The Germ Theory of COVID-19 Pandemic and Similarity Analysis of Genomics – A Machine Learning Approach

K. M. Shivakumar[1], M. Kalpana[2*] and C. S. Sumathi[2]

[1]*Faculty of Agricultural Economics, Centre for Agricultural and Rural Development Studies, Tamil Nadu Agricultural University, Coimbatore, Tamil Nadu, India,*
[2]*Faculty of Computer Science, Anbil Dharmalingam Agricultural College and Research Institute, Tiruchirappalli, Tamil Nadu, India*

*Corresponding author

A B S T R A C T

The COVID - 19 pandemic creates havoc all around the world for human beings impacting their health, livelihood, employment, processing, economy etc. All the countries take various efforts like social distancing, containment, self- quarantine, severe testing but the spread of virus is taken place at an alarming rate. This article just provides the scenario of spread of pandemic COVID 19 in the most affected top 10 countries and also provides the similarity analysis of the RNA sequence of the virus affected victims in some of these countries to help them the causal agent and their genetic makeup in taking up preventive and quarantine measures.

Introduction

Around the world, the novel acute respiratory syndrome coronavirus 2 spreads panic causing severe health hazards and mortality. Governments, policy makers, frontline health and sanitation officials and common public all are united against battle with this pandemic virus. Efforts are taken to combat the magnitude of the problem with several preventive mechanisms such as wearing masks, sanitation procedures, social distancing, quarantine measures, containment etc., as strategies. But the spread of virus is increasing day by day causing colossal damage to the human community. Pharmaceutical companies are trying to find a new vaccine to provide solution for this dreadful infection of this Corona virus. The pattern and the spread of the disease is still an enigma as both the symptomatic and asymptomatic carriers of the virus are the major source for infection which further deepens the crisis.

The increased growth in population with constraints such as food, space etc. The spread of disease can be modeled with logistic equation. The differential equation was introduced by Verhulst, a demographic mathematical model which is used for the development of number of members in biological population like bacteria, animals, birds, humans etc.,(1). Many literatures regarding the epidemic growth of the novel coronavirus disease (COVID-19) uses statistical analysis and model to give the future scenarios. Simple exponential growth models are used to known about the growth process of coronavirus (2-5). The logistic response model is used in decision making process for infectious disease. The number of patients at different stages and the medical recourses required for certain period are estimated by the model. The factors such as number of deaths are also evaluated by the model(6). The deaths associated with COVID-19 pandemic greatly outnumbered other two deadly races of coronaviruses i.e., SARS-CoV and MERS-CoV, which caused a huge threat to the global public health and economics(7). Coronavirus disease is a dreadful infectious disease caused by a virus of the Beta coronavirus family named severe acute respiratory syndrome coronavirus 2, previously referred to as 2019 novel coronavirus (2019-nCoV)(8,9). Till date, no approved vaccines or specific antivirals are available for COVID-19(10, 11). Previous SARS pandemic was effectively controlled and managed by conventional preventive measures, including travel restrictions and patient isolation. Currently, these measures are applied in around the world with the COVID-19 outbreak; however, their effectiveness depends on how rigorous they are managed(12, 13). It follows that the methods enabling reliable prediction of spreading of COVID-19 would be of great benefit in persuading public opinion why it is crucial to adhere to these measures in the past

decade(14). Modeling pandemic such as COVID-19 is extremely important in determining their possible future consequences with proper modeling the spread and the effect of such a disease can be supremely important in developing a new vaccine (15).

Though the efforts are taken to contain the spread through periodical lockdowns and constant monitoring of the containment zones, countries are struggling to find a solution for the trade - off between protecting health Vs creating livelihood opportunities. Most of the population is wage earning migrant labourers who are most vulnerable both economical and health aspects. They don't have access to medicines and health care provisions as loss of employment, loss of income, curbing the movement and huge dependent family make the tasks very critical. Further, this pandemic creates social and psychological problems during the lockdown periods which aggravate the problem still worse. The borders have been secured across the provinces, districts, states and countries to curb the movements of public, goods and services depending upon the severity of the infections. As usual, the spread of infections has been modelled all around the world using logistic and exponential growth functions as the initial phase of any infection has exponential growth and as time passes, the preventive measures will reduce the growth of infection. The later stage of the infection rate is modelled by the logistic growth model as the preventive measures would weaken the infection rate. In order to assess the nature and spread of COVID virus across countries, similarity analysis has been carried out using the RNA sequence analysis of the samples extracted from the victims belonging to some of the most affected nations. Such a study would help us to find appropriate preventive solution for the races of the virus.

Materials and Methods

Growth rate of any infection has been modelled based on the exponential or the logistic growth model by the epidemiologists in most of the cases. It is customary practice to model the growth of infection by the exponential model as the rate of growth will be ever increasing in the initial stages as there would be no vaccine to prevent it in the beginning. It is a mathematical function that can be adopted during the initial phases of the outbreaks of pandemic/epidemic but it is not representing the perfect situation as the preventive methods and cured people would not spread virus anymore and infection rate would come to halt at a late stage.

In most of the cases, epidemiologists use logistic growth models for the entire period of the infection. It is a mathematical function tracing the growth trajectory of the infection rate from early surge and went on increasing to a certain stage and become flat at the later stage because of the intervention of the vaccines and health care provisions. It is represented by

y (t) = c / (1+a e^–bt)

Where,

y(t): number of infected cases at given time 't'

c: limiting value

b>0

Number of cases at the beginning will take values c/ (1+a)

Maximum growth rate at 't' is found at ln(a)/b or y(t) =c/2

dy/dt = ry. (1-y/c)

If y=c, growth will be 0.

It has been found that by forecasting the predicted vs actual cases, the maximum growth occurred in a day, doubling period, peak period growth will be found out for the most affected countries. Figure 1 shows the COVID 19 infected cases of top 10 worst affected countries of the World.

Using Scipy non- linear least square optimization values of a can be estimated to minimize least square errors for the model. For the initial values of a, b and c, random exponential for the reference Wuhan centre, China has been taken up.

For assessing similarity between the RNA sample extracted from the COVID 19 victims, text based format of protein sequence namely FASTA files have been collected from NCBI, WHO, Kaggle and analyzed using Biopython.

Results and Discussion

The top 10 COVID -19 infected countries of the world has been obtained from the WHO database June, 2020. The initial loading values for the parameters a, b and c has been modelled based on the exponential growth of the infections happened at the epicentre of this pandemic - Wuhan centre of the Republic of China. Assuming the spread of virus in the initial phase of the epidemic is exponential, the parameters for a, b and c for the major affected countries have been modelled. As defined earlier, c is the maximum carrying capacity of the infections and t is the maximum growth rate. Parameter y is the maximum growth of infections happening at given time 't'. From Table 1, we can infer that Spain and Italy are the countries where the maximum growth rate of infections occurred from the date of first infection recorded in the respective countries. Because of severe containment measures taken in India, it has

594

taken much time to attain the maximum growth rate when compared to other major infected countries. Growth rate of infection for the remaining countries tread in between these upper and lower bounds pegged by the European countries and India. But the maximum infected cases at the growth rate t was found higher for USA and Brazil as the severity of infections and spread of virus was very rapid in these two countries followed by India. The fastest growth moments and doubling period for the most affected countries were also modelled out and presented in Table 2. As discussed in the previous section, the maximum number of infections at peak growth time 't' happened for USA, Brazil and India.

**Table.1** Parameter Value for COVID 19 infections of major countries

Country	A	B	C	T	Y
India	4458.65	0.0549	1138403	152.80	569201
Iran	24.70	0.0347	269577	92.35	134788
Italy	322.43	0.0886	233445	65.18	116722
Chile	1267.18	0.0661	420592	108.01	210296
Brazil	1203.60	0.0621	2031398	114.10	1015699
USA	308.21	0.0528	2478634	108.38	1239312
UK	752.66	0.0755	302012	87.66	151006
Spain	2051.36	0.0578	237208	64.71	118604
Russia	2663.60	0.0690	662036	114.26	331018
Peru	344.71	0.0673	316997	86.70	158498

Table.2 Fastest growth moments for COVID 19 infections of major countries

Country	FGM	DP	Time
India	569201	133	20
Iran	134785	53	80
Italy	116722	52	100
Chile	210296	91	29
Brazil	1015699	95	31
USA	1239312	84	77
UK	151006	71	81
Spain	118604	55	96
Russia	331018	97	55
Peru	158498	70	47

Table.3 Similarity analysis of RNA sequence of major affected countries

Country Pair	Similarity of RNA (%)
China -India	99.7789
China-USA	99.9430
China-Iran	99.9497
China-Italy	99.9497
India-USA	99.9765
India-Iran	99.9765
India-Italy	99.9765
USA-Iran	99.9989
USA-Italy	99.9989
Iran-Italy	99.9989

Fig. 1 COVID -19 Infected Cases of Top 10 Worst Affected Countries of the World
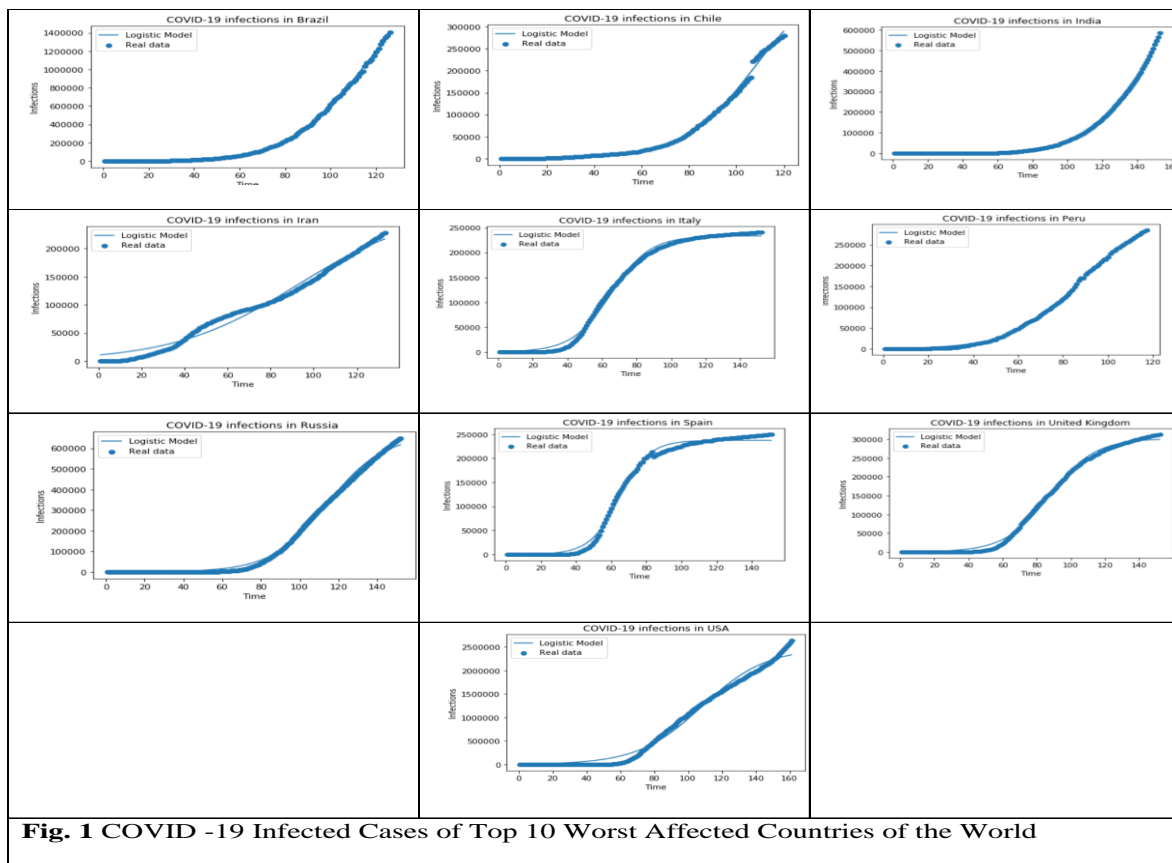
But the doubling rate of inspection was very faster for Italy and Iran as just within 52 and 53 days of first infection recorded respectively, the COVID-19 infected cases have been doubled sooner. Though India has taken much larger time for the peak infections i.e., 133rd day from the date of first infection,

the doubling period was very fast such that within just 20 days, the infected cases would be doubled as per the logistic growth model built for this pandemic followed by Chile and Brazil with 29 and 31 days required for the doubling of infections. For remaining countries, after the fastest growth moments

happened, periodic lock- downs coupled with severe testing procedures contained the doubling rate for most of the European countries.

In order to find out the similarity between the COVID -19 virus infected casualty in the worst affected countries, the protein sequence analysis of the RNA extracted from the diseased human from various countries have been analyzed and tested for similarity. The country pair of each these analyses were presented in Table 3.

The similarity of RNA per cent revealed for all the country pairs indicated that these countries are affected by the same sub type of this pandemic virus and hence a common vaccine should be the cure for all these countries. But, a detailed study with more number of RNA analyses would be better to find out the scope for any such mutations that deviate from the existing race which created havoc around the world.

From the results obtained from the logistic model of modelling the growth of COVID 19 germ, the maximum number of infections, the peak growth rate, doubling period, fastest growth moment, countries like India, Brazil and Chile should take rapid steps to contain the pandemic as the time taken for the doubling the peak rates are very quick just a matter of month.

The similarity of RNA analysis revealed that for all the country pairs tested the behavioral pattern of amino acids are almost identical which would lead a global vaccine development program to address this pandemic at the earliest.

References

1 Boyce W.E. and DiPrima R.C. Elementary Differential Equations and Boundary Value Problems. Von Hoffmann Press.2001.

2 Tuite AR, Fisman DN. Reporting, Epidemic Growth, and Reproduction Numbers for the 2019 Novel Coronavirus (2019-nCoV) Epidemic. Ann Intern Med. 2020; 172(8):567-568. https://www.acpjournals.org/doi/10.7326/M20-0358

3 Zhao S, Cao P, Gao D, Zhuang Z, Chong MKC, Cai Y. Epidemic growth and reproduction number for the novel coronavirus disease (COVID-19) outbreak on the Diamond Princess cruise ship from January 20 to February 19, 2020 : A preliminary data driven analysis. SSRN. 2020. https://doi.org/10.1101/2020.02.26.20028449

4 You C, Deng Y, Hu W, Sun J, Lin Q, Zhou F, *et al.,* Estimation of the Time-Varying Reproduction Number of COVID-19 Outbreak in China. International Journal of Hygiene and Environmental Health. 2020. https://doi.org/10.1101/2020.02.08.20021253

5 Zhang S, Diao M, Yu W, Pei L, Lin Z, Chen D. Estimation of the reproductive number of Novel Coronavirus (COVID-19) and the probable outbreak size on the Diamond Princess cruise ship: A data-driven analysis. Int J Infect Dis. 2020; 93:201-204. https://doi.org/10.1016/j.ijid.2020.02.033

6 Chen W., Gurnet A.,Ruiz A. Modelling the Logistics Response to a General Infectious Disease, IFAC-PapersOnLine. 2015; 48(3):180-186. https://doi.org/10.1016/j.ifacol.2015.06.078

7 Wu P, Hao X, Lau EHY, Wong JY, Leung KSM, Wu JT, *et al.,* Real-time tentative assessment of the epidemiological characteristics of novel coronavirus infections in Wuhan, China, as at 22 January 2020. Eurosurveillance. 2020; 25.

8   Liu Y., Gayle A. A., Wilder-Smith A and Rocklov J. The reproductive number of COVID-19 is higher compared to SARS coronavirus, Journal of Travel Medicine. 2020; 27(2): 1-4. https://doi.org/10.1093/jtm/taaa021

9   Center for Disease Control and Prevention (4th March 2020), Coronavirus Di s e a s e 2019 (COVID-19) - How It Spreads.https://www.cdc.gov/coronavirus/2019-ncov/prepare/transmission.html.

10 Velavan T. P. and Meyer C. G. The COVID-19 epidemic, Tropical Medicine & International Health. 2020; 25(3): 278–280. https://doi.org/10.1111/tmi.13383

11 Letko M., Marzi A., and Munster V. Functional assessment of cell entry and receptor usage for SARS-CoV-2 and other lineage B betacoronaviruses, Nature Microbiology. 2020; 5(4): 562–569.

https://www.nature.com/articles/s41564-020-0688-y

12 World Health Organization. Coronavirus disease 2019 (covid-19) Situation Report–29. 2020.

13 World Health Organization. Report of the WHO-China Joint Mission on Coronavirus Disease 2019 (COVID-19). 2020.

14 Hui D. S., Azhar E. I, Madani T. A. *et al.,* The continuing2019-nCoV epidemic threat of novel coronaviruses to global health — The latest 2019 novel coronavirus outbreak in Wuhan, China, International Journal of Infectious Diseases. 2020; 91: 264–266. https://doi.org/10.1016/j.ijid.2020.01.009

15 Vynnycky E. and White R. An Introduction to Infectious Disease Modelling, OUP oxford, 2010.