

Complexity in SARS-CoV-2 genome data: Price theory of mutant isolates

Saurav Mandal^{1#}, R.K. Sanayaima Singh^{2#}, Saurabh Kumar Sharma², Md. Zubbair Malik^{1*}, and R.K. Brojen Singh^{1*}

¹*School of Computational and Integrative Sciences,*

Jawaharlal Nehru University, New Delhi-110067, India.

²*School of Computer and Systems Sciences, Jawaharlal Nehru University, New Delhi-110067, India.*

SARS-CoV-2 is a highly virulent and deadly RNA virus causing the Covid-19 pandemic and several deaths across the world. The pandemic is so fast that any concrete theory of sudden widespread of this disease is still not known. In this work, we studied and analyzed a large number of publicly available SARS-CoV-2 genomes across the world using the multifractal approach. The mutation events in the isolates obey the Markov process and exhibit very high mutational rates, which occur in six specific genes and highest in *orf1ab* gene, leading to virulent nature. $f(\alpha)$ analysis indicated that the isolates are highly asymmetric (left-skewed), revealing the richness of complexity and dominance by large fluctuations in genome structure organization. The values of H_q and D_q are found to be significantly large, showing heterogeneous genome structure self-organization, strong positive correlation in organizing the isolates, and quite sensitive to fluctuations in and around it. We then present multiple-isolates hosts-virus interaction models, and derived Price equation for the model. The phase plane analysis of the model showed *asymptotic stability* type bifurcation. The competition among the mutant isolates drives the trade-off of the dominant mutant isolates, otherwise confined to the present hosts.

Keywords: SARS-CoV-2; Multifractal analysis; Hurst exponent; Generalized dimension; Price equation.

Introduction

The special concern about the Covid-19 is the type of pneumonia, which is deadly and highly infectious [1] causing the sudden outbreak of this disease from Wuhan, Hubei Province, China [1, 2] and, spread to different parts of the world that led to WHO to declare this epidemic as an emergency in global public health [3]. Coronaviruses (CoVs) belongs to *Coronaviridae* subfamily (which have $\alpha, \beta, \gamma, \delta$ -CoVs genera) in the family *Coronaviridae* of the *Nidovirales* order [4]. The Covid-19 causing virus *SARS-CoV-2* is the seventh member in β -CoV genus [5], which can infect humans [6], and it's genetic manipulation studies, and sequence analysis has led to the claim that it is not laboratory manipulated virus [6, 7]. It's genome is +ssRNS (single-stranded positive-sense RNA virus) [4, 5] having 30kb genome size [8] which becomes 100-160nm in compact form [4, 6, 7]. Further, it's genome has been found 80% similarity with that of SARS-CoV, whereas around 96% similarity with that of bat coronavirus [9].

Clinical symptoms of Covid-19 include dry cough, sore throat, headache, fever, problems in breathing, dyspnea [3], which may lead to developing problems of organ failure, severe pneumonia, septic shock, acute respiratory distress syndrome, etc. to some patients [10]. However, SARS-CoV-2 enter and infect target host cells (lung, intestine, kidney, and blood vessels) through the receptor angiotensin-converting enzyme 2 (ACE2), and ACE2 expression is found to be increased in diabetes (type 1 or type 2) [11]. Further, hypertension is treated with ACE inhibitors and angiotensin II type-I receptor blockers (ARBs) [11], resulting in an increase in ACE2 expression [12]. Hence, diabetes and hypertension patients might have a severe risk of developing Covid-19 as these patients use ACE inhibitors and ARBs as a drug which upregulate ACE2 expression [13]. Moreover, there are number of reports and unsolved cases regarding the infection of this virus to different people at different geographical conditions, people of different ages, sex and pregnant woman, demographical parameters, etc. [3, 14, 15].

There is no concrete theory so far why SARS-CoV-2 has a high affinity to human host cells, which in turn led to extremely fast human-to-human transmission as compared to other members of CoVs and even among seven members in β -CoV. One belief is that SARS-CoV-2 has receptor binding protein (RBD) that has high cohesive towards human receptor ACE2 [7, 9], it's spike protein has functional polybasic cleavage site at the S1-S2 junction which is quite variable [7, 9, 16]. Then it might allow jumping to the human host cells via comfortable adap-

*Electronic address: zubbairmalik@jnu.ac.in, brojen@jnu.ac.in (Corresponding authors)

tation for fast transmission [7]. However, theory of the origin of this virus and pandemic caused by it are still debatable.

Winsor theory of Covid-19 pandemic

The initial growth of the infected people of Covid-19 pandemic in any country across the world follow interesting Malthusian exponential growth law with time [17] as shown in Fig. 1 adapted from Roser et. al. world data [18]). If $N(t)$ is the time-dependent confirmed SARS-CoV-2 infected population (Fig. 1A), or confirmed death population (Fig. 1B), one can see that each curve of each country follow one of the following growth law at different growth stages [17, 19–21],

$$\frac{1}{N} \frac{dN}{dt} = \Gamma(N, t); \quad \Gamma(N, t) = \begin{cases} C, & \text{Malthus growth law} \\ C - DN, & \text{Logistic growth law} \\ Ce^{-Dt}, & \text{Gompertz growth law} \\ CN \log \left[\frac{D}{N} \right] \mathcal{H}(t), & \text{Winsor growth model} \end{cases} \quad (1)$$

where, C and D are constants, and $\mathcal{H}(t)$ is a function of t . The initial growth rate of Covid-19 pandemic is dominated by *Malthus growth rate* [17], and if we consider significant amount of time of the pandemic, coronavirus infected or death population nicely follow *logistic* [19] or more systematically *Gompertz growth law* [20]. The spread of every disease epidemic or pandemic generally increase monotonically (following Malthus law [17]) due to availability of large susceptible population till it reach threshold value [22] (follow logistic or Gompertz law [19, 20]). After reaching this threshold size of epidemic, it will gradually decrease the size of epidemic due to removal of all susceptible populations and gradual decrease of the virus virulence, die out slowly [22]. In such scenario, the Winsor growth model [21] is quite suitable to explain and forecast, because, the function $\mathcal{H}(t)$ in the solution of Winsor model in equation (1), $N(t) = De^{-e^{\mathcal{H}(t)}}$, can have the change in sign, and can be expanded through Taylor series,

$$\mathcal{H}(t) = \sum_{i=0}^{\infty} \frac{1}{i!} \left[\frac{d^i \mathcal{H}}{dt^i} \right] t^i \sim u - vt + O(t^2) \quad (\text{one can see the change in sign in the second term}).$$

Thus, we have Winsor growth law, $N(t) = De^{-e^{u-vt}}$, which is modified form of Gompertz model, can well explain the pandemic. The uncontrolled fast growth of $N(t)$ due to lack of proper checking in the $N(t)$ outgrowth may cause inevitable shortage of food, trapping in (starvation- $N(t)$ outgrowth) cycle which may lead to fast decrease in $N(t)$, the condition known as *Malthusian catastrophe* given by, $\left(\frac{dN}{dt} \right) C$ [17, 23]. However, happening of this condition is unlikely, even though the dynamical process could lead to the sacrifice of a certain population of infected people $N_U(t)$, who is most probably unhealthy/non-immune population out of $N(t)$, and the most fitted one $N_F(t) = N(t) - N_U(t)$ will survive. But, the question of why the growth rates of both infected and death are different in different countries at the epidemic network level or genetic level and how to control the epidemic/pandemic size are still open questions.

Mutation effect: Evolution of SARS-CoV-2 divergence

Mutations are considered as the basic units of species evolution [24], which are sometimes harmful and sometimes beneficial [26] but play important roles in genetic diversity via natural selection [25]. RNA viruses are generally mutation prone viruses during virus-host interaction, and their mutation rates are high ($10^{-4} - 10^{-6}$) due to lack of proofreading activity, and the small increase in mutation rates in them may cause inability to evolve with the host [26]. It is also found that this high mutation rate intensifies virulence and coevolution of the virus with the host [24]. Further, these mutations in them are accountable for viral genome replication [27]. Hence high mutation rates in them have some significance, namely, in helping adaptability to the host [27], in enhancing the degree of viral genome replication kinetics [28] etc. leading to increase in their virulence causing human pandemic.

We have mined publicly available complete genomic data of various SARS-CoV-2 isolates (42 isolates in total) for seven countries (Supplementary TABLE 1). Taking Wuhan seafood market isolate with accession number NC_045512.2 as the reference genome, sequence comparison of these isolate genomes of SARS-CoV-2 with the reference genome is done, and clustered the isolates as phylogenetic tree (Fig. 2). The phylogenetic tree thus obtained (Fig. 2) indicates that there are approximately three main clusters of isolates (shades colors in Fig. 2) distributed across the countries. Surprisingly, it is also found that some of the isolates obtained in other countries (LR757996, MN988668, MN988669, MN996528, MN996530, MT019532) are completely similar to that of the reference genome. Further, the distribution of the virus isolates across these countries are found to be random.

We, then, calculated mutation rates (mutation per unit nucleotide) of the mined isolates, and found that the mutation rate (m_r) of these isolates are very high ($m_r \in [10^{-4} - 10^{-5}]$), which could be a reason of causing the pandemic. The calculated mean value of the mutation per gene of the virus with error bars given by first and third quartiles (Fig. 2A) indicates that mutations take place at some specific genes only (mainly in *orf1ab*, *S*,

ORF3a, *M*, *ORF8*, *N* genes). The results show that m_r of *orf1ab* and *ORF3a*, such that $m_{orf1ab} \rangle m_{ORF3a}$, and these two genes could be key genes for enhancing virus virulence and transmission [24, 27] which may lead to a disease epidemic. Further, the mean values of mutation rates of all isolates in each country are calculated and found the overall mutation rates S. Korea, Sweden, Japan, and the USA are large in comparison. The confirmed Covid-19 positive case, as well as death rate in these countries, are found to be comparatively large in comparison with other countries (Fig. 1), indicating an increase in mutation rate enhances the virulence of SARS-CoV-2 virus [24, 28]. Further, we calculated mutation rates of all six types of mutations $N_i \rightarrow N_{i-1}$, where, $i = 1, 2, 3, 4$, $N_i \in [A, C, T, G]$ and, for the sake of simplicity, $N_i \rightarrow N_{i-1}$ and $N_{i-1} \rightarrow N_i$ are taken to be the same in the calculation (Fig. 3B). The results show that the average mutation rates of these mutation types in each country are found to be random and follow the Markov process [29]. It is observed that the transition is more frequent than transversion. During the evolution of genomes, the transition is more frequently observed as compared to transversion as the former usually does not lead to altering the protein, but the later has a higher impact on the change in the protein structure and function [30]. Some isolates of Wuhan, China contains transversion indicating that there is rapid evolution leading to adaptation with the new environment for survival or possibly making these isolates more virulent compared to their counterpart. On the other hand, *LR757995* out of a total of 15 isolates only consist of $C \rightarrow T$ mutation type at genomic location 8781. This specific mutation is seen to mutate along with mutation from $T \rightarrow C$ at genomic position 28143. These two mutations at positions 8781 and 28143 are highly prevalent in the USA (14 out of 16) and Japanese isolates (3 out of 4). Besides the isolates available from South Korea, Nepal, Sweden, one isolate from Taiwan and Australia does not contain these two specific mutations. It is probable that the genomic isolates spreading in the USA are less virulent compared to the reference Wuhan virus.

On the other hand, out of all Japanese isolates available, only one isolate showed transversion mutation at genomic location 29704. This mutation is at the coding region and codes for the 3' stem-loop II-like motif (also known as s2m) whose function is not yet fully known, but have been found to have a role in viral replication [31]. But we observed that most of these isolates have transitional mutation $C \rightarrow T$ at genomic location 8781. This mutation falls in the coding region, namely, *nsp4* and codes for transmembrane domain 2(TM2). Maximum mutation type is observed to be $C \rightarrow T$ in most genomic isolates falling in the *orf1ab* gene (Fig. 3A and B).

Complexity in SARS-CoV-2 isolates:

We analyze the genomic complexity of the SARS-CoV-2 isolates using multifractal analysis [32, 33] to get an insight into the impact of genome variability and to understand the virulence of the virus isolates in a different perspective. In this technique, we first convert the symbolic genome sequence to time series like DNA walk by taking purine(A or G) as step-up (+1) and pyrimidine as step down (−1) [see the *Methods*]. DNA walk is a cumulative sum, which means that the effect of a single point mutation can lead to different walk altogether, and hence we can quantify them using Multifractal parameters [see the *Methods*]. Then multifractal analysis can be done from the DNA walk of each viral isolate to calculate various multifractal parameters. The *DNA walks* of all SARS-CoV-2 RNA virus isolates collected from the patients' data from various countries (42 isolates) as a function of order parameter q obey Markov process (Fig. 4A). The calculated *fluctuation function* $F_q(s)$ of order q as a function of normalized s [$s = \text{int}(S/L)$, L is the length of the DNA walk of the virus genome] for all SARS-CoV-2 virus isolates follow overall power-law behavior $F_q(s) \sim s^{H_q}$, where, H_q is the Hurst exponent of order q [34] (Fig. 4B) indicating fractal nature in the virus genome organization. Further, the calculated H_q as a function of q is found to be q and s dependent (Fig. 4B and Fig. 5D) indicating the scaling nature in $F_q(s)$ is inter-event dependent showing multifractal nature [32]. Since the F_q behavior indicate *local* multiple fractal natures at various local domains of s with various local Hurst exponents, $H_q^{[1]}, H_q^{[2]}, \dots, H_q^{[n]}$, $F_q(s)$ exhibits self-affine multifractal property via observable variable $z(s)$ [32, 33, 35] given by,

$$\begin{aligned} \frac{z(sx)}{z(x)} &= F_q(s) = F_q(s_1, s_2, \dots, s_n); \quad F_q(s) \sim s^{H_q} \quad \forall 0 < s_1, s_2, \dots, s_n < 1 \\ F_q(s_1, s_2, \dots, s_n) &= F_q^{[1]}(s_1) F_q^{[2]}(s_2) \dots F_q^{[n]}(s_n); \quad F_q^{[n]}(s_k) \sim s_k^{H_q^{[k]}} \\ H_q &= \sum_{k=1}^n H_q^{[k]} \end{aligned} \quad (2)$$

not clear And each local segment s_k satisfy fractal law [33]. We then calculated the average F_q^A of all virus isolates in each country with error bars indicated by first and third quartiles (Fig. 4C). We found that F_q^A in S. Korea is much higher than in other countries, and F_q^A variability is much more in Wuhan and then in Japan. The high value of F_q^A indicates the presence of multiple isolates in Wuhan, and induces large variability in the virus genome could

cause large variability in F_q^A which might be harm to the hosts in the disease epidemic.

We then calculated singularity function $f(\alpha)$ of all the virus isolates (Fig. 4D) and found the curves collapse roughly at $\alpha \sim 1$ and left and right hand parts of of the curves around the value of $\alpha = \alpha^*$ at which $f(\alpha^*) \rightarrow f_{max}$ (f_{max} is the maximum value of f) is quite biased and even left parts of the curves about α^* are quite small. Since $f(\alpha^*) \rightarrow f_{max}$ for $\alpha^* \sim 1.13 \pm 0.012$ for all curves, the virus isolates are of heterogeneous structure driving genome organization complex [36] (Fig. 4D). To understand the behavior of f around α^* , we can expand $f(\alpha)$ around α^* by Taylor series expansion,

$$f(\alpha) = \sum_{k=0}^{u^*} \frac{1}{k!} c_k \Delta \alpha^k; \quad c_k = \frac{1}{k!} \frac{d^k f(\alpha^*)}{d\alpha^k}, \quad \Delta \alpha = [\alpha - \alpha^*] \quad (3)$$

where, u^* is the degree of $f(\alpha)$ after truncation. Then the symmetry of $f(\alpha)$, which can characterize the complexity in the system, is defined by a skew parameter χ ,

$$\chi = \frac{\Delta \alpha_R}{\Delta \alpha_L}; \quad \Delta \alpha_R = \alpha_{max} - \alpha^*, \quad \Delta \alpha_L = \alpha_{min} - \alpha^* \quad (4)$$

where, $f(\alpha)$ is right-skewed if $\chi > 1$, left-skewed if $\chi < 1$ and perfect symmetry if $\chi = 1$ [32]. It is clearly seen from Fig. 4D that $\Delta \alpha_R \gg \Delta \alpha_L$ indicating asymmetric nature of χ . Hence, we calculated average $\Delta \alpha^A \approx \Delta \alpha_R^A$ of each country taking into account all the virus isolates in the country (Fig. 4E), and found that the value of $\Delta \alpha^A$ is much higher in Sweden, Australia, Taiwan and USA indicating richness in genome structure showing strong signature of multifractality, leading to the complex behavior of the virus. However, the variability of $\Delta \alpha^A$ is quite large in Wuhan, which could be due to the presence of multiple isolates indicating virulence of the virus leading to epidemic [37, 38]. Then we calculated average χ_A for the countries we have taken and found that $\chi^A < 1$ (left-skewed) revealing the dominance of the virus genome organization and scaling in multifractality by large fluctuations [32, 36] (Fig. 4F). The values of χ^A are found to similar for all the countries, but the variability is much large for Wuhan, Taiwan and Japan which could be dependent of the large fluctuations in the virus isolates on demography.

Now, we calculated generalized dimension D_q for the virus isolates from their corresponding DNA walks using,

$$D_q = \lim_{r \rightarrow 0} \frac{1}{q-1} \frac{\ln[\sum_{k=1}^N U_k^q(r)]}{\ln(r)} \quad (5)$$

where, $U_k^q(r) = \lim_{N \rightarrow \infty} \frac{N_k}{N}$ is the probability that k th DNA walk segment of length scale r will have N_k observations, and obeys $U_k^q(r) \sim r^\lambda$ [34], where, λ is Holder exponent [39]. The results of D_q shows that D_q is quite dependent on q (Fig. 5A) indicating rich heterogeneous structure in the virus isolates [32], where, local structures in the genome could have various properties and functions. The changes in those local structure may induce drastic change in genome as well as phenotypic behavior of the virus isolates. We then calculated relative change of the isolates with respect to reference virus genome $D_q - D_q^C$ with q (Fig. 5B) shows significant changes in the parameters and $D_q - D_q^C > 0$ indicating increase in complexity due to increase in heterogeneous structure as compared to reference virus genome. The complexity in virus isolate in Japan is quite high (D_q^A is highest) as compared to other countries (Fig. 5C), and variability in complexity in Wuhan is much more as compared to other countries.

We, then, calculated H_q as a function of q (Fig. 5D) by using the MFDFA approach described in *Methods*. Since the H_q values are quite variable with respect to q , the isolate genomes show high heterogeneous structures. Further, since $H_q > 0.5$, genomic signal in the whole virus isolate genome sequences indicate positive correlations in them [40] indicating possibility of strong self-organizing behavior in each virus isolate [32, 41]. The calculated change in H_q with respect to reference virus genome $H_q - H_q^C > 0$ is positive (Fig. 5E), the correlations in the virus genome increases in the mutated isolates which could be one reason for increase in virulence in virus isolates. The average value in H_q^A and variability are much more in Wuhan virus isolates as compared to other countries (Fig. 5F). The multifractal analysis of the virus isolates across the countries reveals that mutated virus isolates have increased virulence as compared to the reference virus. Further, presence of multiple isolates in a host also enhances virulence of the virus. The virus isolate is quite sensitive to fluctuations and might correlated to increase in virus mutations which cause heterogeneity in genome structure exhibiting strong multifractal properties.

Price theory of evolution of mutant isolates in infected disease

One of most important basic reasons of outbreak of an epidemic could be due to sudden trade-off of multiple viral

species modulated by high mutation effects, which obey Markov process, due to virus-hosts interaction [26, 42]. This epidemic process could be dependent on various factors, namely, virus virulence, transmission (passive and active) etc [43]. Let us consider M mutationally different isolates of the virus in the hosts, and x_i be the quantitative trait of i th individual virus or parasite ($0 \leq x_i \leq 1$) in the hosts, then the dynamics of mean of this variable (\bar{x}) is given by Price equation [43–45],

$$\frac{d\bar{x}}{dt} = cov(x, \omega) + E(\omega \Delta x) + E \left[\frac{dx^s}{dt} \right] = \bar{\omega} \Delta \bar{x} \quad (6)$$

where, $cov(x, \omega) = \sum_i q_i \Delta x_i \Delta \omega_i$ and $E(\omega \Delta x) = \sum_i q_i \omega_i \Delta x_i$ are the covariance and expectation value functions, with q_i as the frequency and x^s as the expected value of survived individual virus during $[t, t + dt]$ respectively [44]. The model we consider incorporates the mechanism of infecting susceptible hosts S by a virus isolate of type ' i ' (y_i is the concentration of the i th isolate) and then turn into new dominating mutant isolate type ' $j = i + 1$ ' due to mutation event in hosts-virus interaction [46]. The model allows the presence of different isolate types during the interaction with the host causing more virulent harming the hosts which could be one of the most important factors which could drive the virus transmission among the hosts [37, 38]. Then the dynamics of i th isolate variable y_i is given by,

$$\frac{dy_i}{dt} = y_i [\omega_i S - d - \mu] + \mu_1 \sum_{j=1}^{i-1} p_{ij} y_j - \mu_2 \sum_{j=i+1}^M p_{ij} y_j \quad (7)$$

where, $i = 1, 2, \dots, M$, ω_i corresponds to fitness parameter, which is the transmission rate at which hosts infected by isolate ' i ' infect S , d is the death rate of hosts, μ_1 is the rate of generating self mutant ' i ', μ_2 is the rate of generating new dominating mutant isolate ' j ' and p_{ij} is the probability that infectious isolate ' i ' turned into isolate ' j ' and $\mu = \mu_2 - \mu_1$. Then we can define, $y = \sum_{i=1}^N y_i$, and frequency, $q_i = \frac{y_i}{y}$, and equation (7) becomes, $\frac{dq_i}{dt} = \frac{1}{y} \frac{dy_i}{dt} - \frac{y_i}{y^2} \frac{dy}{dt}$. Now substituting (7) to this expression, using $\sum_{i=1}^M \sum_{j=1}^M p_{ij} y_j = 1$ and after simplification, we arrived at the following equation,

$$\frac{dq_i}{dt} = S q_i [\omega_i - \bar{\omega}] + \mu_1 \sum_{j=1}^{i-1} p_{ij} q_j - \mu_2 \sum_{j=i+1}^M p_{ij} q_j - \mu_1 \bar{q}_{i-1} + \mu_2 [1 - \bar{q}_i] \quad (8)$$

where, $\bar{q}_i = \sum_{j=1}^i p_{ij} q_j$. Now, the mean \bar{x} can be obtained from $\bar{x} = \sum_{i=1}^N x_i q_i$. Then using to the equation (8), simplified the obtained equation and rearranging the terms, we have the dynamics of \bar{x} of the model by the following Price equation,

$$\frac{d\bar{x}}{dt} = S \times cov(x, \omega) + \mu \bar{x} [M - 1] \quad (9)$$

The fitness parameter of individual virus isolate type i of j th parent is composed of two forces arises from competition in the host and increase virulence, given by, $\omega_{ij} = \frac{x_{ij}}{x_i} (1 - x_i)$ [43]. Hence, the fitness of the i th mutant isolate group will be given by, $\omega_i = (1 - x_i)$. Now, this ω_{ij} become one important parameter that determine virus isolate trade-off which can in general represented by, $\omega = x f(x)$ [42], such that, $f(x) \approx \sum_i \frac{df^i(x_0)}{dx} (\Delta x)^i \sim 1 - \alpha x + \lambda x^2 + O(\Delta x^2)$, where, $\Delta x = x - x_0$, $f(x_0) - x_0 \frac{df(x_0)}{dx} + x_0^2 \frac{d^2 f(x_0)}{dx^2} = 1$, $\alpha = 2 \frac{d^2 f(x_0)}{dx^2} - \frac{df(x_0)}{dx}$ and $\lambda = \frac{d^2 f(x_0)}{dx^2}$. Then, one can approximate fitness parameter to be, $\omega \sim x [1 - \alpha x + \lambda x^2]$. The optimized fitness of the virus can be obtained by calculating the equilibrium state of ω and is dependent on various factors to control the virus transmission/diffusion [42, 47].

Now, if one consider κ_i is the genetic variability (additive in nature) of the i th mutant isolate group, then, x_i can be expressed as, $x_i = x + \epsilon \kappa_i$, where, $\epsilon \ll 1$ [43]. Now, we have, $\Delta x = x_i - x = \epsilon \kappa_i$, and $\Delta \omega = \omega_i - \omega = (1 - x_i) - (1 - x) = -\epsilon \kappa_i$. Substituting these values, we have, $cov(x, \omega) = \sum_{i=1}^M q_i \Delta x_i \Delta \omega_i = -\epsilon^2 \sum_{i=1}^M q_i \kappa_i^2 = -\epsilon^2 \Lambda_g$, where, Λ_g is the genetic variance of the mutant isolates in the group. Now, putting back to the equation (12) and rearranging the terms, we have,

$$\frac{d\bar{x}(t)}{dt} - \mu \bar{x}(t) [M - 1] = -\epsilon^2 S \Lambda_g \quad (10)$$

This is special Bernoulli differential equation with constant coefficient and inhomogeneous term, which can be solved easily by the technique of calculating integrating factor (IF), where, $IF = e^{-\mu(M-1)t}$, or by simple separation of variables method. Then after solving the equation (10), the expression for \bar{x} is given by,

$$\bar{x}(t) = \frac{\epsilon^2 S \Lambda_g}{\mu(M-1)} + \left[\bar{x}_0 - \frac{\epsilon^2 S \Lambda_g}{\mu(M-1)} \right] e^{\mu(M-1)t} \quad (11)$$

where, \bar{x}_0 is the amount of \bar{x} at $t = 0$.

Theorem 1: Optimal trait value has the properties: $\bar{x}^* \propto \Lambda_g$ and $\bar{x}^* \propto \frac{1}{\mu}$.

Proof: The optimal trait value \bar{x}^* can be obtained by taking $\frac{d\bar{x}(t)}{dt} = 0$ in equation (10), or by taking $\lim_{t \rightarrow 0} x(t)$ in equation (11). It is given by,

$$\bar{x}^* = \frac{\epsilon^2 S \Lambda_g}{\mu(M-1)} \quad (12)$$

From this equation (12), we can directly obtain, $\bar{x}^* \propto \Lambda_g$, and $\bar{x}^* \propto \frac{1}{\mu}$.

Lemma 1: Phase plane analysis of the virus-pathogen interaction system modeled by (10) shows that mutational rate drives the system an "Asymptotic stability".

Proof: The equation (10) can be written as $\frac{d\bar{x}}{dt} = a\bar{x} - b = F(x, a)$, where, $a = \mu[M-1]$ and $b = \epsilon^2 S \Lambda_g$. Fixed point of the dynamical system is given by, $\bar{x}^* = b/a$, and hence, $\frac{d\bar{x}}{dt} = a[\bar{x} - \bar{x}^*]$. The stability of the dynamical system around a fixed point can be analyzed by calculating, $\lambda = \left. \frac{dF(x, a)}{d\bar{x}} \right|_{\bar{x} \rightarrow \bar{x}^*} = a$, such that, for $\lambda > 0$, \bar{x}^* is unstable, whereas, for $\lambda < 0$, \bar{x}^* is stable [49]. Phase plane analysis of this dynamical system can be done as follows: (a) for $a > 0$: (i) $\frac{d\bar{x}}{dt} > \frac{b}{a}$ if $\bar{x}_0 > \frac{b}{a}$, such that $\bar{x} \rightarrow \infty$ departing from $\bar{x}^* = \frac{b}{a}$; (ii) for $a < 0$: $\frac{d\bar{x}}{dt} < \frac{b}{a}$ when $\bar{x}_0 < \frac{b}{a}$, indicating $\bar{x} \rightarrow -\infty$; and (iii) for $\bar{x}_0 = \frac{b}{a}$ shows $\frac{d\bar{x}}{dt} = \bar{x}^* = \text{constant}$. Therefore, the system is unstable, and the system will diverge for slight perturbation to it. Further, \bar{x}^* , in this case, is unstable fixed point because $\lambda > 0$ for $a > 0$. This condition corresponds to the Malthusian explosion. Following a similar argument for the case $a < 0$, it can be shown that \bar{x}^* is a stable fixed point. Hence, the bifurcation diagram of this dynamical system exhibit asymptotic stability (Fig. 6). Since, $\mu = \mu_2 - \mu_1$, we have, $a = (\mu_2 - \mu_1)[M-1]$. The bifurcation diagram indicates that for $\mu_2 > \mu_1$, \bar{x}^* is unstable fixed point. This is the condition where the rate of generating new dominating mutant isolates is much more than that of generating its own isolates. This condition could lead to the trade-off of the new mutant isolates away from the original hosts. This could be the origin of the spreading of viral infection due to the generation of new mutant viral isolates. This claim can be correlated to the notion that since mutations in RNA virus is high [26, 27] causing harm to the hosts and hence, transmitted by diffusing from one host to another causing infection to other hosts [28]. Otherwise, the old viral isolates will stick to the old hosts ($\mu_2 < \mu_1$, \bar{x}^* is a stable fixed point since $\lambda < 0$) without diffusing from the hosts. Further, the characteristics time scale, which is the time required to determine significant change in \bar{x} around fixed point \bar{x}^* , is given by, $\tau = \frac{1}{\frac{dF(\bar{x}^*)}{d\bar{x}}} = \frac{1}{\mu[M-1]}$.

Lemma 2: Competition might drive the migration of virus isolates to different hosts.

Proof: The criteria of competitive or cooperative of any two mutant isolates modeled by the dynamical system, $\frac{dy_i}{dt} = F_i[y_i, y_j; t]$, is given by [50]: 1. The mutant isolates 'i' and 'j' are competitive in a region R if, $\frac{\partial F_i}{\partial y_j}(y) \geq 0$; $i \neq j \forall y \in R$; 2. The two mutant isolates are cooperative if, $\frac{\partial F_i}{\partial y_j}(y) < 0$; $i \neq j \forall y \in R$. Applying this criteria to the model

we considered (7), $\frac{dy_i}{dt} = F_i(y_i, y_j; t)$, we have,

$$\frac{\partial F_i}{\partial y_j} = \mu_1 \sum_{j=1}^{i-1} p_{ij} \delta_{ij} - \mu_2 \sum_{j=i+1}^M p_{ij} \delta_{ij} = \mu_1 p_{j;1 \leq j \leq i-1}^{[1]} - \mu_2 p_{j;i+1 \leq j \leq M}^{[2]} \quad (13)$$

such that, 'i' and 'j' mutant isolates competitive if $\mu_1 p_{j;1 \leq j \leq i-1}^{[1]} \geq \mu_2 p_{j;i+1 \leq j \leq M}^{[2]}$. This condition allows the two mutant isolates compete among themselves, and if the new mutant 'j' dominates over 'i', then there could be high chance of more virulence due to it's mutation and may trade-off from the present hosts to the others. However, if $\mu_1 p_{j;1 \leq j \leq i-1}^{[1]} < \mu_2 p_{j;i+1 \leq j \leq M}^{[2]}$, then the two mutant isolates will cooperate to each other and may confine to the present hosts. This condition could happen when the virulence of the two mutant isolate types are nearly similar due

to not much difference in the mutation profiles.

Conclusion

SARS-CoV-2 is a highly virulent RNA virus causing pandemic across the world. Some of the most important factors of the origin of this virulence could be due to high mutational events in the virus genome when hosts-virus interaction takes place, and the presence of multiple virus isolates in the host. The Covid-19 pandemic is all of a sudden first epicenter from Wuhan, China, and spread all over the world, killing nearly two lakh people across the world. The pandemic is so fast, no concrete theory of the widespread of this disease, and still no vaccine is discovered yet. In this work, we studied and analyzed a large number of publicly available SARS-CoV-2 genomes collected from patients from different parts of the world to explore the possibility of the origin of this pandemic. Our study of mutational profiles of SARS-CoV-2 isolates across the world show very high mutational rates in these isolates, which led the isolates more virulent, causing significant harm to the hosts and enhance viral transmission to hosts. The mutations in these isolates also take place mainly in six specific genes out of eleven genes, and the highest mutation takes place in the *orf1ab* gene and the mutation events follow the Markov process. It is also found from the mutational profile study that transition is much more frequent than transversion.

We, then, analyzed the SARS-CoV-2 isolates using the multifractal approach to understand the complexity in the isolates genome variability and structural organization. The calculated multifractal parameters reveal that structural variability in the isolates is caused by the presence of multiple isolates in the hosts. These isolates are highly asymmetric (left-skewed) in the singularity function, indicating the richness of complexity and dominance by large fluctuations in genome structure organization imparting virulence in the virus. The large values obtained for Hurst exponent and generalized dimension indicates heterogeneous genome structure organization and strong positive correlation in the organization of SARS-CoV-2 isolates. The genome structures of the isolates are far from equilibrium, self-organized, and quite sensitive to fluctuations in and around it. This could also be the origin of multiple isolates in a certain geographical area. The presence of multiple isolates in hosts may lead the Covid-19 pandemic.

We then study a simple model of hosts-virus interaction by taking into consideration multiple isolates in it and applied Price theory to understand the epidemic. The theory indicates that the fitness parameter of the isolates, which is dependent on various factors, namely virulence, fluctuations, etc., is quite sensitive and affected by fluctuations, multiple isolate interactions with the host, etc. However, Price theory needs to be investigated by taking into account various factors such as quarantine, exposed, health care, optimized economy, and many other factors.

Materials and Methods

The detailed methods of this study are given below.

Conversion of genomic sequence into DNA Walk

The whole genome of every isolate as well as the reference genome are converted in *DNA walk* $\{x_i; i = 1, 2, \dots, N\}$ by considering purine (A or G) as step up ($x_i = +1$) and pyrimidine (C or T) as step down ($x_i = -1$) [48]. The *DNA walk* is considered to be a non stationary time series data due to its stochastic behavior and is the input considered for various parameters. This DNA walk is considered as the profile, $Y(i) = \sum_{j=1}^i (x_j - \langle x \rangle)$, where, $\langle x \rangle$ is the mean, and $i = 1, 2, \dots, N$.

Multifractal Detrended Fluctuation Analysis (MF-DFA) The Multifractal Detrended Fluctuation Analysis (MF-DFA) is a powerful method for analysing and studying fractal characteristics in non-stationary time series data[34]. Using the numerical technique mentioned in Kantelhardt et al.[32, 33] various multifractal parameters that describe the fractal nature in the time series can be calculated such as generalized dimension(D_q), Hurst exponent(H_q), singularity spectrum($f(\alpha)$) etc. We calculate these parameters for all the three types of isolates. The procedure of MF-DFA has five steps which are summarized.

- Step 1. We consider the nonstationary time series data as the input to MF-DFA algorithm for calculating all the multifractal parameters. In the first step in MF-DFA the DNA walk time series signal x_j having length N is considered and profile representing the time series is calculated as below,

$$Y(i) = \sum_{j=1}^i (x_j - \langle x \rangle) \quad (14)$$

where $\langle x \rangle$ is the average value and $i = 1, 2, 3, \dots, N$.

- step 2. The profile $Y(i)$ is divided in non-overlapping segments of length (l) of same size as $N_s \equiv \text{int}(N/s)$. It is to note that as the length N of the time series is not usually the multiple of examined time scale s therefore at the tail part of profile some small part could be remaining. Keeping this in mind the same protocol is redone from the other end of the time series. Therefore $2N_s$ segments are calculated altogether.
- step 3. Using the least-square method of fitting calculate the local trend for every segment($2N_s$). We calculate the variance as follows,

$$F^2(s, \eta) = \frac{1}{s} \sum_{i=1}^s \{Y[(\eta - 1)s + i] - y_\eta(i)\}^2 \quad (15)$$

where for every segment $\eta, \eta = 1, 2, 3, \dots, N_s$ and

$$F^2(s, \eta) = \frac{1}{s} \sum_{i=1}^s \{Y[N - (\eta - N_s)s + i] - y_\eta(i)\}^2 \quad (16)$$

for $\eta = N_s + 1, \dots, 2N_s$ and $y_\eta(i)$ is the polynomial that will be fitted in the segment η .

- step 4. The fluctuation function of q^{th} order is estimated using the average of all the segments as,

$$F_q(s) = \left\{ \frac{1}{2N_s} \sum_{\eta=1}^{2N_s} [Y[(\eta, s)]^{q/2}] \right\}^{1/q} \quad (17)$$

where, q an index variable is any real valued number except zero[33].

- step 5. The scaling behaviour of fluctuation $F_q(s)$ is calculated after analysing the log-log plots ($F_q(s)$ with respect to s) for a single q value and if time-series is has long range power-law correlation then $F_q(s)$ will increase for larger values of s following power law as,

$$F_q(s) \sim s^{H_q} \quad (18)$$

The $H(q)$ is the generalised hurst exponent and in our analysis we have calculated the $H(q)$ by considering q from $q_{min} = -5$ to $q_{max} = +5$. The value of $H(q)$ changing with the change in value of q represents that our DNA walk data multifractal nature and it is to note that if varying the value of $H(q)$ with respect to q is constant then it indicates that time series is monofractal in nature.

The relationship between $H(q)$ and mass exponent is calculated for the various isolates as follows,

$$\tau(q) = qH_q - 1 \quad (19)$$

Another important multifractal parameter is the singularity spectrum for characterizing a multifractal time series. It is related to $\tau(q)$ with the Legendre transformation. α and $f(\alpha)$ are calculated as below,

$$\alpha = \frac{\tau(q)}{dq} \quad (20)$$

and

$$f(\alpha) = q\alpha - \tau(q) \quad (21)$$

Here, α is called the Holder exponent also called singularity strength and $f(\alpha)$ is the singularity spectrum and can be simply expressed in the form of q and $H(q)$ as the following equation[33].

$$\alpha = H(q) + qH'(q) \text{ and } f(\alpha) = q[\alpha - H(q)] + 1 \quad (22)$$

The generalised multifractal dimension(D_q) is another multifractal parameter and is defined as follows[33],

$$D_q \equiv \frac{\tau(q)}{q-1} = \frac{qH(q)-1}{q-1} \quad (23)$$

here, $\tau(q)$ is Mass exponent and $H(q)$ is Hurst exponent. where, The $f(\alpha)$ is the subset dimension of the series characterised by α whereas the α is called Holder exponent or singularity strength. We performed the alignment of all the genomes and find out the point mutation present in all the isolates with respect to the reference Wuhan genome and calculated the various multifractal parameters.

Acknowledgement

SM (Project ID: 2019-5116, File No: ISRM/11(16)/2019) is financially supported by Indian Council of Medical Research(ICMR). MZM financially supported by Department of Health and Research, Ministry of Health and Family Welfare, Government of India under young scientist FTS No. 3146887. R.K.B.S. acknowledges UPE-II, sanction no. 101, India, for providing financial support.

Author Contributions

RKB and MZM conceptualized and designed the model. SM, RKSS, SKS, MZM and RKBS did the computational experiment and prepared the figures from the results. SM, RKSS, SKS, MZM and RKBS wrote the manuscript. All authors read, checked and approved the manuscript.

Competing interests

The authors declare that they have no competing interests.

-
- [1] Zhu N, Zhang D, Wang W, et al. A novel coronavirus from patients with pneumonia in China, 2019. *N Engl. J. Med.*, **382**, 727-733 (2020).
 - [2] Huang C, Wang Y, Li X, et al. Clinical features of patients infected with 2019 novel coronavirus in Wuhan, China. *Lancet*, **395**, 497-506 (2020).
 - [3] Sohrabi, C., Alsafi, Z., O'Neill, N., Khan, M., Kerwan, A., Al-Jabir, A., Iosifidis, C., and Agha, R. World Health Organization declares global emergency: A review of the 2019 novel coronavirus (COVID-19). *Int. J. Surgery*, **76**, 71-76 (2020).
 - [4] Fehr AR and Perlman S. Coronaviruses: an overview of their replication and pathogenesis. *Methods Mol Biol.*, **1282**, 1-23 (2015).
 - [5] Cui, J., Li, F. and Shi, Z.-L. Origin and evolution of pathogenic coronaviruses. *Nat. Rev. Microbiol.*, **17**, 181192 (2019).
 - [6] Corman, Victor M., Doreen Muth, Daniela Niemeyer, and Christian Drosten. Hosts and sources of endemic human coronaviruses. *In Advances in virus research*, **100**, 163-188. (2018).
 - [7] Andersen KG, Rambaut A, Lipkin WI, Holmes EC and Garry RF. The proximal origin of SARS-CoV-2. *Nature Medicine*, **26**, 450-452 (2020).
 - [8] Shi, M., Lin, X. D., Tian, J. H., Chen, L. J., Chen, X., Li, C. X. & Buchmann, J. Redefining the invertebrate RNA virosphere. *Nature*, **540(7634)**, 539-543 (2016).
 - [9] Zhou, P., Yang, X. L., Wang, X. G., Hu, B., Zhang, L., Zhang, W. & Chen, H. D. A pneumonia outbreak associated with a new coronavirus of probable bat origin. *Nature*, **579(7798)**, 270-273 (2020).
 - [10] N.Chen, M. Zhou, X. Dong, et.al., Epidemiological and clinical characteristics of 99 cases of 2019 novel coronavirus pneumonia in Wuhan, China: a descriptive study, *Lancet*, **395**, 507-513 (2020).
 - [11] Wan Y, Shang J, Graham R, Baric RS, Li F. Receptor recognition by novel coronavirus from Wuhan: An analysis based on decade-long structural studies of SARS. *J Virology*, **94**, e00127-20 (2020).
 - [12] Li XC, Zhang J, Zhuo JL. The vasoprotective axes of the renin-angiotensin system: physiological relevance and therapeutic implications in cardiovascular, hypertensive and kidney diseases. *Pharmacol Res.*, **125**, 2138 (2017).
 - [13] Fang L, Karakiulakis G, and Roth M. Are patients with hypertension and diabetes mellitus at increased risk for COVID-19 infection? *Lancet Respir. Med.*, **8**, e21 (2020).
 - [14] Lunn, P. et al. Using behavioural science to help fight the coronavirus. *J. Behavio. Pub. Admin.*, **3**, (2020).
 - [15] Khachfe HH, Chahrour MA, Sammour J, et al.: An epidemiological study on COVID-19: a rapidly spreading disease. *Cureus.*, **12**, e7313 (2020).
 - [16] Wu, F. et al. A new coronavirus associated with human respiratory disease in China. *Nature*, **579**, 265269 (2020).
 - [17] Thomas Malthus, An Essay on the Principle of Population. *New York: Penguin*, (1798).
 - [18] Roser, M., H. Ritchie, and E. Ortiz-Ospina (2020). Coronavirus Disease (COVID-19) Statistics and Research, Our World in Data at <https://ourworldindata.org/coronavirus>.
 - [19] Ludwig, Wilhelm, "Vergleichende Untersuchung uber Wachstumsgesetze," *Biol. Zentralbl.*, **49**, 735-758 (1929).
 - [20] Gompertz, Benjamin. On the Nature of the Function Expressive of the Law of Human Mortality, and on a New Method of Determining the Value of Life Contingencies. *Phil. Trans. Roy. Soc.*, 513-585 (1825).
 - [21] Winsor, C.P. The Gompertz curve as a growth curve. *Proc. Natl. Acad. Sc. Proc. Nat. Acad. Sc.* **8**, 1-7 (1932).

- [22] O. Kermack and A.G. McKendrick. Contributions to the mathematical theory of epidemics, part 1, *Proc. Roy. Soc. London Ser. A*, 115 700-721 (1927).
- [23] Tellmann, U. Catastrophic populations and the fear of the future: Malthus and the genealogy of liberal economy. *Theory, Culture & Society*, **30(2)**, 135155 (2013).
- [24] Duffy S. Why are RNA virus mutation rates so damn high? *PLoS Biol.*, **16**, e3000003 (2018).
- [25] Baer CF. Does mutation rate depend on itself. *PLoS Biol.*, **6**, e52 (2008).
- [26] Loewe L, Hill WG. The populations of mutations: good, bad and indifferent. *Phil. trans. Royal Soc. London*, **365**, 115367 (2010).
- [27] Vignuzzi, M., Stone, J.K., Arnold, J.J., Cameron, C.E., and Andino, R. Quasispecies diversity determines pathogenesis through cooperative interactions in a viral population. *Nature*, **439**, 344348 (2006).
- [28] Fitzsimmons W, Woods RJ, McCrone JT, Woodman A, Arnold JJ, Yennawar M, et al. A speed-fidelity trade-off determines the mutation rate and virulence of an RNA virus. *PLoS Biol.*, **16(6)**, e2006459 (2018).
- [29] C. Gardiner, *Stochastic methods: a handbook for the natural and social sciences* 4th ed.(2009). Springer.
- [30] Collins, D. W. & Jukes, T. H. Rates of transition and transversion in coding sequences since the human-rodent divergence. *Genomics*, **20(3)**, 386-396 (1994).
- [31] Jonassen, C. M., Jonassen, T. O. & Grinde, B. A common RNA motif in the 3' end of the genomes of astroviruses, avian infectious bronchitis virus and an equine rhinovirus.
- [32] Mandal, S., Roychowdhury, T., Chirom, K., Bhattacharya, A. & Singh, R. B. Complex multifractal nature in *Mycobacterium tuberculosis* genome. *Scientific reports*, **7(1)**, 1-13(2017).
- [33] Kantelhardt, J. W., Zschiegner, S. A., Koscielny-Bunde, E., Havlin, S., Bunde, A., & Stanley, H. E. Multifractal detrended fluctuation analysis of nonstationary time series. *Physica A: Statistical Mechanics and its Applications*, **316(1-4)**, 87-114 (2002).
- [34] Halsey, T. C., Jensen, M. H., Kadanoff, L. P., Procaccia, I. & Shraiman, B. I. Fractal measures and their singularities: The characterization of strange sets. *Physical Review A*, **33(2)**, 1141(1986).
- [35] Calvert, L., Fisher, A. and Mandelbrot, B. The Multifractal Model of Asset Returns. Discussion papers of the Cowles Foundation for Economics. *Yale University: Cowles Foundation*, **1141166**, 1227 (1997).
- [36] Darko, S., Dusan, S., Tatijana, S. and Stanley, H. E. Multifractal analysis of managed and independent float exchange rates. *Physica A*, **428**, 1318 (2015).
- [37] Pepin KM, Lambeth K, Hanley KA. Asymmetric competitive suppression between isolates of dengue virus. *BMC Microbiol.*, **8**, 28 (2008).
- [38] Gomez, P., Sempere, R. N., Elena, S. F., and Aranda, M. A. Mixed infections of Pepino mosaic virus isolates modulate the evolutionary dynamics of this emergent virus. *J. Virol.*, **83**, 12378-12387 (2009).
- [39] Mallat, S. and Hwang, W. L. Singularity detection and processing with wavelets. *IEEE Trans. Infor. Theor.*, **28**, 617 (1992).
- [40] P. Norouzzadeh, B. Rahmani, A multifractal detrended fluctuation description of Iranian rial-US dollar exchange rate. *Physica A*, **367**, 328336 (2006).
- [41] Heylighen, F. The science of self-organization and adaptivity. *The encyclopedia of life support systems*, **5**, 253280 (2001).
- [42] Frank S A. Models of parasite virulence. *Quarterly Review of Biology*, **71(1)**, 37-78 (1996).
- [43] Frank S A. A kin selection model for the evolution of virulence. *Proc. Royal Soc. London B*, **250**, 195-197 (1992).
- [44] Day T, Parsons T, Lambert A, Gandon S. The Price equation and evolutionary epidemiology. *Phil. Trans. R. Soc. B*, **375**, 20190357 (2020).
- [45] Price GR. Selection and covariance. *Nature* **227**, 520521 (1970).
- [46] Bonhoeffer S, Nowak M A. Mutation and the evolution of virulence. *Proc. Royal Soc. London*, **258**, 133-140 (1994).
- [47] H. W. Hethcote, The mathematics of infectious diseases. *SIAM Rev.* **42** 599-653 (2000).
- [48] Peng, C. K., Buldyrev, S. V., Goldberger, A. L., Havlin, S., Sciortino, F., Simons, M. & Stanley, H. E. Long-range correlations in nucleotide sequences. *Nature*, **356(6365)**, 168-170(1992).
- [49] S. H. Strogatz. *Nonlinear Dynamics and Chaos: with Applications in Physics, Biology, Chemistry, and Engineering*. Addison-Wesley, Reading, MA, (1994).
- [50] M. Hirsch. Systems of differential equations which are competitive or cooperative, I: Limit sets, *Siam J. Math. Anal.* **13**, 169-179 (1982).

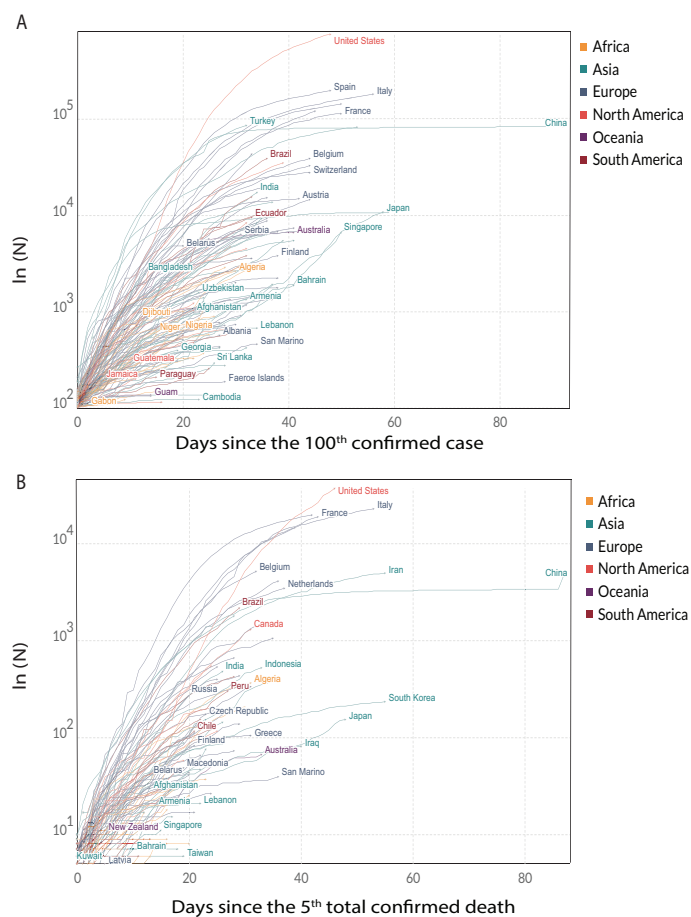


FIG. 1: **Dynamics of Covid-19 pandemic:** The dynamics of country wise confirmed (A) infected and (B) death cases across the world. The extraction of the data is from the database [18].

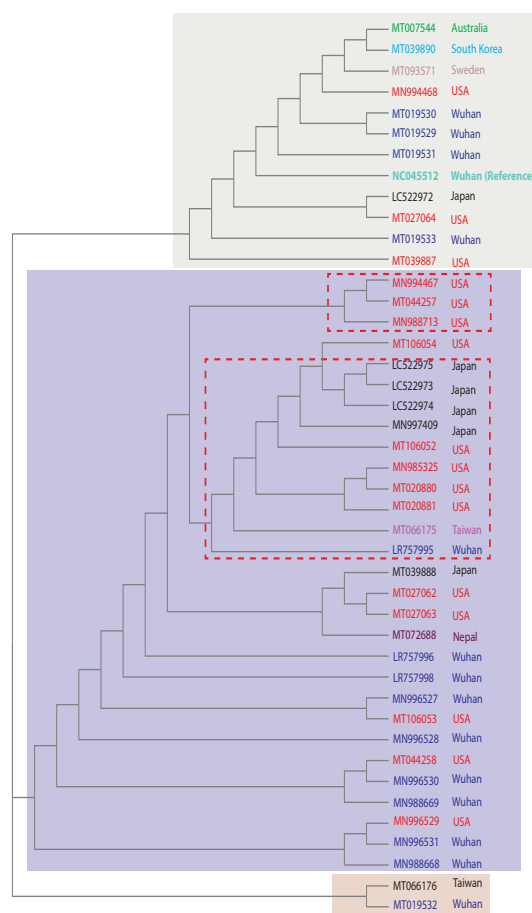


FIG. 2: **Phylogenetic tree of SARS-CoV-2 isolates:** Clustering of SARS-CoV-2 virus isolates collected from the Covid-19 patients samples from eight different countries across the world.

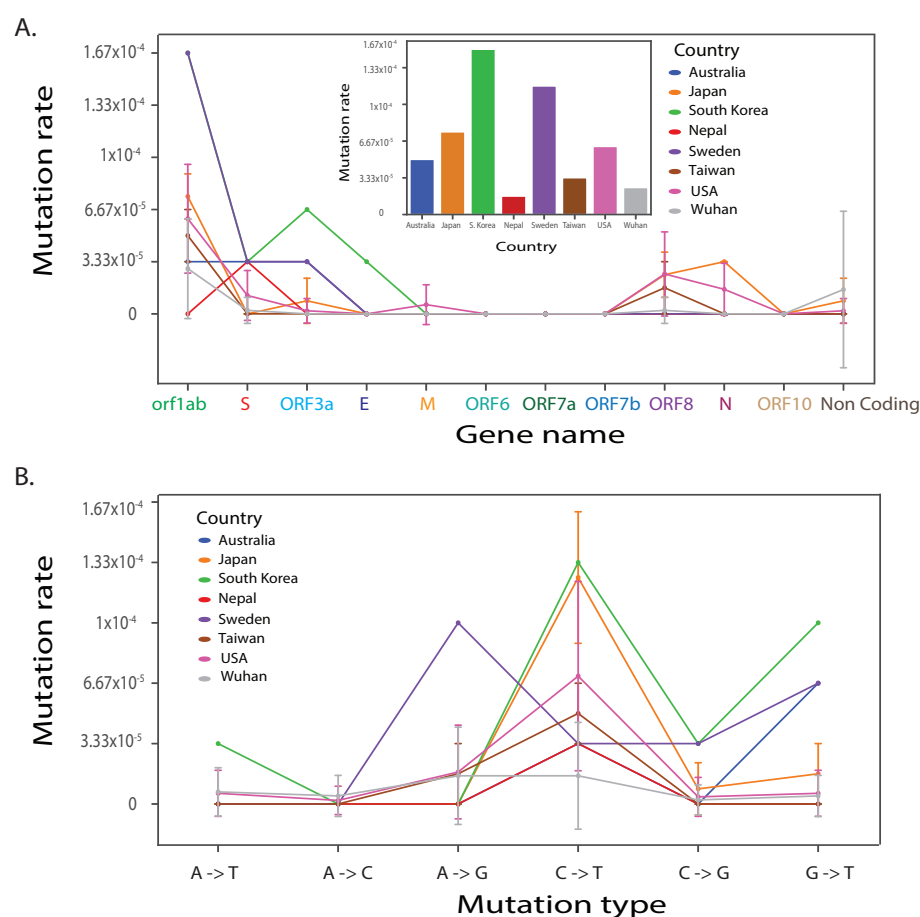


FIG. 3: **Mutation profiles of the SARS-CoV-2 isolates:** (A) Plots of mutation rates verses gene name of SARS-CoV-2. (B) Plots of mutation rate with respect to mutation types.

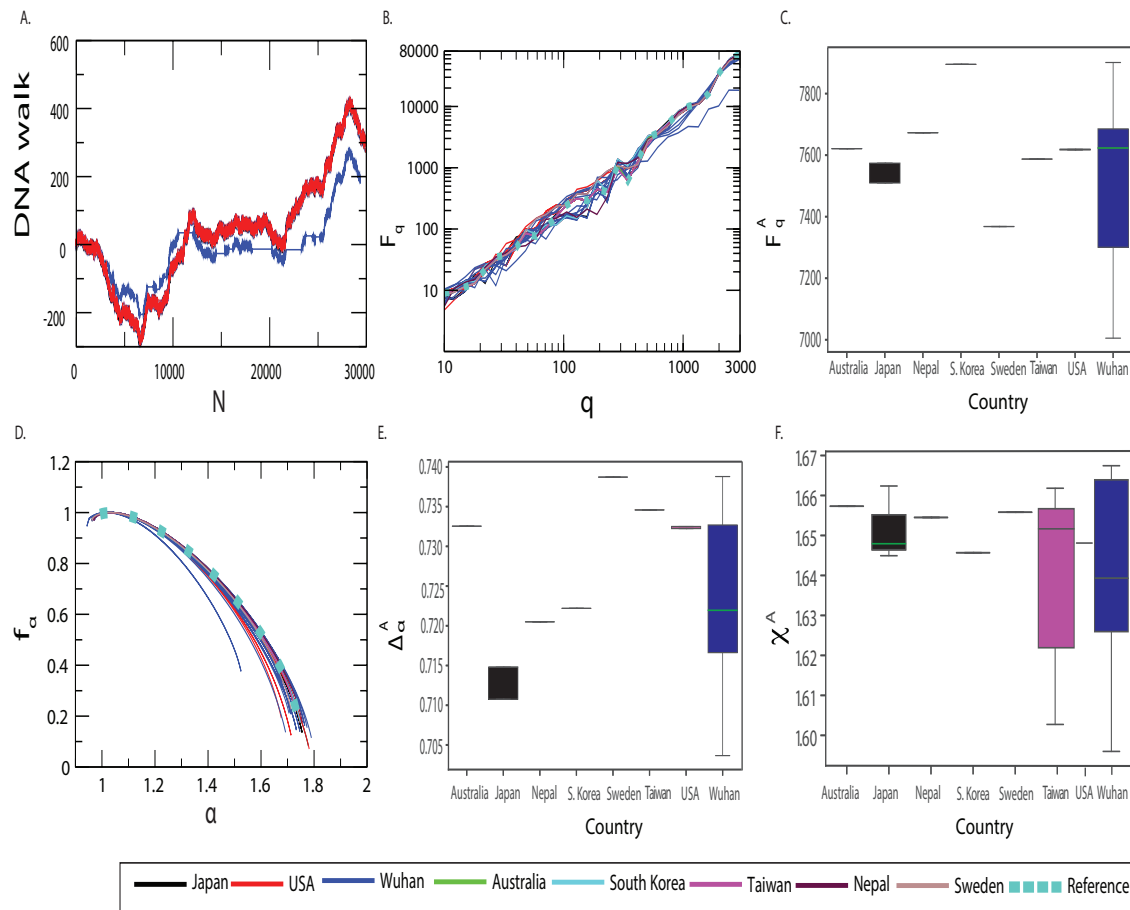


FIG. 4: **Multifractal parameters of the SARS-CoV-2 isolates:** Plots of (A) DNA walks of all the 43 genomic isolates of SARS-CoV-2 where reference genome is shown as dotted light blue. (B) F_q versus q in log-log plot, (C) averaged F_q with respect to respective countries across the world, (D) f_α as a function of α with dotted line being the reference, (E) $\Delta\alpha$ for various countries, and, (F) averaged χ^A for different countries.

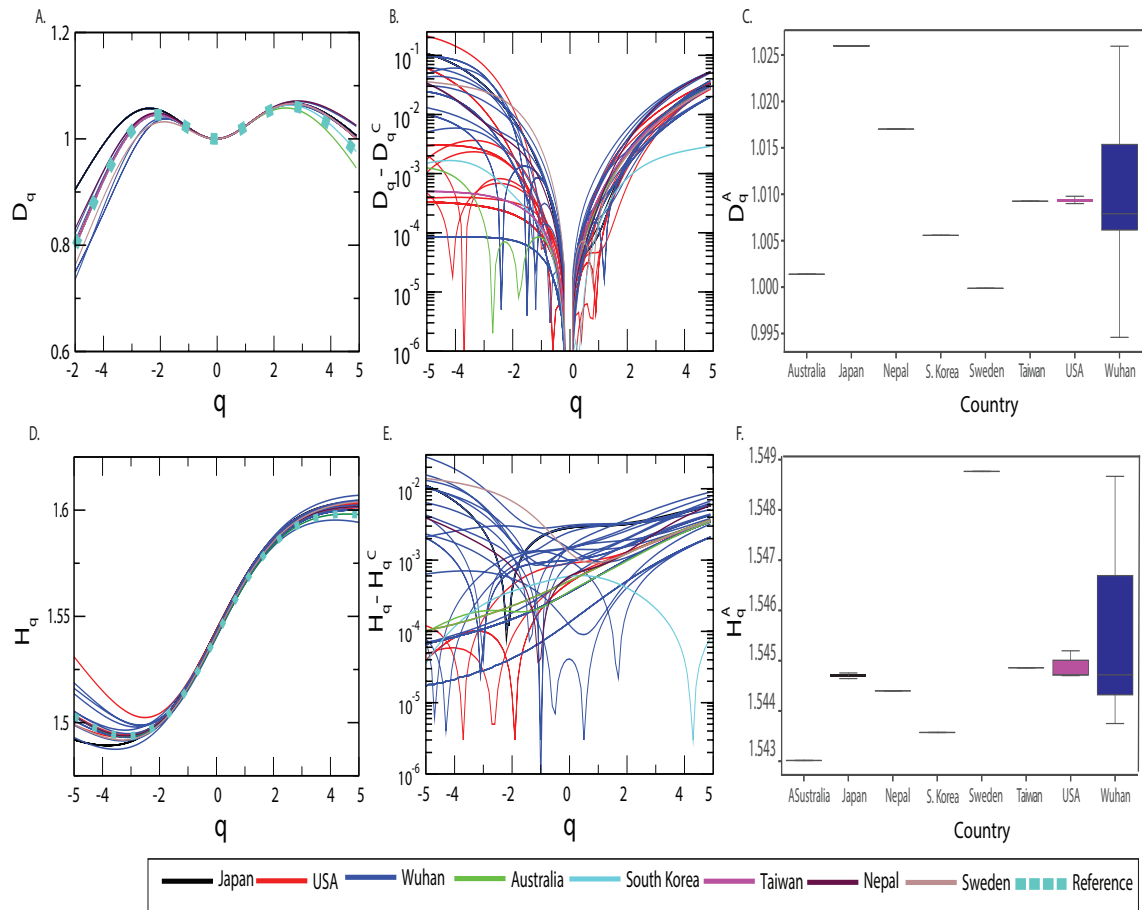


FIG. 5: **Nature of D_q and H_q of the SARS-CoV-2 isolates:** Plots of (A) D_q versus q in log-log plot for 43 genomic isolates of SARS-CoV-2 where reference genome is shown as dotted light blue, (B) averaged D_q versus q in log log plot, (C) averaged D_q as a function of respective country, (D) H_q with respect to q with dotted line being the reference, (E) $H_q - H_q^C$ versus q , and, (F) averaged H_q for the countries we have taken.

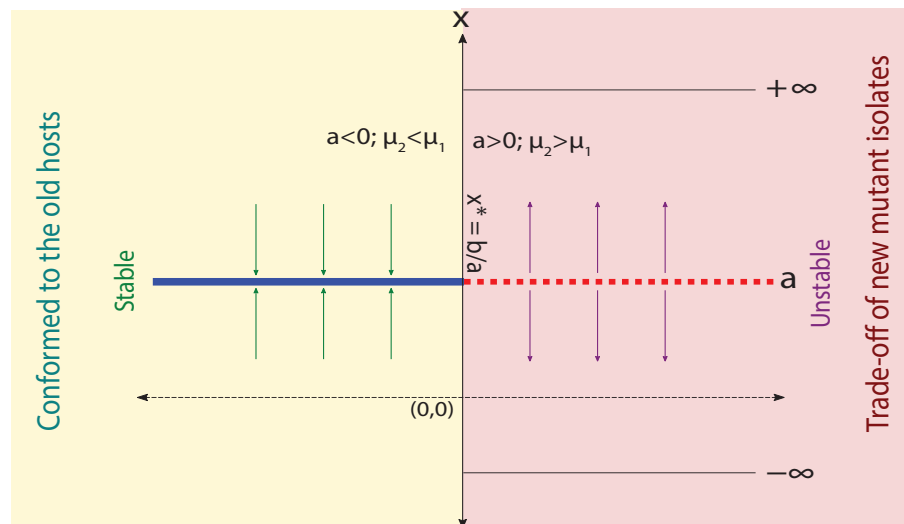


FIG. 6: **Bifurcation diagram of the virus-hosts interaction model.**