spain: Valencia	NBCI		SARS-CoV-2/human/ESP/Valencia5/2020		2697049
CHINA_MN908947.1	Dec-2019	China	NBCI		Wuhan-Hu-1		2697049
CHINA_MN996527.1	30-Dec-2019	China: Wuhan	NBCI		WIV02		2697049
CHINA_MN975262.1	11-Jan-2020	China	NBCI		2019-nCoV_HKU-SZ-005b_2020		2697049
CHINA_MT253696.1	2020-01-23	China: Zhejiang, Hangzhou	NBCI		SARS-CoV-2/human/CHN/HZ-162/2020		2697049
CHINA_MT135042.1	28-Jan-2020	China: Beijing	NBCI		SARS-CoV-2/human/CHN/231/2020		2697049
CHINA_MT121215.1	02-Feb-2020	China: Shanghai	NBCI		SARS-CoV-2/human/CHN/SH01/2020		2697049
CHINA_MT019529.1fasta	23-Dec-2019	China: Hubei, Wuhan	NBCI		BetaCoV/Wuhan/IPBCAMS-WH-01/2019		2697049

[18] https://www.thelancet.com/journals/lancet/article/PIIS0140-6736(20)30183-5/fulltext

[19] https://www.cdc.gov/coronavirus/general-information.html

[20] https://www.cdc.gov/coronavirus/types.html

[21] https://labblog.uofmhealth.org/rounds/how-scientists-quantify-intensity-of-an-outbreak-like-covid-19

[22] https://apps.who.int/iris/bitstream/handle/10665/326126/WHO-MERS-RA-19.1-eng.pdf?ua=1

Author biographies

Dr. Eliana Yepez

Eliana Yepez Eliana has been with Ericsson for more than 10 years and her primary areas of expertise are in channel propagation and statistical analysis of wideband wireless channel measurements. She has a vast experience in the development of algorithms to improve performance of wireless systems. More specifically, her focus has been in the area of MIMO (Multiple Input Multiple Output) antenna element systems. She has worked in the implementation of WiMAX standards, LTE and NR systems, as well as in the development and implementation of physical layer algorithms for simulation tools to improve the throughput performance/cell coverage. Her experience also includes working in the design and implementation of 2.5 GHz radio frequency MIMO channel sounder at Communication Research Center Canada, where she conducted channel measurements campaign to characterize wideband, low mobility, outdoor to indoor wireless channels. She studied the differences and channel capacities of real world measured channels and compared specifically SDMA (Spatial Division Multiple Access) systems with MU-MIMO (multi-user MIMO). She has a Ph.D. degree and an M.Eng. degree in Electrical Engineering in the area of wireless communications and has 2 patents to her name.

Carlos Chaparro

Carlos Chaparro works as a Solution and team lead of the Ericsson Network Manager deployment for customers in the North America Market Area. He is also part of several initiatives in the automation area. Carlos has worked in Customer Service and System integration. He holds a Bachelor of Computer Science degree.

Dr. Kashipati Gonugunta Rao

Kashi Rao works as a Senior Solutions Architect within Digital Services, Ericsson North America and is an industry veteran with over 30 years of experience. He has been with Ericsson for 7 years focusing primarily in the areas of AI/ML, IoT/Packet Core, Cloud, BSS and Media businesses. Prior to Ericsson, he worked for Alcatel-Lucent, Texas Instruments , Cisco Systems and Verizon. He has 20 patents (issued in the United States, Asia, Europe) to his name in AI, computer vision, telecom, video, and VoIP. He has many other patent applications that are in final stages of filing. He has authored over 30 publications with International Journal of Computer Vision, IEEE Conference on Computer Vision and Pattern Recognition, American Association of Artificial Intelligence. Kashi has a Ph.D. in Computer Engineering (Computer Vision, AI) from Univ of Southern California, an MBA from UTD and Bachelor's in Electronics from IIT.

Moloy Kumar Kar

Moloy Kumar Kar works as an End-to-End Solutions Architect within Digital Services, Ericsson North America with primary focus on 5G and Private IoT Networks & Security. His main responsibilities include driving pre-sales engagements, business discussions, solutions architecture and customer success with various Customer Units in the areas of 5G, Private IoT and transport/security. One of Moloy's focus areas is to enable new use cases by driving the evolution of 5G and Private IoT across Massive IoT, Broadband IoT and Critical IoT technology segments. Moloy has been with Ericsson for 14 years and has worked in many areas, from pre-sales consulting, solutions sales, systems integrations, to solutions management, spanning 3G, 4G and 5G technologies. He has worked with radio and core networks, OSS/BSS as well as IP/MPLS, transport, synchronization and security solutions. Moloy holds 3 Master's Degrees – Computer Science, Electrical Engr. specializing in Networks & Telecommunications and an MBA in International Business and preparing to pursue his Ph.D. He recently graduated to become one of our Role Models at Ericsson North America.

Nagaraju Gajula

Nagaraju Gajula is 5G Baseband Software Development manager based out of Austin, Texas. As part of DNEW PDU Development Group, his responsibilities include delivering 5G features in millimeter wave and sub6 based products in Ericsson. He holds master's degree in Computer Science.

Omar Vicencio

Omar Vicencio works in the 5G Smart Factory in Texas, the United States as a fault finding and repair specialist of 5G radios that are manufactured there. His main responsibilities include troubleshooting and fault finding of 5G radios to maintain highest standards of quality. He holds a Master's in Science in Electrical Engineering in the Polytechnic University of Catalonia and went to the Electrical Engineering Graduate School at Stanford University. He earned his Bachelor's in Science in Electrical Engineering from National Autonomous University of Mexico.

Varun Sood

Varun works as a Senior Engineer within Networks, Ericsson North America.

Richard Anthony

Richard has worked for Ericsson for over 25 years (his entire professional career), dating back to a university co-op experience. After early roles in network engineering teams, he has since been in a Product Management role in the areas of CDMA Voice Core, GPRS/UMTS Packet Core, Evolved Packet Core, and defining advanced customer support solutions. He has a Bachelors of Computer Engineering degree from Southern Methodist University.

Dr. Paul McLachlan

Paul McLachlan is Principal Data Scientist at Ericsson's Global Artificial Intelligence Accelerator. He is particularly interested in data privacy and innovation using 5G and edge computing to protect users' data while enabling greater customization and personalization. He is particularly interested in the intersections between 5G and augmented and mixed reality environments as well as digital advertising.