The National Academy of Sciences, India (NASI)

5, Lajpatrai Road, Prayagraj-211002, India

Monthly Summary for the month of June 2020

1. NASI office became fully operational with 100% staff strength, from 1st June 2020, maintaining all the safety measures as per guidelines issued by the GoI to combat COVID-19; also a **brief report on the activities organised by NASI during the lockdown period**, was uploaded on its website, starting with the domain - http://nasi.nic.in/NASI-Report%20on%20March%202020%20Activities-merged.pdf; and onward.

2. Several **Scientific Articles** written/communicated by the NASI Fellows/Members during this period, were distributed among the Fellows/Other Institutions, for the dissemination of information/knowledge on the pandemic COVID-19 (please see a few articles attached herewith as **Annex. 1 a, b, c, d & e**).

3. **World Environment Day (WED-2020)** was celebrated at NASI-HQ, on 5th June, with Social/Physical distancing; it was addressed by Prof. Satya Deo, G.S., NASI, Prof. U C Srivastava, Emeritus Professor at Prayagraj, also telephonically by Prof. Manju Sharma, Past President, NASI and many others. Several other Chapters also celebrated the WED-2020 by organinsing WEBINARS, attended by thousands of participants from many parts of the country. A few clippings/notices/reports are attached as **Annex. 2 'a', 'b' & 'c'**).

4. NASI Delhi Chapter collaborated/organised '**WEBINARS/Online Workshops**', in joint collaboration with several institutions for the UG/PG students and researchers, with kind cooperation/support of its Fellows/Members. The topic of the Webinar (please see the brief programme, as **Annex. 3**) held on June 15-17, 2020 was **"Century of Quantum Mechanics and Still Going Strong"** and the eminent speakers were Prof. Parimal H Vyas, Hon'ble Vice Chancellor, MSUB, Baroda, Prof. S. Lokanathan, Former Professor of Rajasthan University, Prof. Ajoy Ghatak, Former Professor, IIT Delhi, and others. They spoke on Birth of Modern Physics, Evolution of Quantum Theory, Uncertainty Principle, etc. A series of **INTERNATIONAL WEBINARS** has also been started from June 24, 2020 on different topics of Physical Sciences, covering wide areas of Physics and Chemistry, involving the scientists from different countries (please see **Annex. 4**).

5. International Yoga Day, was celebrated by performing 'Yoga at home' on a theme - "Yoga at Home and Yoga with Family."

Several other NASI Chapters also organized online scientific activities.

IJMR_795_20

Indian J Med Res 151, February 2020, pp **-**
DOI: 10.4103/ijmr.IJMR_795_20

Quick Response Code:

Detection of coronaviruses in *Pteropus & Rousettus* species of bats from different states of India

Pragya D. Yadav[1], Anita Shete-Aich[1], Dimpal A. Nyayanit[1], Prachi Pardeshi[1], Triparna Majumdar[1], R. Balasubramaniyam[1,7], Ullas Padinjaremattathil Thankappan[2], Sreelekshmy Mohandas[1], Hitesh Dighe[1], Pradeep Sawant[3], Savita Patil[1], Dilip Patil[4], M.D. Gokhale[5], Basavaraj Mathapati[6], A.B. Sudeep[5], Sreekant Baradkar[1], Abhimanyu Kumar[1], Rutuja Kharde[1], Malvika Salve[1], Yash Joshi[1], Nivedita Gupta[8] & Devendra T. Mourya[†]

[1]Maximum Containment Laboratory, [2]Diagnostic Virology Group, [3]Enteric Virus Group, [4]Animal house, [5]Entomology Group, [6]Poliovirus Group, [†]ICMR-National Institute of Virology, Pune, Maharashtra, [7]ICMR-National Institute of Virology Kerala Unit, Alappuzha, Kerala & [8]Division of Epidemiology & Communicable Diseases, Indian Council of Medical Research, New Delhi, India

Received March 26, 2020

***Background & objectives*: Bats are considered to be a natural reservoir for many viruses, of which some are potential human pathogens. In India, an association of *Pteropus medius* bats with the Nipah virus was reported in the past. It is suspected that the recently emerged severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) also has its association with bats. To assess the presence of CoVs in bats, we performed identification and characterization of bat CoV (BtCoV) in *P. medius* and *Rousettus* species from representative States in India, collected during 2018 and 2019.**

***Methods*: Representative rectal swab (RS) and throat swab specimens of *Pteropus* and *Rousettus* spp. bats were screened for CoVs using a pan-CoV reverse transcription-polymerase chain reaction (RT-PCR) targeting the RNA-dependent RNA polymerase (*RdRp*) gene. A single-step RT-PCR was performed on the RNA extracted from the bat specimens. Next-generation sequencing (NGS) was performed on a few representative bat specimens that were tested positive. Phylogenetic analysis was carried out on the partial sequences of *RdRp* gene sequences retrieved from both the bat species and complete viral genomes recovered from *Rousettus* spp.**

***Results*: Bat samples from the seven States were screened, and the RS specimens of eight *Rousettus* spp. and 21 *Pteropus* spp. were found positive for CoV *RdRp* gene. Among these, by Sanger sequencing, partial *RdRp* sequences could be retrieved from three *Rousettus* and eight *Pteropus* bat specimens. Phylogenetic analysis of the partial *RdRp* region demonstrated distinct subclustering of the BtCoV sequences retrieved from these *Rousettus* and *Pteropus* spp. bats. NGS led to the recovery of four sequences covering approximately 94.3 per cent of the whole genome of the BtCoVs from *Rousettus* bats. Three BtCoV sequences had 93.69 per cent identity to CoV BtRt-BetaCoV/GX2018. The fourth BtCoV sequence was 96.8 per cent identical to Bat CoV HKU9-1.**

***Interpretation & conclusions*: This study was a step towards understanding the CoV circulation in Indian bats. Detection of potentially pathogenic CoVs in Indian bats stresses the need for enhanced screening for novel viruses in them. One Health approach with collaborative activities by the animal health and human health sectors in these surveillance activities shall be of use to public health. This would help in**

the development of diagnostic assays for novel viruses with outbreak potential and be useful in disease interventions. Proactive surveillance remains crucial for identifying the emerging novel viruses with epidemic potential and measures for risk mitigation.

Key words Bats - coronavirus - India - next-generation sequencing - phylogenetic - reverse transcription-polymerase chain reaction

A large number of emerging infectious diseases are known to be zoonotic in origin. In the last two decades, many viruses have been identified from bat species[1]. Bats have been recognized as the natural reservoirs of a variety of pathogenic viruses such as Rabies, Hendra, Marburg, Nipah and Ebola virus[2]. Bats are known to harbour coronaviruses (CoVs) and serve as their reservoirs. Alpha-CoV (α-CoVs) and beta-CoV (β-CoVs) have been detected in bats in Asia, Europe, Africa, North and South America and Australasia[3]. In the last two decades, bat CoVs (BtCoVs) garnered considerable attention as potential human pathogens[4,5]. Severe acute respiratory syndrome (SARS)-CoV-2 causing the current pandemic [CoV disease 2019 (COVID-19)] is also a member of the same genus and found to be similar to bat-derived CoV strain RATG13[6]. SARS-CoV-2 is reported to be 96 per cent identical to BtCoV at the whole genome level, and related viruses were identified in the previously sampled bat population in China[7].

CoVs are enveloped, single-stranded, positive-sense RNA viruses with a comparatively large genome size of 26 to 32 kb, classified under the family *Coronaviridae* in the order Nidovirales[8]. According to the International Committee on Taxonomy of Viruses (ICTV), they are classified into four genera, namely, α-CoV, β-CoV, γ-CoV and δ-CoV)[9]. β-CoVs are further classified into four different lineages [lineage A (L_A), lineage B (L_B), lineage C (L_C) and lineage D (L_D)][10]. Most of the human CoVs are either zoonotic in origin or circulate in animals[11]. CoVs can cause a wide range of infections, including respiratory tract infections, gastroenteritis, hepatitis and encephalomyelitis in their respective hosts. It is believed that many of the currently circulating α-CoVs and β-CoVs of mammals have evolutionary links to CoVs from bats[1].

India has a diverse population of bats; around 117 species of bats have been recorded, with around 100 subspecies coming under 39 genera belonging to eight families (*Pteropodidae*, *Rhinolophidae*, *Hipposideridae*, *Megadermatidae*, *Rhinopomatidae*, *Emballonuridae*, *Molossidae* and *Vespertilionidae*)[12]. The Indian Council of Medical Research-National Institute of Virology (ICMR-NIV) at Pune, India, has detected several viruses in bats, including the Nipah virus in *Pteropus medius*, Malsoor virus, Tioman virus and a novel adenovirus in *Rousettus leschenaultia*[13-15]. Nipah viral RNA antibodies could be detected in *Pteropus* bats from many States of India, and the possible link of transmission from bats could be established during the Nipah outbreak which occurred in Kerala in 2018 and 2019[16,17]. The use of conventional polymerase chain reaction/reverse transcription-polymerase chain reaction (PCR/RT-PCR), as well as metagenomics and next-generation sequencing (NGS) technologies, has led to the discovery of many novel viruses in bats. The identification of new CoVs in bats in several neighbouring Asian countries such as China[3], Sri Lanka[18] and Singapore[19,20] and the growing threats of novel CoV diseases such as COVID-19 led us to investigate *Pteropus* and *Rousettus* bats commonly found in India, for identification and characterization of BtCoVs.

Material & Methods

This study was approved by the Institutional Animal Ethics Committee (IAEC) of ICMR-NIV, Pune (IAEC/2019/MEZ/04). Permissions were also obtained from the Principal Chief Conservators of Forests (PCCF)/wildlife wardens of different States/ Union Territories (UT) (Kerala, Karnataka, Tamil Nadu, Himachal Pradesh, Punjab, Gujarat, Odisha, Telangana, Chandigarh and Puducherry).

Study sites and sample collection: Upon obtaining permission from the respective State authorities, bat-roosting sites in each State/UT were identified with the help of the forest officials. Bats were trapped using mist nets and were chemically restrained using isoflurane anaesthesia. Throat swabs (TS) and rectal swabs (RSs) were collected in virus transport medium (VTM) and were transported to ICMR-NIV, Pune, on dry ice. The specimens were collected from *Pteropus* spp. bats from Kerala, Karnataka, Chandigarh, Gujarat, Himachal Pradesh, Odisha, Puducherry, Punjab, Tamil Nadu

and Telangana and *Rousettus* spp. bats from Kerala, Karnataka, Chandigarh, Gujarat, Odisha, Punjab and Telangana States during 2018-2019. These bats were monitored and released after recovery. Twelve bats that died during the trapping process were transported to ICMR-NIV on dry ice. Necropsy of these bats was carried out in the Biosafety Level 4 (BSL-4) containment facility, and tissue specimens (intestine and kidney) collected were tested.

Detection of bat coronavirus using RT-PCR: RNA was extracted from the bat specimens using the MagMAX pathogen RNA/DNA isolation kit (Invitrogen, USA). RT-PCR was performed using Superscript III one-step RT-PCR (Invitrogen, USA) with Platinum High-Fidelity *Taq* polymerase (Invitrogen, USA) using the published BtCoV-specific primers targeting the conserved region of RNA-dependent RNA polymerase (*RdRp*) gene[21]. The amplicon of 440 bp was separated on 1.5 per cent agarose gel and visualized under VersaDoc MP 4000 ultraviolet transilluminator (Bio-Rad, USA).

Sequencing of the positive coronavirus specimens

Sanger sequencing of bat coronavirus: The RT-PCR products were separated on 1.5 per cent agarose gel, and 440 bp bands were excised. The excised gels were extracted and purified using a QIAQuick gel extraction kit (Qiagen, Hilden, Germany). The purified products were quantified, and chain-terminated PCR reactions were performed using pathogen-specific forward and reverse primers[21] with the BigDye Terminator 3.1 sequencing kit (Applied Biosystems, USA). BigDye reactions were purified using the DyeEx 2.0 spin kit (Qiagen, Germany). The purified chain-terminated reactions were sequenced using the ABI PRISM® 3100 Automated DNA Sequencer (Thermo Fisher Scientific, USA). The sequence data generated were assembled using the Sequencer 5.1 software (Accelrys Inc., USA).

Next-generation sequencing (NGS) of bat coronavirus: Selected bat specimens were used for RNA extraction[22,23]. RNA libraries were prepared and quantified by Qubit® 2.0 Fluorometer (Invitrogen, USA). NEB NextrRNA depletion kit (New England Biolabs, USA) was used to remove host ribosomal RNA and re-quantified using Qubit® 2.0 Fluorometer (Invitrogen, USA). In brief, the RNA library preparation involved fragmentation, adenylation, adapter ligation and amplification. The amplified libraries were quantified using KAPA Library Quantification Kit (KapaBiosystems, Roche, Switzerland) as per the manufacturer's protocol and further loaded onto the Illumina Miniseq NGS platform (Illumina, USA)[22,23].

The FASTQ files generated after the completion of the run were analyzed using CLC Genomics Workbench software version 11 (CLC, Qiagen, Germany). *De novo* assembly programme was used to assemble contiguous sequences (contigs). The contigs generated were analyzed using BLAST (*https://blast.ncbi.nlm.nih.gov/Blast.cgi*) to identify matching sequences. The closest matching sequence from GenBank (*https://www.ncbi.nlm.nih.gov/genbank/*) was used for reference mapping.

Phylogenetic analysis of partial and complete genome sequences of bat coronavirus: The CoV sequences retrieved from RS specimens of *Rousettus* spp. bats (n=4) were aligned with whole-genome sequences from GenBank using the create alignment function of the CLC genomics workbench (*https://digitalinsights.qiagen.com*). Partial *RdRp* gene sequences (~419 bp) retrieved by Sanger sequencing, for both the bat species specimens (genomic location: 14,701-15,120) were used to construct a phylogenetic tree along with the available *RdRp* sequences in GenBank. Phylogenetic analysis was carried out using the neighbour-joining method available in MEGA v7 software[24] using the Kimura 2-parameter nucleotide (nt) substitution model with 1000 bootstrap replicates. The nt divergence for the open reading frame (*ORF*) 1a polyprotein (*ORF* 1a), *ORF* 1b polyprotein (*ORF* 1b), spike protein (*S)*, nucleocapsid phospoprotein (*N)*, envelope protein (*E)* and membrane glycoprotein (*M)* genes was estimated using the Kimura 2-parameter model as implemented in the MEGA software. The sequences retrieved in the current study, along with those downloaded from GenBank, were grouped into the genus. AQ1

The viruses from the β-CoV genus were further grouped into lineages, L_A, L_B, L_C and L_D, to estimate the evolutionary divergence over the respective gene sequence pairs between groups using the MEGA software[24]. The distance was estimated using a Kimura 2-parameter model with uniform rates among the sites. The bootstrap of 500 replicates was used to estimate the variation in the model.

Results

The TS and RS specimens for 78 *Rousettus* spp. bats were collected in VTM from seven States (Kerala, Karnataka, Chandigarh, Gujarat, Odisha, Punjab and Telangana). The TS and RS specimens of

508 *Pteropus* spp. bats were also collected in VTM from 10 States/UTs in India (Kerala, Karnataka, Chandigarh, Gujarat, Himachal Pradesh, Odisha, Puducherry, Punjab, Tamil Nadu and Telangana). During the trapping process, 12 (8 *Rousettus* and 4 *Pteropus* spp.) bats died. Organ specimens (intestine and kidney) were collected from these bats (TS and RS specimens of these 12 bats were included in the total number of samples).

Detection of bat coronavirus using RdRp gene RT-PCR: Four of the 78 RS of *Rousettus* spp. bats screened for the BtCoV were found positive. All the positive RS samples belonged to Kerala State. Intestinal specimens of two bats were also found to be positive for the BtCoV. One bat (MCL-19-Bat-606), from Kerala, was tested positive in both the intestinal specimen and the RS. The second bat (MCL-20-Bat-76), from Karnataka, was tested positive only in the intestinal specimen. Altogether, five *Rousettus* spp. bats were positive for the BtCoV. All TS specimens from *Rousettus* spp. were found negative for BtCoV (Table I).

Twenty one of the 508 RSs from *Pteropus* spp. bats screened were tested positive for the BtCoV (Table I). These positive bats belonged to Kerala (n=12), Himachal Pradesh (n=2), Puducherry (n=6) and Tamil Nadu (n=1). The TS specimens of the same bats were tested negative for BtCoV. The TS specimens of RS-negative (n=42) bats were also screened and found to be negative (Table I). A total of 25 bats from both the species were found positive.

Sequencing of the positive coronavirus specimens

Sanger sequencing of bat coronavirus: Using the Sanger sequencing protocol, partial *RdRp* sequences of BtCoV were retrieved from two (out of four amplicons) specimens of *Rousettus* spp. One of the sequences (MCL-19-bat-588/2) showed close identity to BtCoV HKU9-5-2 (AN): HM211099.1; sequence identity (SI): 99.2 per cent, whereas the second *RdRp* sequence (MCL-20-bat-76/10) had an SI of 98.8 per cent with BtCoV HKU9-1 (AN: EF065513.1), both from China.

Sanger's sequencing protocol led to retrieval of eight partial *RdRp* sequences which belonged to *Pteropus* spp. These bats were collected from Kerala (n=5) and Tamil Nadu (n=3) states. One of the three partial *RdRp* sequences from Tamil Nadu had 97.93 per cent SI with BtCoV/B55951/Pte_lyl/CB2-THA (AN: MG256459.1, Thailand). The other two sequences had a minimum of 99.48 per cent SI with the CoV PREDICT_CoV-17/PB072 (AN: KX284942.1, Nepal). One of the five partial *RdRp* sequences from Kerala had 98.88 per cent SI with BtCoV/B55951/Pte_lyl/CB2-THA (AN: MG256459.1, Thailand). The remaining four partial *RdRp* sequences had >97 per cent SI with CoV PREDICT_CoV-17/PB072 (AN: KX284942.1, Nepal).

Next-generation sequencing (NGS) of bat coronavirus: NGS was performed on 10 specimens [4 RS, 2 kidney

Table I. Bat coronavirus positivity in bat specimens screened using RNA-dependent RNA polymerase (*RdRp*) gene-specific reverse transcription-polymerase chain reaction (RT-PCR) in different states

Place of collection	Number of positive/number tested for different bat species for BtCoV *RdRp* gene-specific RT-PCR			
	Pteropus bats (%)		*Rousettus* bats (%)	
	Rectal swabs	Throat swabs	Rectal swabs	Throat swabs
Kerala	12/217 (5.53)	0/21 (0.00)	4/42 (9.52)	0/4 (0.00)
Karnataka	0/78 (0.00)	NT	0/4 (0.00)	0/4 (0.00)
Chandigarh	0/27 (0.00)	NT	0/6 (0.00)	0/6 (0.00)
Gujarat	0/30 (0.00)	NT	0/18 (0.00)	0/18 (0.00)
Odisha	0/30 (0.00)	NT	0/2 (0.00)	0/2 (0.00)
Punjab	0/14 (0.00)	NT	0/2 (0.00)	0/2 (0.00)
Telangana	0/30 (0.00)	NT	0/4 (0.00)	0/4 (0.00)
Himachal Pradesh	2/29 (6.89)	0/6 (0.00)	NA	NA
Puducherry	6/23 (26.09)	0/10 (0.00)	NA	NA
Tamil Nadu	1/30 (3.33)	0/5 (0.00)	NA	NA
	21/508 (4.13)	0/42 (0.00)	4/78 (5.13)	0/40 (0.00)

NT, not tested; NA, not available; BtCoV, bat coronavirus

and 4 intestinal tissue) of the five *Rousettus* bats to retrieve the complete genome of the BtCoV. Kidney and intestine tissues of the bats from Karnataka State (MCL-20-Bat-76) and RS along with intestine tissue of bats from Kerala State (MCL-19-Bat-606) were used for sequencing and analysis.

Two different viruses were retrieved based on the BLAST analysis of the sequences from the kidney and intestine tissues of the bats from Karnataka. Kidney specimen of MCL-20-Bat-76 had an SI of 94 per cent and query coverage (QC) of 94 per cent with CoV BtRt-BetaCoV/GX2018 (AN: MK211379.1), whereas the intestine tissue of the MCL-20-Bat-76 had an SI of 96.8 and 95 per cent QC with the BtCoV HKU9-1 (AN: EF065513.1). The sequences from RS and intestine tissue of the MCL-19-Bat-606 from Kerala, had 93.69 and 93.99 per cent SI to CoV BtRt-BetaCoV/GX2018 (AN: MK211379.1), respectively, with 100 per cent QC. Further, 99.8 per cent of the CoV BtRt-BetaCoV/GX2018 sequences were retrieved from the intestine specimen of the MCL-19-Bat-606. The details of the genome recovered reads mapped and the percentage of reads mapped are summarized in Table II.

Phylogenetic analysis of partial and complete genome sequences of bat coronavirus: A neighbour-joining tree was generated using the partial *RdRp* region sequences derived from *Pteropus* and *Rousettus* spp. bat specimens. It was observed that all the BtCoV sequences wereclustered within the L_D sequences of beta CoVs. A distinct subclustering of the sequences retrieved from *Pteropus* and *Rousettus* spp. bats is shown in Fig. 1. The sequences in the light pink colour are retrieved from the *Pteropus* spp., whereas those in the dark pink region belong to *Rousettus* spp. The sequence divergence of 0.35 was observed between *Pteropus* spp. and *Rousettus* spp., which was obtained by averaging over all the sequence pairs between the two species, determining those to be distinct sequences to each species.

The complete genome sequences of four BtCoV obtained from *Rousettus* spp. specimens were used for generating a neighbour-joining tree (Fig. 2). These sequences were also clustered within L_D of β-CoVs as observed for partial *RdRp* sequence tree. These complete genome sequences were grouped into gene pairs to identify the gene with higher and lower divergence. The complete genomes of the Indian BtCoV sequences were grouped under L_D. The evolutionary divergence of *ORF* 1b was <0.54 between the different β-CoV lineages with a maximum score of 0.7 between different BtCoV sequences used in this study (Table III). *E* gene sequences had larger divergence within the β-CoV genus ranging from 2.18 to 0.94. Lineages L_A and L_C had the maximum divergence of 2.18, whereas the L_B and L_C were the least (0.94). *N* gene has an overall higher divergence amongst different lineages (ranging: 2.08-0.75). Overall, evolutionary divergence for the sequences of each gene pair demonstrated that *S*, *N*, *E* and *M* genes from the α- and δ-CoV highly diverged across the different genus. In contrast, the *ORF* 1b was less divergent across the genera (Table III).

Discussion

As per the available information, the BtCoV causing human infection belongs to α- and β-CoV genera of the *Coronaviridae* family. β-CoV genus has five strains known to infect humans[25]. The two human-infecting strains (NL63 and 229E) from α-CoV genus which cause mild-to-moderate respiratory infections are believed to have originated in bats[25]. Two members of the β-CoV genus (HCoV-OC43 and HCoV-HKU1) are known to cause the common cold and lower respiratory tract infections[26]. The other three are now shown to be pathogenic to humans (SARS-CoV-1, MERS-CoV and SARS-CoV-2). The SARS-CoV-1 and SARS-CoV-2 belong to L_B and MERS CoV belongs to L_C of β-CoV genus[27].

Table II. Details of the genome recovered reads mapped and the percentage of reads mapped from the *Rousettus* bat samples

Sample details	Sample type	Virus retrieved	Relevant reads	Percentage of reads	Percentage of genome recovered
MCL-20-Bat-76	Kidney	Coronavirus BtRt-BetaCoV/GX2018	1632	0.015	94.39
	Intestine	BtCoV HKU9-1	4499	0.056	95.75
MCL-19-Bat-606	Rectal swab	Coronavirus BtRt-BetaCoV/GX2018	13,973	0.114	99.53
	Intestine	Coronavirus BtRt-BetaCoV/GX2018	10,214,492	93.476	99.87

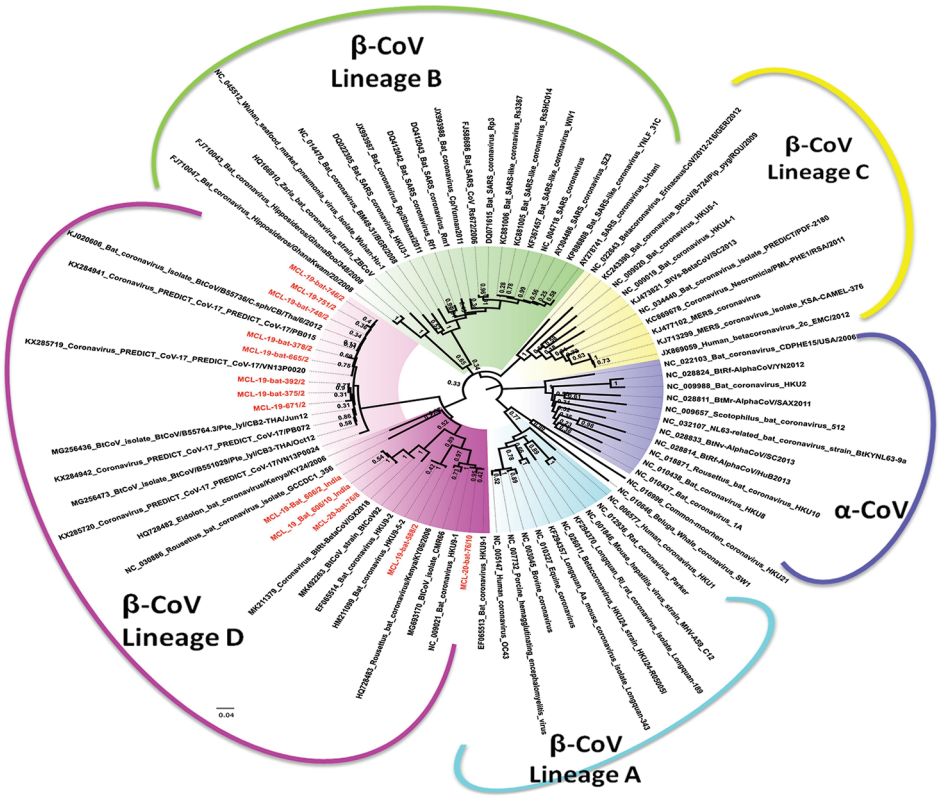


Fig. 1. Neighbour-joining tree for the RNA-dependent RNA polymerase (*RdRp*)-partial sequence (genomic location: 14,701-15,120) generated from Sanger sequencing. The tree was constructed using the *RdRp* sequence available in the GenBank sequences. Kimura 2-parameter model was used as the substitution model to generate the tree. A bootstrap replication of 1000 cycles was performed to generate the tree to assess the statistical robustness.

The phylogenetic analysis for the partial *RdRp* region revealed the presence of distinct BtCoVs in both the bats. The genomic sequences retrieved from the Indian sequences form a distinct cluster. The three CoV_BtRtBetaCoV/GX2018 sequences retrieved from the Indian *Rousettus* bats were 5.8-6.7 per cent different from the reference sequence, which was retrieved from *Rhinolophus affinis*. The two CoV_BtRtBetaCoV/GX2018 sequences retrieved from different bats were 1.2 per cent different from each other. The effect of host influence on the nt usage of the virus cannot be denied; however, it needs to be explored further in detail.

Bats are reservoirs for viruses with human pathogenic potential[28,29], and are known to harbour a broad range of CoVs[1]. The global distribution of bats, along with the different types of cell receptors present within them, favours virus replication, and is a possible link to their intraspecies transmission. The interspecies spill-over of a BtCoV to humans is thought to occur through an intermediate host, in which the virus replicates through yet completely unidentified routes. In India, regions of the Western Ghats, particularly in Kerala, are reported to have habitat for diverse bat populations. The reports of pathogenic human viruses from bat specimens demand enhanced methods to monitor human exposure to various bat species. Investigations in unexplored regions/States should be focused on gaining further insights into CoV diversity within Indian bat populations.

Earlier, we had reported the presence of pathogenic viruses such as the Nipah virus in *Pteropus* bats in India[16]. In the present scenario of changing demography and ecological manipulations, it is challenging to have checks on the encounters of bats with other animals and humans. Therefore, active and continuous surveillance remains crucial for outbreak alerts for bat-associated viral agents with epidemic potential, which would be helpful in timely interventions.

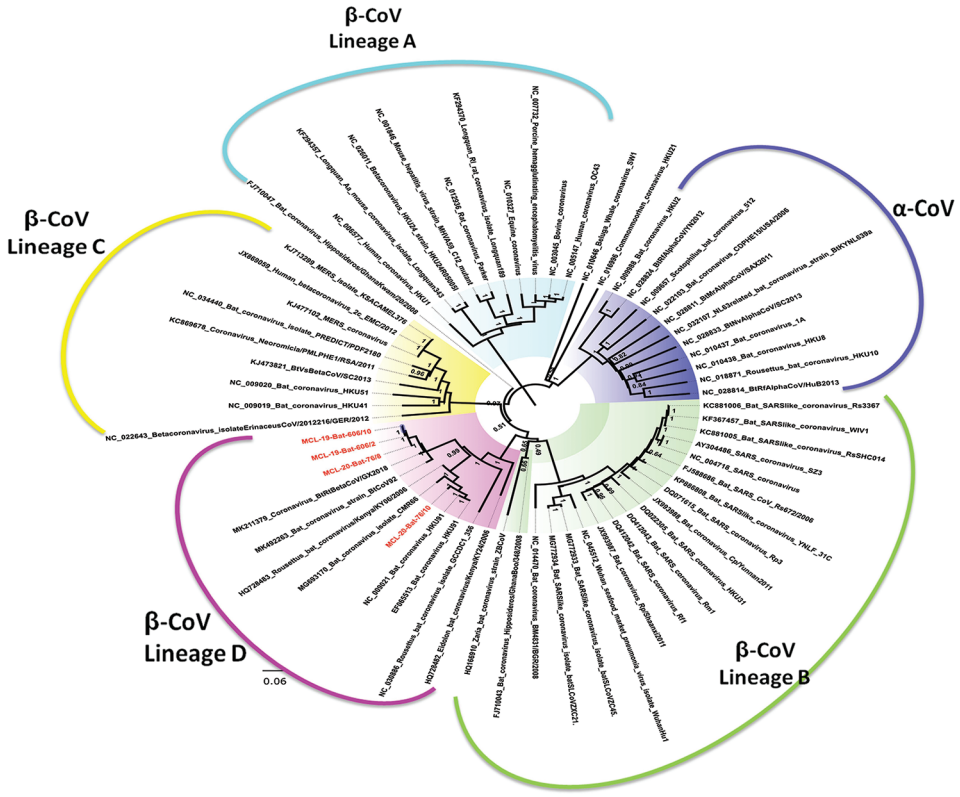


Fig. 2. Phylogenetic tree for the complete genome tree: A neighbour-joining tree was generated using the representative complete genome sequence available in the GenBank sequences. Kimura 2-parameter model was used as the substitution model to generate the tree. A bootstrap replication of 1000 cycles was performed to generate the tree to assess the statistical robustness.

Although CoVs in the subfamily *Coronavirinae* do not usually produce clinical symptoms in their natural hosts (bats), accidental transmission of these viruses to humans and other animals may result in respiratory, enteric, hepatic or neurologic diseases of variable severity. It is still not understood as to why only certain CoVs can infect people.

There is a need of proactive surveillance of zoonotic infections in bats. Detection and identifications of such aetiological agents will provide leads for the development of diagnostic along with preparedness and readiness to deal with such emergent viruses thereby quickly containing them. The detection and identification of such viruses from bats also recommends cross-sectional antibody surveys (human and domestic animals) in localities where the viruses have been detected. Similarly, if epidemiological situation demands, evidence-based surveillance should also be conducted. There is a need of developing strong mechanisms for working jointly with various stakeholders such as wildlife, poultries, animal husbandry and human health departments.

In conclusion, our study showed detection of pathogenic CoVs in two species of Indian bats. Continuous active surveillance is required to identify the emerging novel viruses with the epidemic potential.

***Acknowledgment*:** Authors acknowledge the encouragement and support extended by Prof. (Dr) Balram Bhargava, Secretary, Department of Health Research, Ministry of Health & Family Welfare, Government of India, Director-General, Indian Council of Medical Research (ICMR), New Delhi, and Dr Raman Gangakhedkar, Division of Epidemiology & Communicable Diseases, ICMR, New Delhi. Authors acknowledge the Principal Conservators of Forests of Kerala, Karnataka, Tamil Nadu, Himachal Pradesh, Punjab, Gujarat, Odisha, Telangana, Chandigarh and Puducherry and other concerned officials for their permission to capture bats. Authors are also thankful to Servshri Rajen Lakra, Prasad Sarkale, Pranita Gawande, Kaumudi Kalele, Shrimati Ashwini Waghmare, Nikita Bankar and Shri Sanjay Thorat for providing technical support. Authors acknowledge the contribution of the field team for the hard work performed by Servshri Sanjay Gopale, Annasaheb Suryawanshi, Manjunath Holeppanavar, Ganesh Chopade, Madhav Acharya, D.K. Singh and Sachin Daigude for technical assistance during the study.

Table III. Evolutionary divergence for *ORF* 1b, *S*, *N* and *M* genes for the retrieved sequences with other reference sequences. The lower right-check hand matrix of the table depicts the divergence and the upper left-check matrix of the matrix (blue colour) depicts the variation observed in the bootstrap replication

N gene	Alpha	Delta	Gamma	L_A	L_B	L_C	L_D	*M* gene	Alpha	Delta	Gamma	L_A	L_B	L_C	L_D
Alpha		0.15	0.09	0.10	0.08	0.08	0.09	Alpha		0.11	0.12	0.05	0.06	0.05	0.06
Delta	2.08		0.11	0.16	0.09	0.11	0.10	Delta	1.50		0.26	0.08	0.16	0.10	0.11
Gamma	1.57	1.49		0.08	0.08	0.09	0.08	Gamma	1.53	1.84		0.10	0.12	0.11	0.09
L_A	1.84	1.73	1.37		0.05	0.05	0.06	L_A	0.92	1.24	1.30		0.06	0.05	0.05
L_B	1.48	1.37	1.32	1.09		0.03	0.04	L_B	1.05	1.51	1.37	0.92		0.05	0.05
L_C	1.57	1.52	1.42	1.07	0.75		0.04	L_C	0.99	1.35	1.27	0.80	0.82		0.05
L_D	1.64	1.46	1.36	1.27	0.90	0.97		L_D	0.99	1.42	1.23	0.84	0.79	0.82	
ORF 1b	Alpha	Delta	Gamma	L_A	L_B	L_C	L_D	*ORF* 1a	Alpha	Delta	Gamma	L_A	L_B	L_C	L_D
Alpha		0.01	0.01	0.01	0.01	0.01	0.01	Alpha		0.02	0.02	0.01	0.02	0.02	0.03
Delta	0.70		0.01	0.02	0.01	0.01	0.01	Delta	1.32		0.03	0.02	0.03	0.03	0.04
Gamma	0.62	0.67		0.01	0.01	0.01	0.01	Gamma	1.14	1.33		0.02	0.03	0.02	0.04
L_A	0.61	0.69	0.60		0.01	0.01	0.01	L_A	1.22	1.01	1.30		0.02	0.02	0.04
L_B	0.60	0.70	0.65	0.54		0.01	0.01	L_B	1.26	1.42	1.41	1.19		0.01	0.02
L_C	0.58	0.69	0.62	0.53	0.50		0.01	L_C	1.35	1.41	1.44	1.19	0.97		0.03
L_D	0.60	0.67	0.61	0.53	0.50	0.52		L_D	1.26	1.27	1.39	1.09	0.90	1.03	
S gene	Alpha	Delta	Gamma	L_A	L_B	L_C	L_D	*E* gene	Alpha	Delta	Gamma	L_A	L_B	L_C	L_D
Alpha		0.02	0.02	0.03	0.03	0.03	0.02	Alpha		0.12	0.18	0.09	0.15	0.15	0.12
Delta	0.86		0.03	0.04	0.04	0.06	0.03	Delta	1.14		0.47	0.22	0.41	0.28	0.17
Gamma	1.14	0.96		0.04	0.05	0.06	0.04	Gamma	1.59	1.64		0.22	0.24	0.32	0.19
L_A	1.36	1.28	1.43		0.03	0.03	0.02	L_A	1.03	1.58	1.57		0.23	0.21	0.25
L_B	1.33	1.23	1.34	1.19		0.04	0.02	L_B	1.24	1.75	1.40	1.83		0.11	0.14
L_C	1.42	1.32	1.46	1.17	1.03		0.03	L_C	1.37	1.64	1.83	2.18	0.94		0.17
L_D	1.34	1.24	1.41	1.16	1.00	1.11		L_D	1.25	1.42	1.52	1.95	1.16	1.37	

ORF 1a, open reading frame 1a polyprotein; *ORF* 1b, open reading frame 1b polyprotein; *S*, spike glycoprotein; *N*, nuclocapsid phospoptotein; *M*, membrane glycoprotein; *E*, envelope protein

AQ2

***Financial support & sponsorship*:** This work was supported by intramural funding by the Indian Council for Medical Research, New Delhi.

***Conflicts of Interest*:** None.

References

1. Banerjee A, Kulcsar K, Misra V, Frieman M, Mossman K. Bats and coronaviruses. *Viruses* 2019; *11*. pii: E41.
2. Plowright RK, Eby P, Hudson PJ, Smith IL, Westcott D, Bryden WL, *et al*. Ecological dynamics of emerging bat virus spillover. *Proc Biol Sci* 2015; *282* : 20142124.
3. Wong ACP, Li X, Lau SKP, Woo PCY. Global epidemiology of bat coronaviruses. *Viruses* 2019; *11* . pii: E174.
4. Stephenson J. SARS in China. *JAMA* 2004; *291* : 2534.
5. Memish ZA, Mishra N, Olival KJ, Fagbo SF, Kapoor V, Epstein JH, *et al*. Middle East respiratory syndrome coronavirus in bats, Saudi Arabia. *Emerg Infect Dis J* 2013; *19* : 1819-23.
6. Andersen KG, Rambaut A, Lipkin WI, Holmes EC, Garry RF. The proximal origin of SARS-CoV-2. *Nature Med* 2020. [doi: https://doi.org/10.1038/s41591-020-0820-9].
7. Zhou P, Yang XL, Wang XG, Hu B, Zhang L, Zhang W, *et al*. A pneumonia outbreak associated with a new coronavirus of probable bat origin. *Nature* 2020; *579* : 270-3.
8. Fehr AR, Perlman S. Coronaviruses: an overview of their replication and pathogenesis. *Methods Mol Biol* 2015; *1282* : 1-23.
9. de Groot RJ, Baker SC, Baric R, Enjuanes L, Gorbalenya AE, Holmes KV, *et al*. *Coronaviridae* - Positive sense RNA viruses - Positive Sense RNA Viruses (2011). International Committee on Taxonomy of Viruses (ICTV); 2019. Available from: *https://talk.ictvonline.org/ictv-reports/ictv_9th_report/positive-sense-rna-viruses-2011/w/posrna_viruses/222/* coronaviridae, accessed on February 19, 2020.
10. de Groot RJ, Baker SC, Baric RS, Brown CS, Drosten C, Enjuanes L, *et al*. Commentary: Middle East respiratory syndrome coronavirus (MERS-CoV): Announcement of the Coronavirus Study Group. *J Virol Am Soc Microbiol* 2013; *87* : 7790-2.
11. Woo PCY, Lau SKP, Huang Y, Yuen KY. Coronavirus diversity, phylogeny and interspecies jumping. *Exp Biol Med (Maywood)* 2009; *234* : 1117-27.
12. Raghuram H, Jain M, Balakrishnan R. Species and acoustic diversity of bats in a palaeotropical wet evergreen forest in Southern India. *Curr Sci* 2014; *107* : 631-41.
13. Raut CG, Yadav PD, Towner JS, Amman BR, Erickson BR, Cannon DL, *et al*. Isolation of a novel adenovirus from *Rousettus leschenaultii* bats from India. *Intervirology* 2012; *55* : 488-90.
14. Yadav P, Sarkale P, Patil D, Shete A, Kokate P, Kumar V, *et al*. Isolation of Tioman virus from *Pteropus giganteus* bat in North-East region of India. *Infect Genet Evol* 2016; *45* : 224-9.
15. Yadav P, Deoshatwar A, Shete A, Tandale B, Patil D, Dalal S, *et al*. Serosurvey of Malsoor virus among *Rousettus leschenaulti* bat & human population residing nearby Robber's cave, Mahabaleshwar, Maharashtra, India. *Indian J Med Res* 2017; *146* : 545-7.
16. Yadav P, Sudeep A, Gokhale M, Pawar S, Shete A, Patil D, *et al*. Circulation of Nipah virus in *Pteropus giganteus* bats in northeast region of India, 2015. *Indian J Med Res* 2018; *147* : 318-20.
17. Yadav PD, Shete AM, Kumar GA, Sarkale P, Sahay RR, Radhakrishnan C, *et al*. Nipah virus sequences from humans and bats during Nipah outbreak, Kerala, India, 2018. *Emerg Infect Dis* 2019; *25* : 1003-6.
18. Kudagammana HDWS, Thevanesam V, Chu DKW, Eriyagama NB, Peiris JSM, Noordeen F. Coronaviruses in guano from *Pteropus medius* bats in Peradeniya, Sri Lanka. *Transbound Emerg Dis* 2018; *65* : 1122-4.
19. Mendenhall IH, Kerimbayev AA, Strochkov VM, Sultankulova KT, Kopeyev SK, Su YCF, *et al*. Discovery and characterization of novel bat coronavirus lineages from Kazakhstan. *Viruses* 2019; 11 : pii: E356..
20. Lim XF, Lee CB, Pascoe SM, How CB, Chan S, Tan JH, *et al*. Detection and characterization of a novel bat-borne coronavirus in Singapore using multiple molecular approaches. *J General Virol* 2019; *100* : 1363-74.
21. Poon LLM, Chu DKW, Chan KH, Wong OK, Ellis TM, Leung YHC, *et al*. Identification of a novel coronavirus in bats. *J Virol* 2005; *79* : 2001-9.
22. Yadav PD, Albariño CG, Nyayanit DA, Guerrero L, Jenks MH, Sarkale P, *et al*. Equine Encephalosis virus in India, 2008. *Emerg Infect Dis* 2018; *24* : 898-901.
23. Yadav PD, Nyayanit DA, Shete AM, Jain S, Majumdar TP, Chaubal GY, *et al*. Complete genome sequencing of Kaisodi virus isolated from ticks in India belonging to Phlebovirus genus, family *Phenuiviridae*. *Ticks Tick Borne Dis* 2019; *10* : 23-33.
24. Kumar S, Stecher G, Tamura K. MEGA7: Molecular Evolutionary genetics analysis version 7.0 for bigger datasets. *Mol Biol Evol* 2016; *33* : 1870-4.
25. Lim YX, Ng YL, Tam JP, Liu DX. Human coronaviruses: A review of virus-host interactions. *Diseases* 2016; *4*. pii: E26.
26. European Centre for Disease Prevention and Control. *Factsheet for health professionals on coronaviruses*. Available from: *https://www.ecdc.europa.eu/en/factsheet-health-professionals-coronaviruses*, accessed on March 19, 2020.
27. Cui J, Li F, Shi ZL. Origin and evolution of pathogenic coronaviruses. *Nat Rev Microbiol* 2019; *17* : 181-92.
28. Smith I, Wang LF. Bats and their virome: An important source of emerging viruses capable of infecting humans. *Curr Opin Virol* 2013; *3* : 84-91.
29. Calisher CH, Childs JE, Field HE, Holmes KV, Schountz T. Bats: Important reservoir hosts of emerging viruses. *Clin Microbiol Rev* 2006; *19* : 531-45.

For correspondence: Dr Devendra T. Mourya, ICMR-National Institute of Virology, Pune 411 021, Maharashtra, India
e-mail: dtmourya@gmail.com

Author Queries???
AQ1: Is it membrane or matrix membrane
AQ2: Is it matrix or membrane protein

Genomics of Indian SARS-CoV-2: Implications in genetic diversity, possible origin and spread of virus

Mainak Mondal[#], Ankita Lawarde[#] and Kumaravel Somasundaram*

Department of Microbiology and Cell Biology, Indian Institute of Science, Bengaluru 560 012, India

World Health Organization (WHO) declared COVID-19 as a pandemic disease on 11 March 2020. Comparison of genome sequences from diverse locations allows us to identify the genetic diversity among viruses which would help in ascertaining viral virulence, disease pathogenicity, origin and spread of the SARS-CoV-2 between countries. The aim of this study is to determine the genetic diversity among Indian SARS-CoV-2 isolates. Initial examination of the phylogenetic data of SARS-CoV-2 genomes (*n* = 3123) from different continents deposited at GISAID (Global Initiative on Sharing All Influenza Data) revealed multiple origin for Indian isolates. An in-depth analysis of 558 viral genomes derived from samples representing countries from USA, Europe, China, East Asia, South Asia, Oceania, Middle East regions and India revealed that most Indian samples are divided into two clusters. A1 sub-cluster showed more similarity to Oceania and Kuwait samples, while A2 sub-cluster grouped with South Asian samples. In contrast, cluster B grouped with countries from Europe, Middle East and South Asia. Viral clade analysis of Indian samples revealed a high occurrence of G clade (D614G in spike protein; 37%), which is a European clade, followed by I clade (V378I in ORF1ab; 12%), which is an Oceania clade with samples having Iran connections. While A1 cluster is enriched with I clade, the cluster B is enriched with G clade type. Thus our study identifies that the Indian SARS-CoV-2 viruses are enriched with G and I clades in addition to 50% samples with unknown genetic variations. The potential origin to be countries mainly from Europe, Middle East Oceania and South Asia regions, which strongly imply the spread of virus through most travelled countries. The study also emphasizes the importance of pathogen genomics through phylogenetic analysis to discover viral genetic diversity and understand the viral transmission dynamics with eventual grasp on viral virulence and disease pathogenesis.

Keywords: COVID-19, genetic diversity, pandemic, SAR-CoV-2, severe acute respiratory syndrome.

*For correspondence. (e-mail: skumar1@iisc.ac.in)
[#]Contributed equally.

A novel corona virus (SARS-CoV-2) causes acute respiratory disease (Coronavirus disease 2019; COVID-19), which was initially found in China but now it is spread all over the world[1]. The total number of COVID-19 cases diagnosed so far exceeds 4.1 million worldwide as on 11 May 2020 with the number almost reaching 68,000 in India[2,3]. SARS-CoV-2 is an enveloped, non-segmented positive-sense RNA virus with a large genome of approximately 30 kb in length[1].

A total of 17,878 viral isolates have been sequenced and deposited online as on 11 May 2020 (ref. 4). Genome sequence analysis of viral genome between countries would help to understand the origin and also the severity of the disease process itself. The sequence information for 173 Indian viral isolates is available in the Global Initiative on Sharing All Influenza Data (GISAID) database[4–6]. In this study, we carried out systematic analysis of genome sequences of Indian SARS-CoV-2 isolates and inferred the possible source of origin and important genetic variants of Indian viruses.

Samples and methods

Sample collection

We have collected 173 genome sequences from Indian clinical samples deposited at GISAID as on 8 May 2020 (ref. 4). In addition, we also collected another 421 representative genome sequences of samples from USA (75), Europe (80), China (75), East Asia (64), South Asia (41), Oceania (75) and Middle East (11). Among Indian viral isolates, seven viral genome sequences that belong to passaged virus through cell lines (Vero CCL81 isolate P1) and one sequence with incomplete genome were excluded in this study. The viral sequences having complete and high coverage (*n* = 137) were alone selected for phylogeny analysis. However, for viral clade analysis, an additional set of 28 Indian viral genomes with low coverage was also used. Genome accession and sample data information can be found in 'SupplymentaryData.xlsx'.

Phylogenetic tree analysis

A total of 558 complete genomes were taken for alignment using MAFFT version 7.402 at CIPRES Science

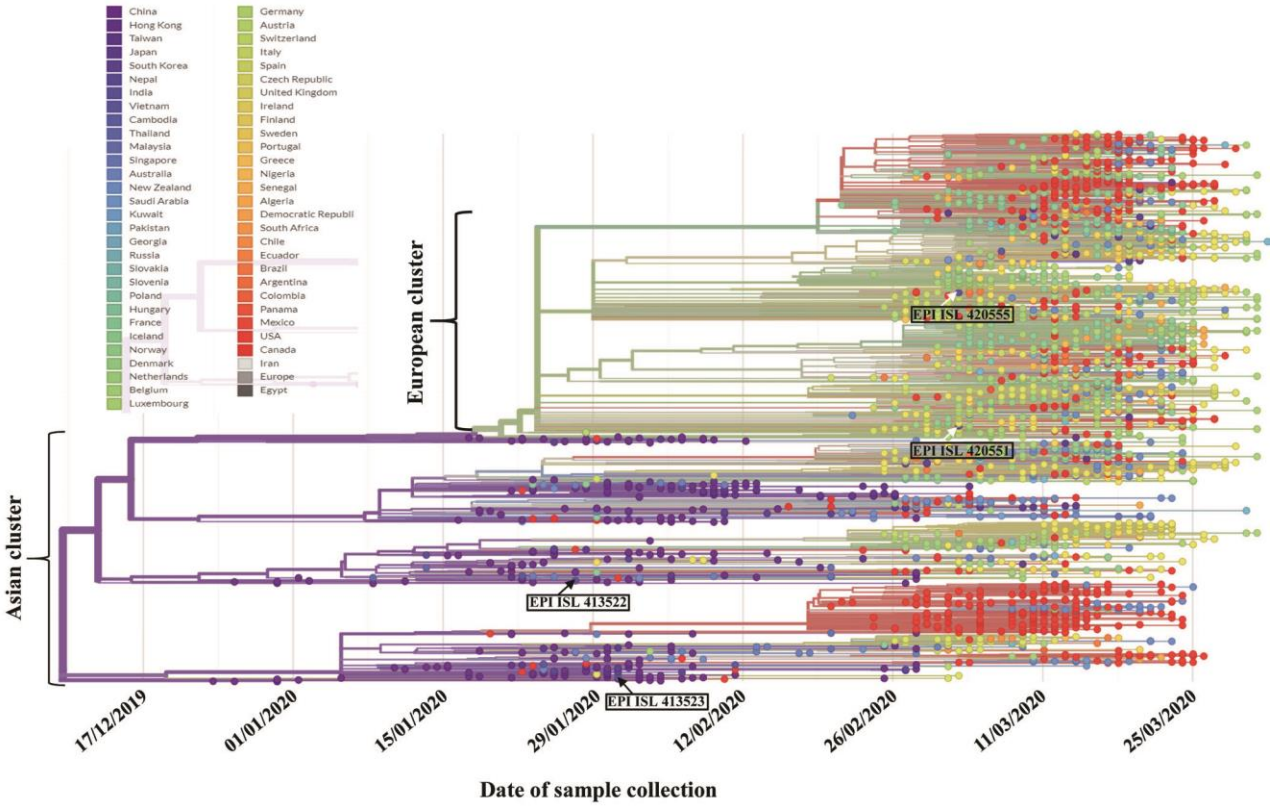


Figure 1. A modified view of phylogenetic analysis (rectangular view) of genome sequences of SARAS-CoV-2 (n = 3123) taken from https://www.gisaid.org/epiflu-applications/next-hcov-19-app/. The list of countries from where samples were used is given with their colour code.

Gateway[7]. Phylogenetic analysis by maximum likelihood (ML) method was carried out using IQ tree version 1.6.12 (ref. 8). TIM + F + R2 having lowest BIC score (109220.216) was selected as best substitution model out of 279 substitution model fitted. Analysis was carried out with 10^3 ultrafast bootstrap replicates. The tree file obtained was visualized using Figtree version 1.4.4 (ref. 9).

Viral clade analysis

A reference Wuhan isolate and all the other viral genome sequences (n = 586) were obtained and used for this analysis. Individual genes namely *ORF1ab*, *S*, *ORF3* and *ORF8* were extracted from the whole genome. The genes were aligned using CLUSTAL Omega algorithm[10] and translated to amino acid sequences. The aligned protein coding genes was visualized in BioEdit version 7.2.5 (ref. 11).

Results and discussion

To identify the origin of Indian isolates of SAR-CoV-2 virus, we examined the phylogenetic data from GISAID[4]. The phylogenetic data from 3123 samples, which included 4 Indian isolates, available at GISAID website were analysed (Figure 1; Supplementary Figure 1). It is of our interest to note that there are two major clusters – Asian cluster is represented by purple and related colours, while the European cluster is represented by greenish yellow. The Indian samples, represented by the arrows (black and white) clustered with both Asian and European clusters. While the Indian samples with black arrows were isolated during January 2020, the other two samples with white arrows were isolated during March 2020 (more details later).

Further to precisely map the origin of Indian SARS-CoV-2 isolates, we carried out an independent phylogenetic analysis using a selected set of samples representing most regions and countries where the COVID-19 infection rate is high. The samples which were collected earliest during this pandemic in each of the countries were only considered. The set consisted of 558 samples as detailed earlier. The analysis shows interesting features about the possible source of origin of Indian SARS-CoV-2 samples (Figure 2; Supplementary Figure 2). In particular, the Indian samples are located away from China/East Asian samples and are divided between two clusters, A and B. While the sub-cluster A1 consists of mostly Oceania/Kuwait samples besides a large number of Indian samples (n = 19), the A2 sub-cluster is enriched with South Asian samples along with Indian samples (n = 53). The cluster B consists mostly of European and few numbers of Middle East/South Asian samples besides a large number of

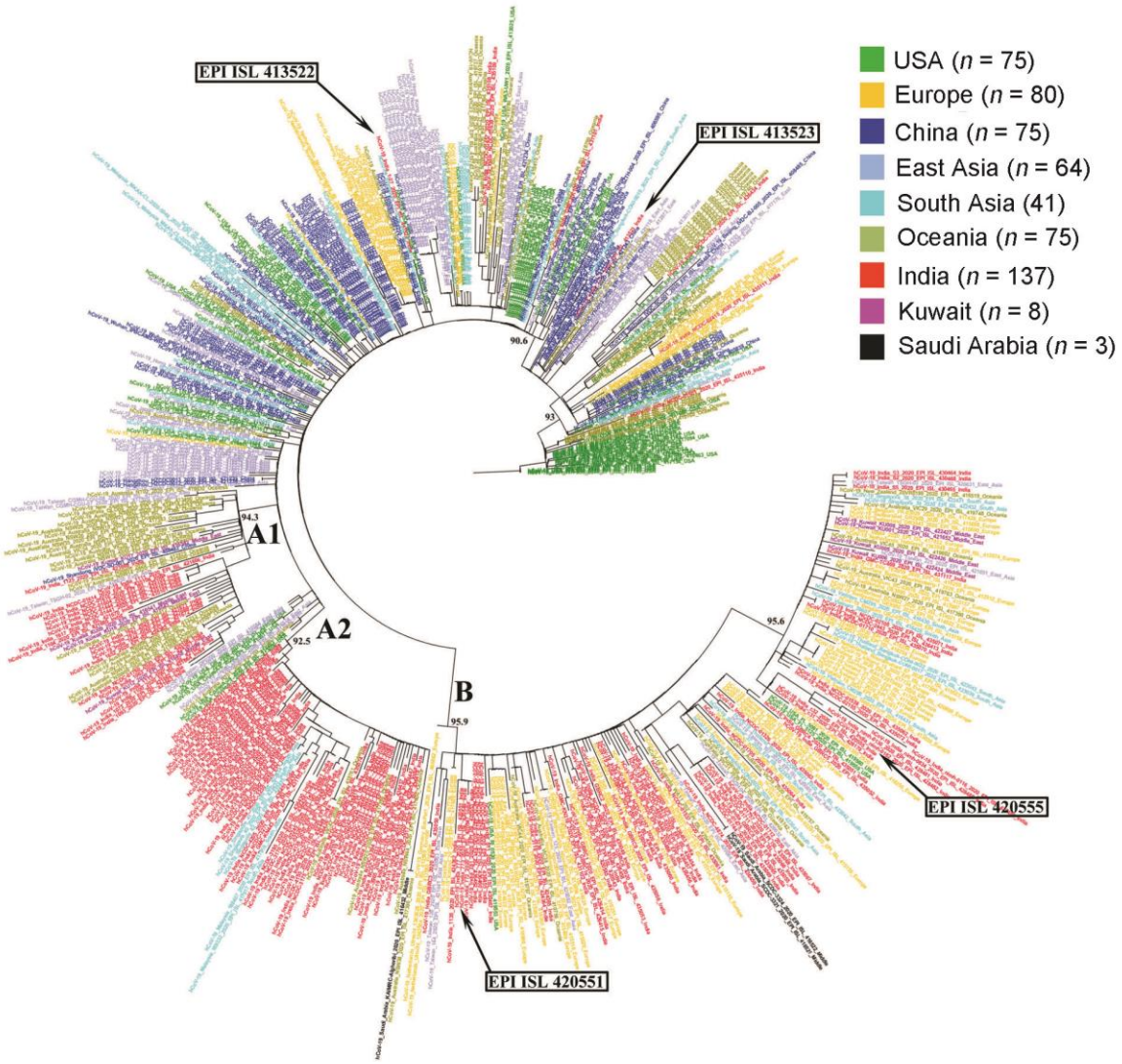


Figure 2. Maximum likelihood (ML) phylogenetic tree was constructed using whole genome sequences obtained from 558 individual SARS-Cov-2 viral isolates. The branch was tipped according to strains with respective country. The nodes represent bootstrap values. The taxa were coloured according to different countries. The colour code for different regions/countries is given.

Indian samples (n = 57). The analysis revealed that most Indian SARS-CoV-2 viruses (129 out of 137) show more similarity to that of specific countries. In cluster A, Indian samples show more similarity to the viruses found in Oceania, Kuwait and South Asian samples, while in the cluster B, Indian samples show more similarity to mainly European and few numbers of Middle East/South Asian samples. These results indicate that majority of Indian SARS-CoV-2 viruses have originated from Europe, Middle East, South Asia and Oceania regions. The remaining Indian isolates (n = 8) have grouped with other clusters which contained most samples from China and East Asia. This indicates that these viruses might have been introduced by Indian travellers from China and its neighbouring countries as they show close resemblance to ancestral Chinese virus. Indeed, a recent study reported that two viruses (EPI_ISL_413522 and EPI_ISL_413523) from this group were isolated from patients who travelled from Wuhan, China[5]. A similar correlation for the remaining six samples could not be made as their travel information is not available.

Given the fact that the date of sample collection probably coincides with the time of disease occurrence, the collection dates of different Indian samples provide some hint at the origin and spread of virus. The first two Indian viral isolates (EPI_ISL_413522 and EPI_ISL_413523) that were collected during January 2020 from patients who travelled from Wuhan, China showed more similarity with China/East Asia viral isolates. This conclusion is well supported by the fact that the initial outbreak of SARS-CoV-2 virus in Wuhan happened during December 2019 (ref. 12). A large majority of Indian viral isolates (n = 129), which were collected during March/April 2020, show more similarity with samples from Europe,

Middle East, South Asia and Oceania regions. The delay in the occurrence of majority of Indian COVID-19 cases probably indicates the time taken for the virus to spread from China to other countries from where the Indian travellers would have contracted the virus.

It is interesting to note that two samples (EPI_ISL_420551 and EPI_ISL_420555), that are seen with European cluster in Figure 1, grouped with the cluster B, are enriched with European and few numbers of Middle East/South Asian samples according to Figure 2. Similarly, other two samples (EPI-ISL_413522 and EPI_ISL_413523) (collected from patients who had travel history from Wuhan, China) that grouped with samples from China were also identified to be associated with the major Asian cluster in Figure 1. Thus, the results of our independent phylogenetic analysis (according to Figure 2) match with that of analysis done by GISAID.

According to specific variations in different viral proteins compared to Chinese ancestral SARS-CoV-2 virus, GISAID identified three clades of SARS-CoV-2 namely, G, V and S clades[4,13]. G clade is characterized by D614G (A23403G) in S protein and largely encompasses sequences from Europe. V clade is characterized by G251V (G26144T) in ORF3 and mostly includes Asian and European sequences. S clade is characterized by the presence of L84S (C8782T) in ORF8 and mostly comprises sequences from North America. Recently, a new clade of SARS-CoV-2 carrying V378I (G1397A) in ORF1ab has been linked to travellers returning from Iran to Australia[14]. We then studied the genomes of Indian SARS-CoV-2 viruses to find out the association between Indian samples and different clades. The sample set (n = 558) which was used for phylogenetic analysis and an additional set of Indian samples with low coverage (n = 28) was subjected to clade analysis which revealed several interesting facts (Supplementary Figures 3 and 4). The samples from China failed to classify with any of the clades except a significant proportion of S clade which signifies the ancestral nature of Chinese viruses. While most groups had significant proportion of unclassified samples, East Asia and South Asia samples were found to split among G, V and S clades. Oceania samples were represented in all clades with a significant high proportion of I and S clades. In contrast to these groups, European samples showed rather a very high proportion of G clade type and USA samples showed a high proportion of S clade type. While 50% of Indian samples do not belong to any of the clades, a significant enrichment in G clade (35.15%) and I clade (12.12%) is seen (Figure 2; Supplementary Figures 3 and 4). The unclassified Indian samples do not appear to be more similar to Chinese ancestral viruses as they are grouped in A and B clusters (according to Figure 2) and found separated from ancestral samples. It is interesting to note that all I clade Indian samples are part of A1 cluster (according to Figure 2) and all G clade Indian samples are part of B cluster (according to Figure 2) (Supplementary Figure 2). Further studies are needed to identify the specific genetic variation(s) unique to this unclassified Indian samples. We conclude from this analysis that there is a higher occurrence of G clade (35.15%) and I clade (12.12%) among Indian samples. The type of unique variations specific to the remaining 50% of Indian samples is yet to be identified.

We have also analysed the type of viral clades in different states of India (Supplementary Table 1; Supplementary Figure 5). This analysis was limited by fewer numbers of viral sequences available for many states. Considering mainly those states where more number of viral sequences are available, G clade is prevalent in many states in particular Delhi, Gujarat, Karnataka, Madhya Pradesh and West Bengal. While a large percentage of I clade samples could not be tied to one state, it is interesting to note that Kargil and Ladakh samples are enriched with I clade.

Our finding that Indian SARS-CoV-2 isolates belong to specific clades may have important consequences with respect to virus transmission rate and virulence; extent of the disease severity and various other aspects of disease pathogenesis. It has been reported that viruses belonging to different clades may differ in their virulence[15]. For example, the G clade viruses carry glycine (G) corresponding to the codon 614 of S protein instead of aspartic acid (D) in other clades. Phylogenetic analysis identified that D614G mutation is originated from ancestral D residue seen in the reference Wuhan virus[16]. This residue is located very close to glycosylation region of the viral spike protein encoded by S gene[17]. It has been proposed that mutations in and around glycosylation region may alter viral spike protein structure and hence the membrane fusion process resulting in varied pathogenicity and transmissibility. Further, the difference in the death rate of COVID-19 patients of East Coast versus West Coast of USA is implicated to their difference in their G clade status[15]. G clade virus has also been identified to be highly transmissible by utilizing multiple mechanisms over its ancestral virus[18,19]. Several other studies also reported that mutations in spike protein of other Corona viruses alter the virulence[20–22].

The presence of multiple clades of SARS-CoV-2 strains in a population may also have serious implications in the accuracy of diagnostic tests that are being employed worldwide. It is not clear whether the diagnostic tests based on detection of antibody or quantifying the viral RNA genome that are in use distinguish these variations. Hence, it is important to develop diagnostic kits based on the type(s) of clades prevalent in an area. Indeed, it is reported that the serology-based rapid tests are ineffective in detecting COVID-19 positive cases in India and elsewhere in the world. The variation created by the presence of different viral clades in a population also needs to be considered seriously in developing vaccines.

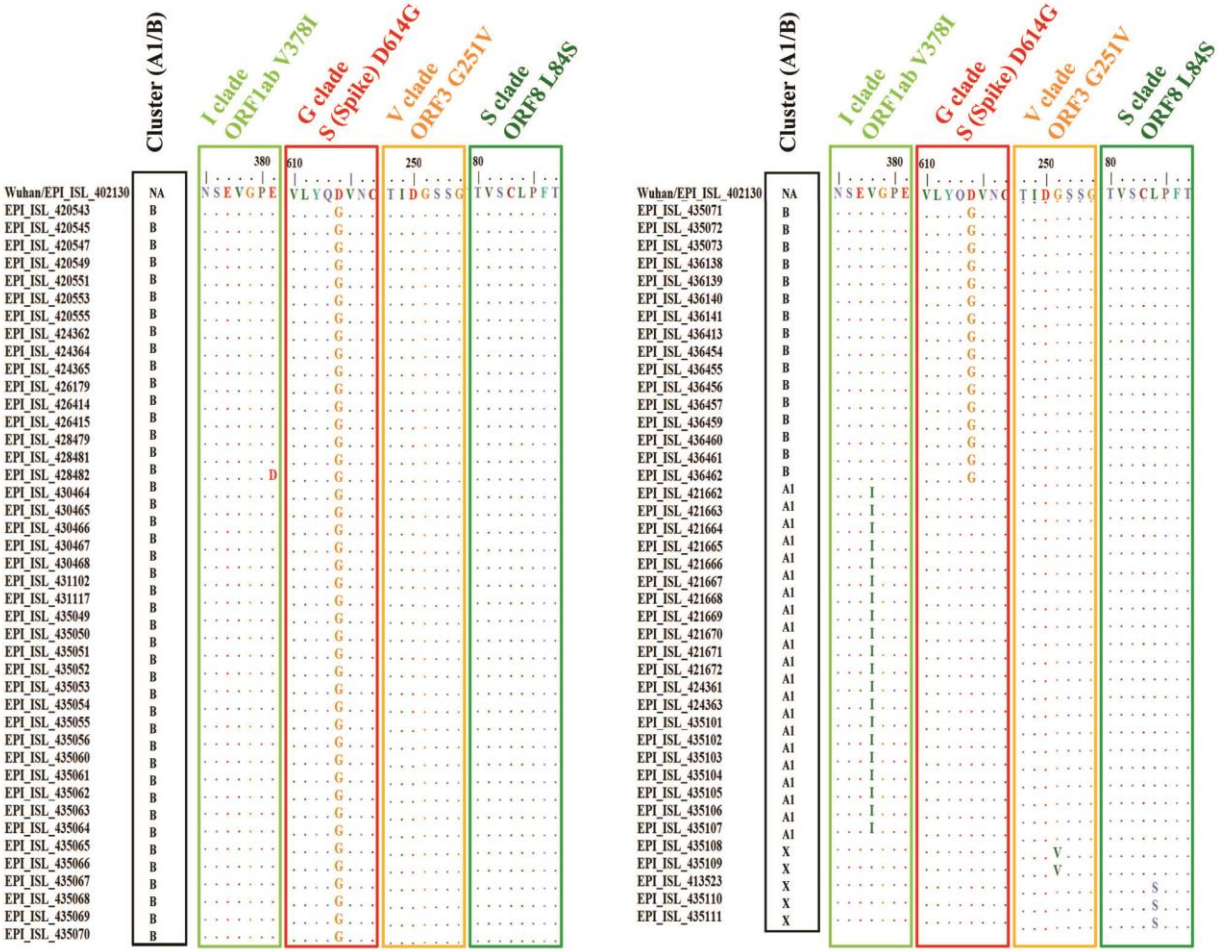


Figure 3. Characterization of clade defining genetic markers of 165viral sequences from the Indian isolates included in this study. Cluster (A1/B) shows the Indian samples belonging to the clusters identified by phylogeny analysis. I clade contains V378I marker in the ORF1ab region; G clade contains D614G marker in the spike protein (S); V clade contains G251V in ORF3 and S clade contains L84S marker in ORF8 region. NA: not applicable; X stands for the samples that do not belong to either A1 or B. Of the analysed Indian samples, 82 of them failed to classify with any of the four clades. The data for the remaining 83 samples is shown.

In particular, mutations in the receptor binding domain (RBD) of S protein are likely to create structural changes thus creating an escape mechanism from antibody recognition[23]. While the structural alterations created by D614G in S protein is not completely understood, it is proposed that the location of this change in the RBD makes it unlikely to affect critical epitopes to be used for vaccine development[23]. It has been also proposed that relaxation models of social distancing should consider the presence of one or more types of viral clades[15].

While this manuscript was under review, a higher occurrence of A2a (45.7%) and A3 (37.1%) types of SARS-CoV-2 in India was reported[24]. Upon comparison, it appears that A2a and A3 types are the same as G and I clades reported in this study. While our study with much higher number of viral genomes also reports the high occurrence of G clade (35.15%) followed by I clade (12.12%), a true picture on the proportion of different viral types in India will emerge only when more number of SARS-CoV-2 genomes are analysed. It is possible that the highly transmissible G clade type may become a dominant form occupying much large proportion in India in the next few months as it was reported in USA[18].

Conclusions

Collectively, we conclude that the G and I clades represent a significant proportion of Indian SARS-CoV-2 viruses based on this limited analysis. The probable source of origin of Indian SARS-CoV-2 viruses is countries from Europe and Oceania regions besides Middle East and South Asian regions. The possible spread of the SARS-CoV-2 virus to India through Middle East countries from Europe and Oceania regions cannot be ruled out. Indeed, both A and B clusters contain samples from Middle East countries. In addition, these samples appear to split between I and G clades (Supplementary Figure 6). In the absence of the information related to travel/contact history of Indian patients, more inference and definite

conclusions on the possible source of origin could not be made at present. Thus our result also indicates that there is a close connection between source of virus and the countries that are most travelled by Indians. The study also highlights the power of rapid viral genome sequencing and public data sharing to improve the detection and management of pandemic diseases such as COVID-19. It is important to point out that most countries in America, Europe, Oceania and East Asia were quick in supporting the advanced scientific studies on the virus and disease process itself in suitable containment facilities with appropriate ethical clearance towards developing novel treatment modalities and preventive vaccines. Needless to say that major countries from emerging economies such as Brazil and India should also support experimental research on SARS-CoV-2 pathogenesis.

Certainly, our analysis has clear limitations, the most important one being that we were able to analyse only a small number of Indian SARS-CoV-2 genomes while the number of COVID-19 cases increased beyond 60,000. Further, travel history of the patients and other clinical parameters are needed to make the conclusions definite. Hence, it is required that more number of Indian isolates of SARS-CoV-2 needs to be sequenced. Nevertheless, our study highlights the need for large-scale community surveillance for SARS-CoV-2 introductions and the spread. More importantly, this work underscores the power of pathogen genomics to identify epidemiological understanding of the virus and the disease.

Author contributions. M.M. and A.L. carried out data downloading and analysis. K.S. executed the whole study and wrote the manuscript.

1. Guo, Y. R. *et al.*, The origin, transmission and clinical therapies on coronavirus disease 2019 (COVID-19) outbreak – an update on the status. *Mil. Med. Res.*, 2020, **7**(1), 11; doi:10.1186/s40779-020-00240-0.
2. https://www.arcgis.com/
3. https://www.mohfw.gov.in/
4. https://www.gisaid.org/
5. Yadav, P. D. *et al.*, Full-genome sequences of the first two SARS-CoV-2 viruses from India. *Indian J. Med. Res.*, 2020; doi: 10.4103/ijmr.IJMR_663_20 [Epub ahead of print].
6. Sardar, R., Satish, D., Birla, S. and Gupta, D., Comparative analyses of SAR-CoV2 genomes from different geographical locations and other coronavirus family genomes reveals unique features potentially consequential to host–virus interaction and pathogenesis. bioRxiv: 2020.03.21.001586; doi:https://doi.org/10.1101/2020.03.21.001586.
7. Miller, M. A., Pfeiffer, W. and Schwartz, T., Creating the CIPRES science gateway for inference of large phylogenetic trees. In 2010 Gateway Computing Environments Workshop (GCE), 2010, pp. 1–8.
8. Trifinopoulos, J., Nguyen, L. T., von Haeseler, A. and Minh, B. Q., W-IQ-TREE: a fast online phylogenetic tool for maximum likelihood analysis. *Nucl. Acids Res.*, 2016, **44**(W1), W232–W235; https://doi.org/10.1093/nar/gkw256.
9. http://tree.bio.ed.ac.uk/software/figtree/
10. Madeira, F. *et al.*, The EMBL-EBI search and sequence analysis tools APIs in 2019. *Nuc. Acids Res.*, 2019, **47**(W1), W636–W641; doi:10.1093/nar/gkz268.
11. Hall, T. A., BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. *Nucl. Acids Symp. Ser.*, 1999, **41**, 95–98.
12. Zhou, P. *et al.*, A pneumonia outbreak associated with a new coronavirus of probable bat origin. *Nature*, 2020, **579**(7798), 270–273; doi:10.1038/s41586-020-2012-7. Epub 3 February 2020.
13. Elbe, S. and Buckland-Merrett, G., Data, disease and diplomacy: GISAID's innovative contribution to global health. *Global Chall.*, 2017, **1**(1), 33–46; doi:10.1002/gch2.1018. eCollection.
14. Eden, J. S. *et al.*, An emergent clade of SARS-CoV-2 linked to returned travellers from Iran. bioRxiv: 2020.03.15.992818; doi: https://doi.org/10.1101/2020.03.15.992818.
15. Brufsky, A., Distinct viral clades of SARS-CoV-2: implications for modeling of viral spread. *J. Med. Virol.*, 2020; doi: 10.1002/jmv.25902 [Epub ahead of print].
16. Wu, F. *et al.*, A new coronavirus associated with human respiratory disease in China. *Nature*, 2020, **579**, 265–269; https://doi.org/10.1038/s41586-020-2008-3.
17. Andersen, K. G., Rambaut, A., Lipkin, W. I., Holmes, E. C. and Garry, R. F., The proximal origin of SARS-CoV-2. *Nat. Med.*, 2020, **26**, 450–452; https://doi.org/10.1038/s41591-020-0820-9.
18. Korber, B. *et al.*, Spike mutation pipeline reveals the emergence of a more transmissible form of SARS-CoV-2. bioRxiv: 2020.05.05; https://doi.org/10.1101/2020.04.29.069054.
19. Bhattacharyya, C., Das, C., Ghosh, A., Singh, A. K., Mukherjee, S., Majumder, P. P., Basu, A. and Biswas, N. K., Global spread of SARS-CoV-2 subtype with spike protein mutation D614G is shaped by human genomic variations that regulate expression of TMPRSS2 and MX1 genes. bioRxiv: 2020.05.05; https://doi.org/10.1101/2020.05.04.075911.
20. Krueger, D. K., Kelly, S. M., Lewicki, D. N., Ruffolo, R. and Gallagher, T. M., Variations in disparate regions of the murine coronavirus spike protein impact the initiation of membrane fusion. *J. Virol.*, 2001, **75**(6), 2792–2802; doi:10.1128/JVI.75.6.2792-2802.2001.
21. Geoghegan, J. L. and Holmes, E. C., The phylogenomics of evolving virus virulence. *Nat. Rev. Genet.*, 2018, **19**, 756–769; https://doi.org/10.1038/s41576-018-0055-5.
22. Ontiveros, E., Enhanced virulence mediated by the murine coronavirus, mouse hepatitis virus strain JHM, is associated with a glycine at residue 310 of the spike glycoprotein. *J. Virol.*, 2003, **77**(19), 10260–10269; doi:10.1128/jvi.77.19.10260-10269.2003.
23. Dearlove, B. *et al.*, A SARS-CoV-2 vaccine candidate would likely match all currently circulating strains. bioRxiv: 2020.04.27; https://doi.org/10.1101/2020.04.27.064774.
24. Biswas, N. K. and Majumder, P. P., Analysis of RNA sequences of 3636 SARS-CoV-2 collected from 55 countries reveals selective sweep of one virus type. *Indian J. Med. Res.*, Special issue on COVID-19 (in press, 28 April 2020).

ACKNOWLEDGMENTS. The results reported here are in whole or part based upon data generated by Global Initiative on Sharing All Influenza Data (GISAID) (https://www.gisaid.org/). We gratefully acknowledge the authors, originating and submitting laboratories of the sequences from GISAID's EpiCoV™ Database. K.S. acknowledges DBT, DST-SERB and ICMR, Government of India for research grant. Infrastructure support by funding from DST-FIST, IISc-DBT partnership and UGC (Centre for Advanced Studies in Molecular Microbiology) to M.C.B. is acknowledged. K.S. is a J. C. Bose Fellow of the Department of Science and Technology.

Received 28 April 2020; revised accepted 22 May 2020

doi: 10.18520/cs/v118/i11/1786-1791

Published in Internet of Things, May 2020

PREDICTING THE GROWTH AND TREND OF COVID-19 PANDEMIC USING MACHINE LEARNING AND CLOUD COMPUTING

Shreshth Tuli
Department of Computer Science and Engineering
Indian Institute of Technology - Delhi
India
shreshthtuli@gmail.com

Shikhar Tuli
Department of Electrical Engineering
Indian Institute of Technology - Delhi
India
shikhartuli98@gmail.com

Rakesh Tuli
Department of Biotechnology Engineering
UIET, Panjab University, Chandigarh
India
rakeshtuli@hotmail.com

Sukhpal Singh Gill
School of EECS
Queen Mary University of London
United Kingdom
s.s.gill@qmul.ac.uk

May 4, 2020

ABSTRACT

The outbreak of COVID-19 Coronavirus, namely SARS-CoV-2, has created a calamitous situation throughout the world. The cumulative incidence of COVID-19 is rapidly increasing day by day. Machine Learning (ML) and Cloud Computing can be deployed very effectively to track the disease, predict growth of the epidemic and design strategies and policy to manage its spread. This study applies an improved mathematical model to analyse and predict the growth of the epidemic. An ML-based improved model has been applied to predict the potential threat of COVID-19 in countries worldwide. We show that using iterative weighting for fitting Generalized Inverse Weibull distribution, a better fit can be obtained to develop a prediction framework. This can be deployed on a cloud computing platform for more accurate and real-time prediction of the growth behavior of the epidemic. A data driven approach with higher accuracy as here can be very useful for a proactive response from the government and citizens. Finally, we propose a set of research opportunities and setup grounds for further practical applications. Predicted curves for some of the most affected countries can be seen at https://collaboration.coraltele.com/covid/.

Keywords COVID-19; Machine Learning; SARS-CoV-2; Coronavirus; Prediction; Cloud Computing

1 Introduction

The novel Coronavirus disease (COVID-19) was first reported on 31 December 2019 in the Wuhan, Hubei Province, China. It started spreading rapidly across the world [1]. The cumulative incidence of the causitive virus (SARS-CoV-2) is rapidly increasing and has affected 196 countries and territories with USA, Spain, Italy, U.K. and France being the most affected [2]. World Health Organization (WHO) has declared the coronavirus outbreak a pandemic, while the virus continues to spread [3]. As on 4 May 2020, a total of 3,581,884 confirmed positive cases have been reported leading to 248,558 deaths [2]. The major difference between the pandemic caused by CoV-2 and related viruses, like SARS and MERS is the ability of CoV-2 to spread rapidly through human contact and leave nearly 20% infected subjects as

[0]**Abbreviations:** ML, Machine Learning; SARS-CoV-2; Severe Acute Respiratory Syndrome Coronavirus 2, COVID-19, Coronavirus disease

symptom-less carriers [4]. Moreover, various studies reported that the disease caused by CoV-2 is more dangerous for people with weak immune system. The elderly people and patients with life threatening diseases like cancer, diabetes, neurological conditions, coronary heart disease and HIV/AIDS are more vulnerable to severe effects of COVID-19 [5]. In the absence of any curative drug, the only solution is to slow down the spread by exercising "social distancing" to block the chain of spread of the virus. This behavior of CoV-2 requires developing robust mathematical basis for tracking its spread and automation of the tracking tools for on line dynamic decision making.

There is a need for innovative solutions to develop, manage and analyse big data on the growing network of infected subjects, patient details, their community movements, and integrate with clinical trials and, pharmaceutical, genomic and public health data [6]. Multiple sources of data including, text messages, online communications, social media and web articles can be very helpful in analyzing the growth of infection with community behaviour . Wrapping this data with ML and Artificial Intelligence (AI), researchers can forecast where and when, the disease is likely to spread, and notify those regions to match the required arrangements. Travel history of infected subjects can be tracked automatically, to study epidemiological correlations with the spread of the disease. Some community transmission based effects have been studied in other works (`https://www.cdc.gov/mmwr/volumes/69/wr/mm6915e1.htm`). Infrastructure for the storage and analytics of such huge data for further processing needs to be developed in an efficient and cost-effective manner. This needs to be organized through utilization of cloud computing and AI solutions [7]. Alibaba developed cloud and AI solutions to help China, fight against coronavirus, predict the peak, size and duration of the outbreak, which is claimed to have been implemented with 98% accuracy in real world tests in various regions of China [8]. FDifferent types of pneumonia can be resolved using ML-based CT Image Analytics Solution, which can be helpful to monitor the patients with COVID-19 [9]. Details can be seen at `https://spectrum.ieee.org/the-human-os/biomedical/imaging/hospitals-deploy-ai-tools-detect-covid19-chest-scans`. The development of vaccine for COVID-19 can also be accelerated by analysing the genome sequences and molecular docking, deploying various ML and AI techniques [10].

Motivation and Our Contributions: ML [11] can be utilized to handle large data and intelligently predict the spread of the disease. Cloud computing [12] can be used to rapidly enhance the prediction process using high-speed computations [7]. Novel energy-efficient edge systems can be used to procure data, in order to bring down power consumption. In this paper, we present a prediction model deployed using FogBus framework [13] for accurate prediction of the number of COVID-19 cases, the rise and the fall of the number of cases in near future and the date when various countries may expect the pandemic to end. We also provide a detailed comparison with a baseline model and show how catastrophic the effects can be if poorly fitting models are used. We present a prediction scheme based on the ML model, which can be used in remote cloud nodes for real-time prediction allowing governments and citizens to respond proactively. Finally, we summarize this work and present various research directions.

Article structure: The rest of the paper is organized as follows: Section 2 presents the prediction model and performance comparison. Section 3 concludes the work and describes the future research opportunities. Section 4 provides details of open repositories for the dataset, code and results.

2 Prediction Model and Performance Comparison

Machine Learning (ML) and Data Science community are striving hard to improve the forecasts of epidemiological models and analyze the information flowing over Twitter for the development of management strategies, and the assessment of impact of policies to curb its spread. Various datasets in this regard have been openly released to the public. Yet, there is a need to capture, develop and analyse more data as the COVID-19 grows worldwide [14, 15].

The novel coronavirus is leaving a deep socio-economic impact globally. Due to the ease of virus transmission, primarily through droplets of saliva or discharge from the nose when an infected person coughs or sneezes, countries which are densely populated need to be on a higher alert [16]. To gain more insight on how COVID-19 is impacting the world population and to predict the number of COVID-19 cases and dates when the pandemic may be expected to end in various countries, we propose a Machine Learning model that can be run continuously on Cloud Data Centers (CDCs) for accurate spread prediction and proactive development of strategic response by the government and citizens.

Dataset: The dataset used in this case study is the Our World in Data by Hannah Ritchie[1]. The dataset is updated daily from the World Health Organization (WHO) situation reports[2]. More details about the dataset are available at: `https://ourworldindata.org/coronavirus-source-data`.

[1]Our World In Data: COVID-19 Dataset; source: `https://github.com/owid/covid-19-data/tree/master/public/data/`

[2]Situation Reports: WHO; source: `https://www.who.int/emergencies/diseases/novel-coronavirus-2019/situation-reports`

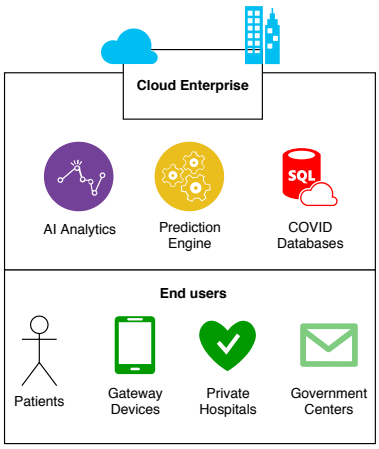


Figure 1: Proposed Cloud based AI framework for COVID-19 related analytics.

Cloud framework: The ML models are built to make a good advanced prediction of the number of new cases and the dates when the pandemic might end. To provide fail-safe computation and quick data analysis, we propose a framework to deploy these models on cloud datacenters, as shown in Figure 1.In a cloud based environment, the government hospitals and private health-centers continuously send their positive patient count. Population density, average and median age, weather conditions, health facilities etc. are also to be integrated for enhancing the accuracy of the predictions. For this case study, we used three instances of single core *Azure B1s* virtual machines with 1-GiB RAM, SSD Storage and 64-bit Microsoft Windows Server 2016[1]. We used the HealthFog [11] framework leveraging the FogBus [13] for deploying multiple analysis tasks in an ensemble learning fashion to predict various metrics, like the number of anticipated facilities to manage patients and the hospitals. We analyzed that the cost of tracking patients on a daily basis, amortized CPU consumption and Cloud execution is 37% and only 1.2 USD per day. As the dataset size increases, computationally more powerful resources would be needed.

ML model: Many recent works have suggested that the COVID-19 spread follows exponential distribution [17, 18, 19]. As per empirical evaluations and previous datasets on SARS-CoV-1 virus pandemic, many sources have shown that data corresponding to new cases with time has large number of outliers and may or may not follow a standard distribution like Gaussian or Exponential [20, 21, 22, 23]. In recent study by Data-Driven Innovation Laboratory, Singapore University of Technology and Design (SUTD)[3], the regression curves were drawn using the Susceptible-Infected-Recovered model [24] and Gaussian distribution was deployed to estimate the number of cases with time. However, in the previously reported studies on the earlier version of the virus, namely SARA-CoV-1, the data was reported to follow Generalized Inverse Weibull (GIW) Distribution [25] better than Gaussian as shown in Figure 2 (details of Robust Weibull fitting follow in the next section). Detailed comparison for SARS-CoV-2 has been described in the next section. This fits the following function to the data:

$$f(x) = k \cdot \gamma \cdot \beta \cdot \alpha^{\beta} \cdot x^{-1-\beta} \cdot \exp\big(-\gamma(\frac{\alpha}{x})^{\beta}\big). \tag{1}$$

Here, $f(x)$ denotes the number of cases with x, where $x > 0$ is the time in number of days from the first case, and $\alpha, \beta, \gamma > 0, \in \mathbb{R}$ are parameters of the model. Now, we can find the appropriate values of the parameters α, β and γ to minimize the error between the predicted cases ($y = f(x)$) and the actual cases ($\hat{y}$). This can be done using the popular Machine Learning technique of Levenberg-Marquardt (LM) for curve fitting [26]. However, as various sources have suggested, in initial stages of COVID-19 the data has many outliers and noise. This makes it hard to accurately predict the number of cases. Thus, we propose an iterative weighting strategy and call our fitting technique "Robust Fitting". A diagrammatic representation of the iterative weighting process is shown in Figure 3.

The main idea is as follows. We maintain weights for all data points (i) in every iteration (n, starting from 0) as w_i^n. First, we fit a curve using the LM technique with weights of all data points as 1, thus $w_i^0 = 1 \; \forall \; i$. Second, we find the weight corresponding to every point for the next iteration (w_i^{n+1})) as:

[1]Azure Cloud VMs: `https://azure.microsoft.com/en-au/pricing/calculator/`

[3]When Will COVID-19 End, DDI Lab, SUTD: https://ddi.sutd.edu.sg/when-will-covid-19-end

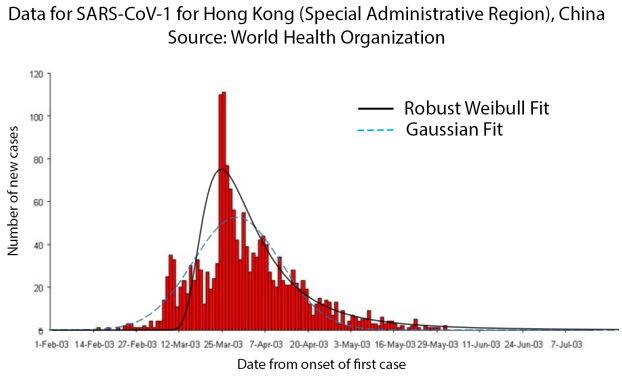


Figure 2: Fit curves for SARS-CoV-1 pandemic for Hong Kong (SAR), China. Data source: WHO epidemic curves (`https://www.who.int/csr/sars/epicurve/epiindex/en/index4.html`)

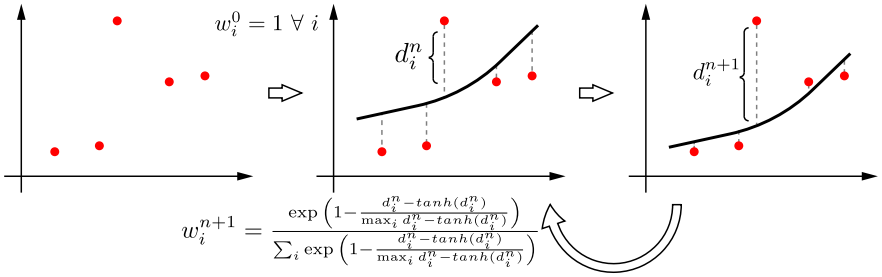


Figure 3: Iterative weighting technique for robust curve fitting.

$$w_i^{n+1} = \frac{\exp\left(1 - \frac{d_i^n - tanh(d_i^n)}{\max_i d_i^n - tanh(d_i^n)}\right)}{\sum_i \exp\left(1 - \frac{d_i^n - tanh(d_i^n)}{\max_i d_i^n - tanh(d_i^n)}\right)}. \tag{2}$$

Simply, in the above equation, we first take $tanhshrink$ function defined as $tanhshrink(x) = x - tanh(x)$ for the distances of all points along y axis from the curve (d_i). This is to have a higher value for points far from the curve and near 0 value for closer points. This, is then standardized by dividing with max value over all points and subtracted from 1 to get a weight corresponding to each point. This weight is then standardized using $softmax$ function so that sum of all weights is 1. The curve is fit again using LM method, now with the new weights w_i^{n+1}. The algorithm converges when the sum total deviation of all weights becomes lower than a threshold value.

Distribution Model Selection: To find the best fitting distribution model for the data corresponding to COVID-19, we studied the data on daily new confirmed COVID cases. Five sets of global data on daily new COVID-19 cases were used to fit parameters of different types of distributions. Finally, we identified the best performing 5 distributions. The results are shown in Table 1. We observe that using the iteratively weighted approach, the Inverse Weibull function fits the best to the COVID-19 dataset, as compared to the iterative versions of Gaussian, Beta (4-parameter), Fisher-Tippet (Extreme Value distribution), and Log Normal functions. When applied to the same dataset, Iterative Weibull showed an average MAPE of 12% lower than non-iteratively weighted Weibull. A step-by-step algorithm for iteratively weighted curve fitting using the GIW distribution (called "Robust Weibull") is given in Algorithm 1.

Analysis and Interpretation: To compare the proposed "Robust Weibull fitting" model, we use the baseline proposed by Jianxi Luo from SUTD [3]. The comparison metrics include Mean Squared Error (MSE), Mean Absolute Percentage Error (MAPE) and Coefficient of determination (R^2). Table 2 shows the model predictions of the spread of the COVID-19 for every major country for which sufficient data was available and model fits had $R^2 > 0.5$ using the proposed model. As shown in the table, the proposed model performs significantly better than the baseline.

Country	MSE					R^2					MAPE				
	Weibull	Gaussian	Beta 4	Fisher-Tippet	Log Normal	Weibull	Gaussian	Beta 4	Fisher-Tippet	Log Normal	Weibull	Gaussian	Beta 4	Fisher-Tippet	Log Normal
World	**2.41E+07**	3.78E+07	2.99E+07	2.89E+07	2.99E+07	**0.98**	0.97	0.98	0.97	0.97	49.14	49.14	50.39	48.12	**46.19**
India	6.97E+03	7.09E+03	6.89E+03	**6.89E+03**	7.00E+03	0.97	0.97	**0.98**	0.97	0.97	**18.33**	18.33	18.49	21.69	20.69
United States	**8.37E+06**	1.11E+07	8.63E+05	9.47E+06	9.78E+06	**0.95**	0.93	0.94	0.93	0.94	**24.33**	24.33	40.23	71.64	111.63
United Kingdom	**2.00E+05**	2.22E+05	2.12E+05	2.02E+05	2.07E+05	**0.95**	0.95	0.95	0.95	0.95	21.46	21.46	20.43	21.52	**17.42**
Italy	**1.56E+05**	3.38E+05	2.10E+05	2.09E+05	2.35E+05	**0.96**	0.92	0.95	0.95	0.94	**14.98**	14.98	20.00	19.62	170.63

Table 1: Preliminary Evaluation of different models. We observe that iterative fitting of Inverse Weibull performs significantly better than iterative fitting of other distributions like Gaussian, Beta (4-parameter), Fisher-Tippet (Extreme Value distribution), and Log Normal. The lowest value of MSE/MAPE and highest values of R^2 among all distributions are shown in bold.

Algorithm 1 Robust Curve Fitting using Iterative weighting

Require:

x : Input sequence of days from first case

y : Number of cases for each day in x

ϵ : Threshold parameter

procedure ROBUST CURVE FITTING

$w^0 \leftarrow$ Unit vector $[1] \times size(x)$

for iteration n from 0, step 1 **do**

$f \leftarrow$ LM(input = x, target = y, weights = w^n)

$d_i \leftarrow |f(x_i) - y_i| \; \forall \; i$

$$w_i^{n+1} \leftarrow \frac{\exp\left(1-\frac{d_i^n - tanh(d_i^n)}{\max_i d_i^n - tanh(d_i^n)}\right)}{\sum_i \exp\left(1-\frac{d_i^n - tanh(d_i^n)}{\max_i d_i^n - tanh(d_i^n)}\right)}$$

if $\sum_i |w_i^n - w_i^{n+1}| < \epsilon$ **then**

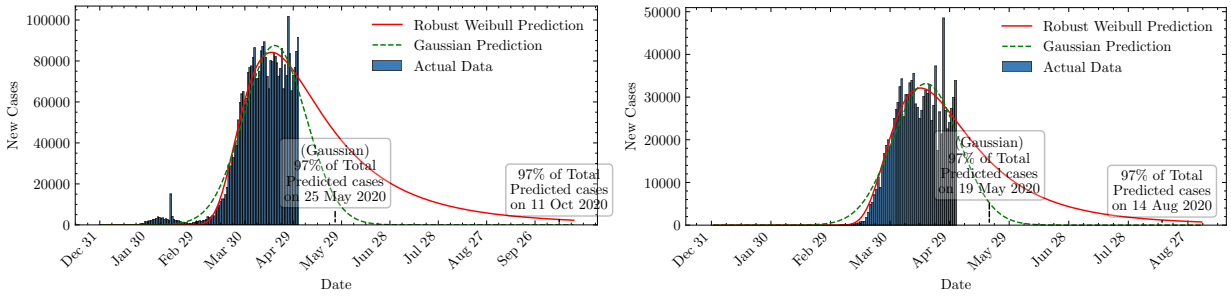
break

end for

end procedure

As shown in Figure 4[4], the predictions of the baseline Gaussian model deployed by SUTD are overoptimistic. Following such models could lead to premature uplifting of the lockdown, causing adverse effect on management of the epidemic. Having better fit models, as proposed here, could help plan a better strategy, based on more accurate predictions and future scenarios.

Figure 5 shows the total predicted number of cases for all countries across the globe. Here we have neglected those countries where the data is insufficient for making predictions, or the number of days for data is less than 30. As shown in Figure 4 explained in model section, the fit curve can be used to predict the number of cases that will have to be dealt by the country, assuming the same trend continues. The figure illustrates that the maximum number of total cases will be in the North America region. The number of cases will also be high in the European continent, Russia and eastern Asia, including China, the original epicenter of the disease.



(i) World (ii) United States

Figure 4: Comparison of predicted dates to reach 97% of the total expected cases by baseline Gaussian and proposed Robust Weibull models. The predicted end date of the pandemic in the baseline model are over-optimistic.

[4]Curves and predictions of all countries have been given in Appendix

Country	Predictions of Robust Weibull Model			Fit comparison metrics					
	Total Cases	Date of last case	97% cases date	MSE (W)	MSE (G)	R^2 (W)	R^2 (G)	MAPE (W)	MAPE (G)
United States	1,937,724	11-Feb-22	14-Aug-20	**9.32E+06**	1.33E+07	**0.95**	0.92	**26.58**	1568.56
Russia	529,687	27-Nov-21	26-Sep-20	**5.50E+04**	5.92E+04	**0.99**	0.98	**24.53**	75.91
India	409,418	29-Oct-24	13-Aug-21	**8.40E+03**	9.11E+03	**0.97**	0.97	**22.38**	80.47
United Kingdom	331,124	31-Jul-21	18-Aug-20	**2.54E+05**	3.19E+05	**0.95**	0.93	**20.14**	211.72
Ukraine	254,087	10-Dec-41	18-Jan-31	**3.31E+04**	3.37E+04	**0.53**	0.52	**1842.80**	2079.70
Italy	253,022	7-Mar-21	27-Jun-20	**1.52E+05**	3.55E+05	**0.96**	0.91	**14.55**	1577.95
Spain	236,737	30-Sep-20	20-Apr-20	**4.67E+05**	6.59E+05	**0.93**	0.90	3682.04	**2917.90**
Turkey	234,218	22-Jun-23	30-Dec-20	2.27E+05	**1.49E+05**	0.92	**0.94**	**30.95**	555.87
Germany	181,369	17-Oct-20	2-May-20	**3.39E+05**	4.50E+05	**0.91**	0.88	1013.65	**582.89**
France	147,795	11-Oct-20	29-May-20	**3.93E+05**	4.14E+05	**0.84**	0.83	**32.36**	134.36
Qatar	143,779	1-Oct-22	18-Mar-21	3.71E+03	**3.46E+03**	**0.93**	0.93	99.14	**90.02**
Canada	139,331	7-Dec-21	31-Oct-20	**2.12E+04**	2.64E+04	**0.95**	0.94	**28.19**	210.56
Belarus	135,375	4-Jun-22	2-Feb-21	**1.28E+04**	1.28E+04	**0.83**	0.83	**1040.39**	1101.48
Iran	126,048	12-Mar-21	15-Jul-20	2.11E+05	**1.88E+05**	0.78	**0.80**	**1847.80**	2313.85
China	84,171	6-Jul-20	27-Mar-20	1.40E+06	**1.28E+06**	0.48	**0.53**	**114.04**	202.88
Sweden	68,671	25-Apr-22	15-Feb-21	**5.33E+03**	5.45E+03	**0.91**	0.91	**20.55**	151.04
Belgium	65,257	19-Nov-20	27-Jun-20	**3.88E+04**	4.10E+04	**0.88**	0.88	**18.76**	134.34
Bangladesh	53,127	19-Apr-22	22-Feb-21	**1.38E+03**	1.60E+03	**0.96**	0.96	**30.89**	118.80
Netherlands	53,057	28-Nov-20	2-Jul-20	**1.07E+04**	1.10E+04	**0.94**	0.94	**16.45**	140.98
United Arab Emirates	46,395	18-Jul-21	5-Nov-20	**3.30E+03**	3.49E+03	**0.91**	0.90	**840.72**	947.02
Portugal	37,302	7-Jun-21	12-Sep-20	**2.62E+04**	3.20E+04	**0.75**	0.70	**41.46**	222.40
Indonesia	35,581	19-Sep-21	20-Dec-20	1.24E+03	**1.22E+03**	**0.93**	0.93	**51.96**	124.06
Poland	35,113	22-Nov-22	1-Aug-21	**3.08E+03**	3.42E+03	**0.87**	0.86	**29.90**	110.45
Switzerland	31,407	26-Jul-20	13-May-20	**1.14E+04**	1.39E+04	**0.92**	0.90	**383.82**	476.28
Bahrain	30,258	21-Mar-23	12-Jan-22	1.05E+03	**1.04E+03**	**0.57**	0.57	**98.32**	102.20
Ireland	27,694	6-Sep-20	12-Jun-20	1.17E+04	**9.49E+03**	0.84	**0.87**	25.91	**21.83**
Singapore	24,088	19-Jul-20	28-May-20	**1.68E+04**	1.69E+04	0.82	**0.82**	**912.31**	1018.38
Dominican Republic	22,193	19-Jun-21	29-Apr-20	**1.79E+03**	1.85E+03	**0.81**	0.81	**304.96**	420.08
Romania	22,102	24-Dec-20	10-Aug-20	**1.98E+03**	2.29E+03	**0.91**	0.90	**16.83**	87.83
Algeria	19,188	6-Feb-22	16-May-21	**3.59E+02**	3.96E+02	**0.86**	0.85	**61.43**	147.61
Israel	18,167	3-Aug-20	26-May-20	**8.81E+03**	1.03E+04	**0.80**	0.77	**37.91**	137.87
Japan	17,614	27-Jul-20	29-May-20	1.12E+04	**1.08E+04**	0.74	**0.75**	**162.70**	202.00
Morocco	16,972	24-May-22	2-Aug-21	1.60E+03	**1.40E+03**	**0.69**	0.73	188.07	**171.78**
Serbia	16,426	24-Jan-21	27-Aug-20	2.49E+03	**2.36E+03**	**0.87**	0.87	**210.28**	229.10
Austria	15,781	9-Jun-20	30-Apr-20	**4.07E+03**	5.33E+03	**0.92**	0.89	**23.08**	34.08
Philippines	14,371	24-Nov-20	2-Apr-20	**5.08E+03**	5.49E+03	**0.65**	0.62	**543.57**	698.02
Denmark	13,282	26-Oct-20	17-Jul-20	1.94E+03	**1.81E+03**	0.81	**0.82**	**18.95**	104.47
Moldova	12,818	6-Feb-22	12-Jun-21	**8.78E+02**	9.57E+02	**0.75**	0.73	**36.51**	68.69
Hungary	11,077	19-Jul-22	22-Nov-21	5.85E+02	**5.66E+02**	0.64	**0.65**	**49.55**	71.00
South Korea	10,780	4-May-20	2-Apr-20	**3.35E+03**	3.88E+03	**0.87**	0.85	**55.81**	68.84
Finland	9,158	21-Dec-20	5-Sep-20	**9.09E+02**	9.11E+02	**0.74**	0.74	**125.43**	188.74
Norway	8,534	23-Jul-20	19-Apr-20	**1.73E+03**	1.79E+03	**0.80**	0.79	**187.65**	211.88
Czech Republic	8,528	14-Jul-20	22-May-20	**1.34E+03**	1.56E+03	**0.85**	0.83	**20.11**	59.31
Malaysia	7,080	6-Aug-20	7-Jun-20	**4.88E+02**	5.73E+02	**0.89**	0.87	**30.30**	112.20
Australia	6,797	17-May-20	21-Apr-20	**2.54E+03**	2.78E+03	**0.81**	0.79	**31.85**	36.77
Oman	4,871	4-Sep-20	23-Apr-20	6.01E+02	**5.98E+02**	**0.66**	0.66	**229.07**	232.19
Iraq	4,113	20-Nov-20	20-Apr-20	**4.99E+02**	5.21E+02	**0.47**	0.45	**299.89**	354.05
Luxembourg	3,887	29-May-20	2-May-20	**5.15E+02**	6.64E+02	**0.83**	0.79	**49.42**	127.81
Thailand	3,044	30-May-20	2-Apr-20	**8.51E+02**	9.01E+02	**0.63**	0.61	**381.04**	399.02
Greece	2,944	7-Jul-20	28-Apr-20	3.69E+02	**3.67E+02**	**0.66**	0.66	**137.87**	127.28
Croatia	2,275	15-Jun-20	18-May-20	**8.44E+01**	1.04E+02	**0.88**	0.85	**20.64**	45.14
World	6,734,075	29-Jan-24	11-Oct-20	**2.91E+07**	4.92E+07	**0.98**	0.96	**47.53**	63.36

Table 2: **Predictions and error comparisons.** Country wise predictions using Robust Weibull model and error comparison between Robust Weibull and baseline Gaussian Model. We predict the total number of cases that will be reached, and the last case date i.e. when the model predicts new cases < 1. We also predict the date when the total number will reach 97% of the total expected cases. Such data is critical to prepare the healthcare services in advance. The fit comparison metrics (with proposed model as W and baseline model as G) show that Mean Square Error (MSE) and the Mean Absolute Percentage Error (MAPE) of the proposed model are lower than baseline for most cases. The coefficient of determination (R^2) is higher for the proposed model for most of the countries. The least MSE/MAPE and highest R^2 values among the two models are shown in bold. Data upto 4 May, 2020 was used to create these results.

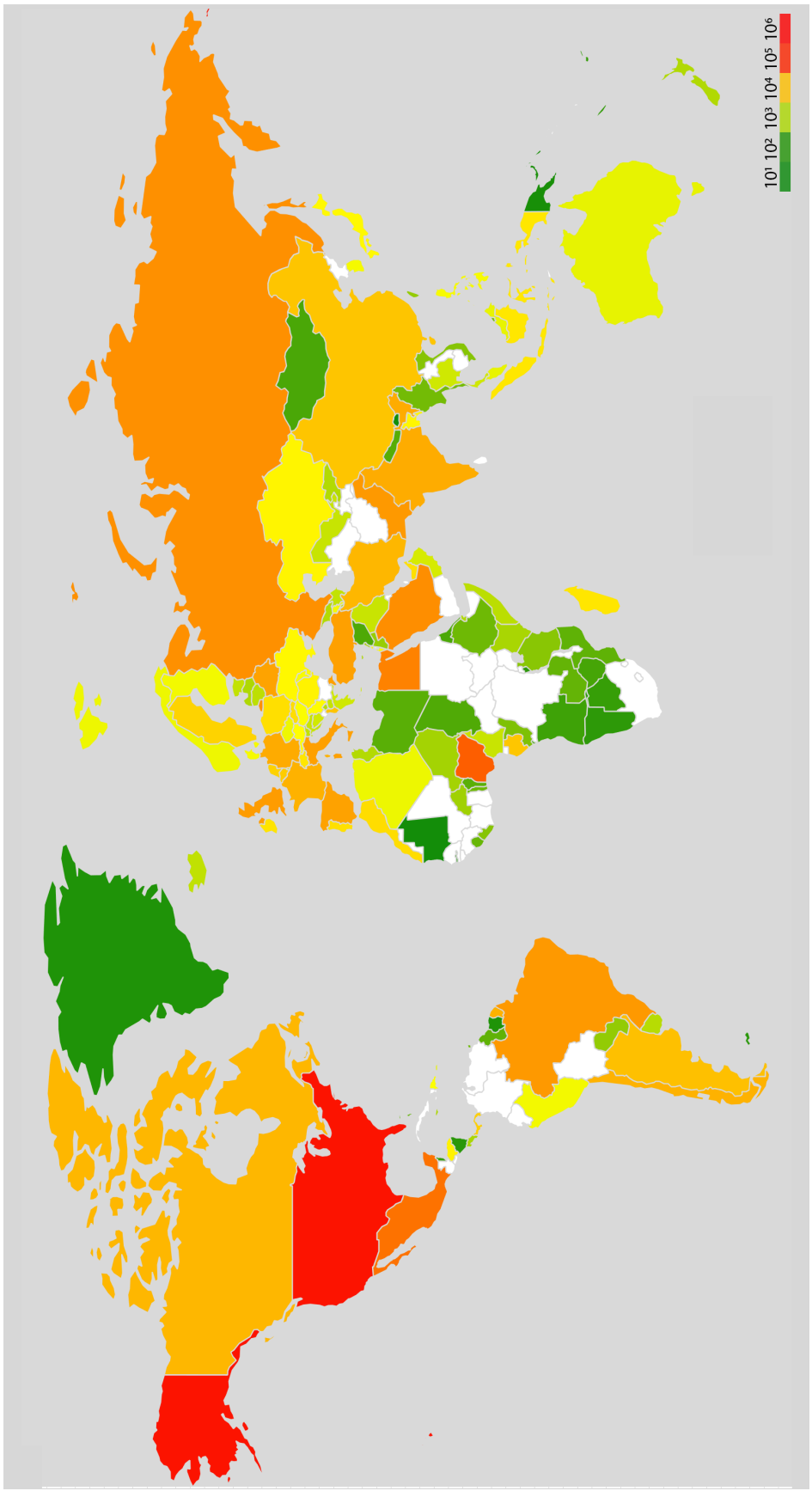


Figure 5: Global heat-map for total predicted cases for different countries as on May 4, 2020 (countries with insufficient data for prediction are shown in white)

Figure 6: Future Research Directions and Open Challenges

3 Conclusions and Future Directions

In this study, we have discussed how improved mathematical modelling, Machine Learning and cloud computing can help to predict the growth of the epidemic proactively. Further, a case study has been presented which shows the severity of the spread of CoV-2 in countries worldwide. Using the proposed Robust Weibull model based on iterative weighting, we show that our model is able to make statistically better predictions than the baseline. The baseline Gaussian model shows an over-optimistic picture of the COVID-19 scenario. A poorly fitting model could lead to a non optimal decision making, leading to worsening of public health situation.

We propose the future directions as follows. Firstly, other important parameters like population density, distribution of age, individual and community movements, level of healthcare facilities available, strain type and virulence of the virus etc., need to be included in the regression model to further enhance the prediction accuracy. Secondly, models like ARIMA [27] can be integrated with Weibull function for further time series analysis and predictions. Thirdly, ML can be utilized to predict the structure and function of various proteins associated with CoV-2 and their interaction with the host human proteins and cellular environment. The contribution of various socio-economic variables that determine the vulnerability, spread and progression of the epidemic can be predicted by developing suitable algorithms. AI based proactive measures can be taken to prevent the spread of the virus to sensitive groups in the society. Real time sensors can be used, for example in traffic camera or surveillance, which track COVID-19 symptoms based on visual imaging and tracking Apps, and inform respective hospitals and administrative authorities for punitive action [28]. Tracking needs to cover all stages from ports of entries to public places and hospitals [29]. The research directions and challenges are summarized in Figure 6.

4 Software Availability

Our prediction model is available online at `https://github.com/shreshthtuli/covid-19-prediction`. The dataset used for this work is the *Our World Dataset*, available at `https://github.com/owid/covid-19-data/tree/master/public/data/`. Few interactive graphs can be seen at `https://collaboration.coraltele.com/covid/`.

References

[1] Chen Wang, Peter W Horby, Frederick G Hayden, and George F Gao. A novel coronavirus outbreak of global health concern. *The Lancet*, 395(10223):470–473, 2020.

[2] Coronavirus - worldometer, link: `https://www.worldometers.info/coronavirus/`, [online accessed] 12 april 2020.

[3] Guangdi Li and Erik De Clercq. Therapeutic options for the 2019 novel coronavirus (2019-ncov), 2020.

[4] Smriti Mallapaty. What the cruise-ship outbreaks reveal about covid-19. *Nature*, 580(7801):18–18, 2020.

[5] Kai Liu, Ying Chen, Ruzheng Lin, and Kunyuan Han. Clinical features of covid-19 in elderly patients: A comparison with young and middle-aged patients. *Journal of Infection*, 2020.

[6] Shi Zhao, Qianyin Lin, Jinjun Ran, Salihu S Musa, Guangpu Yang, Weiming Wang, Yijun Lou, Daozhou Gao, Lin Yang, Daihai He, et al. Preliminary estimation of the basic reproduction number of novel coronavirus (2019-ncov) in china, from 2019 to 2020: A data-driven analysis in the early phase of the outbreak. *International Journal of Infectious Diseases*, 92:214–217, 2020.

[7] Shreshth Tuli, Shikhar Tuli, Gurleen Wander, Praneet Wander, Sukhpal Singh Gill, Schahram Dustdar, Rizos Sakellariou, and Omer Rana. Next generation technologies for smart healthcare: Challenges, vision, model, trends and future directions. *Internet Technology Letters*, page e145.

[8] Chaolin Huang, Yeming Wang, Xingwang Li, Lili Ren, Jianping Zhao, Yi Hu, Li Zhang, Guohui Fan, Jiuyang Xu, Xiaoying Gu, et al. Clinical features of patients infected with 2019 novel coronavirus in wuhan, china. *The Lancet*, 395(10223):497–506, 2020.

[9] Adrien Depeursinge, Anne S Chin, Ann N Leung, Donato Terrone, Michael Bristow, Glenn Rosen, and Daniel L Rubin. Automated classification of usual interstitial pneumonia using regional volumetric texture analysis in high-resolution ct. *Investigative radiology*, 50(4):261, 2015.

[10] Maxwell W Libbrecht and William Stafford Noble. Machine learning applications in genetics and genomics. *Nature Reviews Genetics*, 16(6):321–332, 2015.

[11] Shreshth Tuli, Nipam Basumatary, Sukhpal Singh Gill, Mohsen Kahani, Rajesh Chand Arya, Gurpreet Singh Wander, and Rajkumar Buyya. Healthfog: An ensemble deep learning based smart healthcare system for automatic diagnosis of heart diseases in integrated iot and fog computing environments. *Future Generation Computer Systems*, 104:187–200, 2020.

[12] Sukhpal Singh Gill, Shreshth Tuli, Minxian Xu, Inderpreet Singh, Karan Vijay Singh, Dominic Lindsay, Shikhar Tuli, Daria Smirnova, Manmeet Singh, Udit Jain, et al. Transformative effects of iot, blockchain and artificial intelligence on cloud computing: Evolution, vision, trends and open challenges. *Internet of Things*, 8:100118, 2019.

[13] Shreshth Tuli, Redowan Mahmud, Shikhar Tuli, and Rajkumar Buyya. Fogbus: A blockchain-based lightweight framework for edge and fog computing. *Journal of Systems and Software*, 2019.

[14] Emily Chen, Kristina Lerman, and Emilio Ferrara. Covid-19: The first public coronavirus twitter dataset. *arXiv preprint arXiv:2003.07372*, 2020.

[15] Joseph Paul Cohen, Paul Morrison, and Lan Dao. Covid-19 image data collection. *arXiv preprint arXiv:2003.11597*, 2020.

[16] Yan Bai, Lingsheng Yao, Tao Wei, Fei Tian, Dong-Yan Jin, Lijuan Chen, and Meiyun Wang. Presumed asymptomatic carrier transmission of covid-19. *Jama*, 2020.

[17] Benjamin F Maier and Dirk Brockmann. Effective containment explains sub-exponential growth in confirmed cases of recent covid-19 outbreak in mainland china. *arXiv preprint arXiv:2002.07572*, 2020.

[18] Yi Li, Meng Liang, Xianhong Yin, Xiaoyu Liu, Meng Hao, Zixin Hu, Yi Wang, and Li Jin. Covid-19 epidemic outside china: 34 founders and exponential growth. *medRxiv*, 2020.

[19] Mary Raygoza. Covid-19, exponential growth, and the power of showing up in social solidarity: The math behind the virus. 2020.

[20] Yanping Bai and Zhen Jin. Prediction of sars epidemic by bp neural networks with online prediction strategy. *Chaos, Solitons & Fractals*, 26(2):559–569, 2005.

[21] Ying-Hen Hsieh, Jen-Yu Lee, and Hsiao-Ling Chang. Sars epidemiology modeling. *Emerging infectious diseases*, 10(6):1165, 2004.

[22] Dejian Lai. Monitoring the sars epidemic in china: a time series analysis. *Journal of Data Science*, 3(3):279–293, 2005.

[23] Wendi Wang and Shigui Ruan. Simulating the sars outbreak in beijing with limited data. *Journal of theoretical biology*, 227(3):369–379, 2004.

[24] David Smith and Lang Moore. The sir model for spread of disease: The differential equation model. *Loci.(originally Convergence.) https://www. maa. org/press/periodicals/loci/joma/the-sir-model-for-spread-of-disease-the-differential-equation-model*, 2004.

[25] Felipe RS De Gusmao, Edwin MM Ortega, and Gauss M Cordeiro. The generalized inverse weibull distribution. *Statistical Papers*, 52(3):591–619, 2011.

[26] Jorge J Moré. The levenberg-marquardt algorithm: implementation and theory. In *Numerical analysis*, pages 105–116. Springer, 1978.

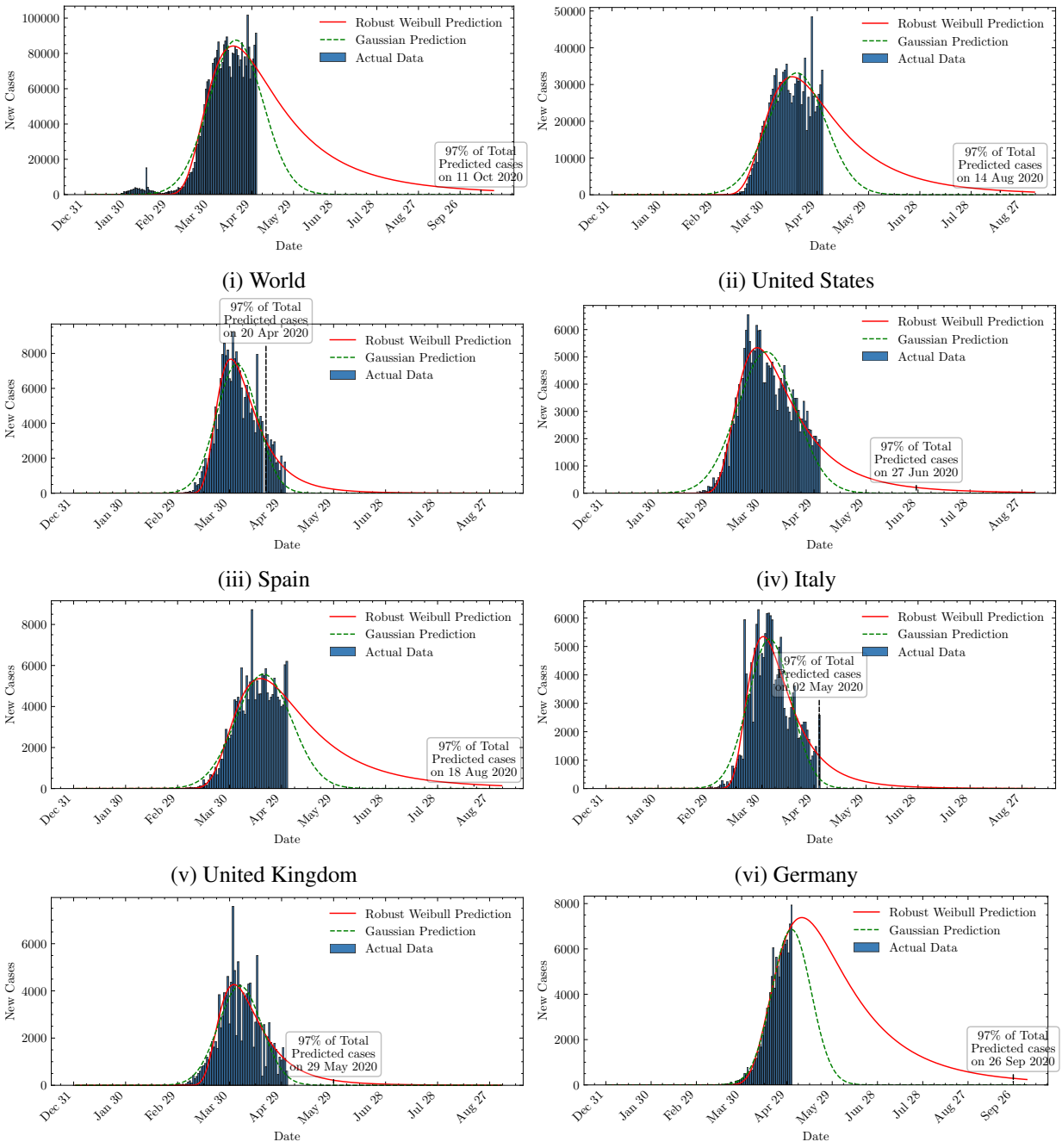
[27] Domenico Benvenuto, Marta Giovanetti, Lazzaro Vassallo, Silvia Angeletti, and Massimo Ciccozzi. Application of the arima model on the covid-2019 epidemic dataset. *Data in brief*, page 105340, 2020.

[28] Halgurd S Maghdid, Kayhan Zrar Ghafoor, Ali Safaa Sadiq, Kevin Curran, and Khaled Rabie. A novel ai-enabled framework to diagnose coronavirus covid 19 using smartphone embedded sensors: Design study. *arXiv preprint arXiv:2003.07434*, 2020.

[29] Shuai Wang, Bo Kang, Jinlu Ma, Xianjun Zeng, Mingming Xiao, Jia Guo, Mengjiao Cai, Jingyi Yang, Yaodong Li, Xiangfei Meng, et al. A deep learning algorithm using ct images to screen for corona virus disease (covid-19). *medRxiv*, 2020.

Appendix: Real data from WHO with predicted curves of proposed and baseline models

All data uptil 4 May 2020 has been used to generate results.



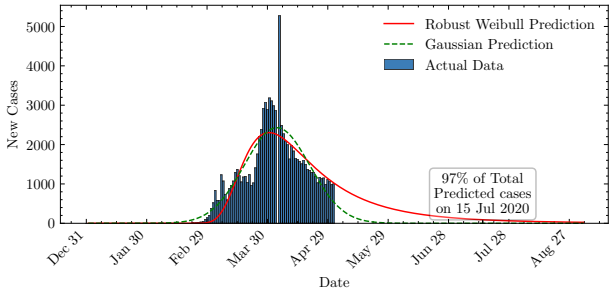
(i) World (ii) United States

(iii) Spain (iv) Italy

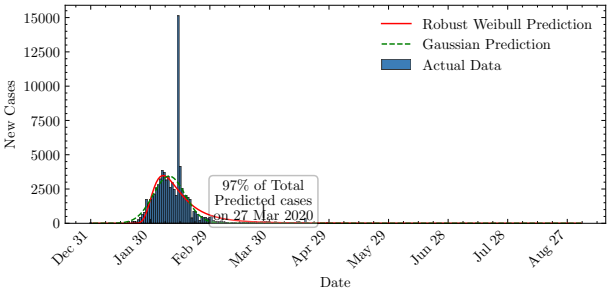
(v) United Kingdom (vi) Germany

(vii) France (viii) Russia

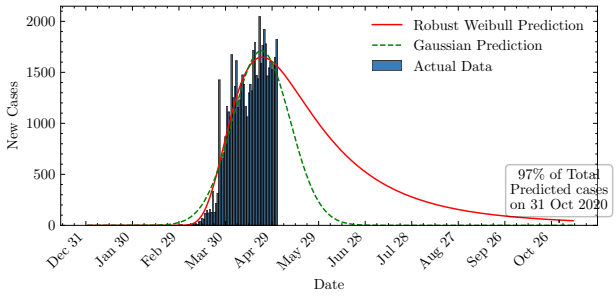
Figure 7: New cases for different countries (continued on next page)



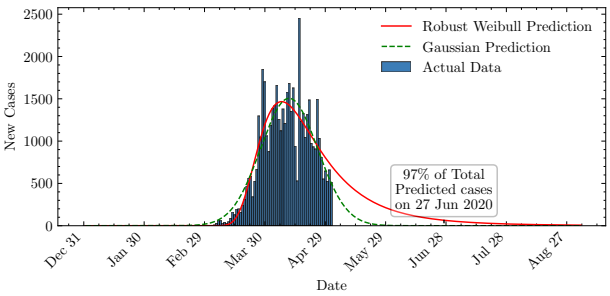
(ix) Iran



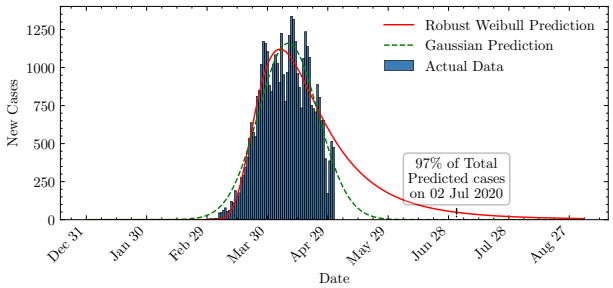
(x) China



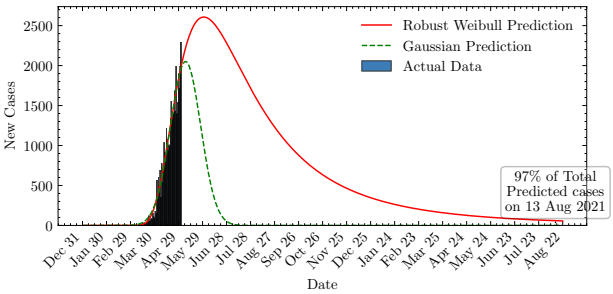
(xi) Canada



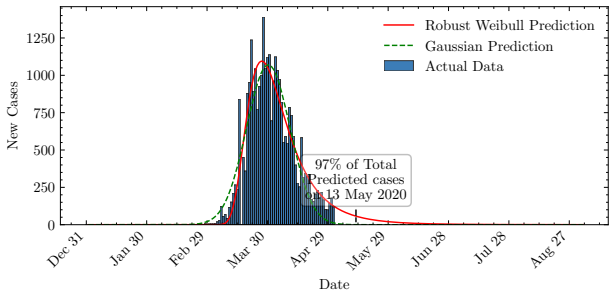
(xii) Belgium



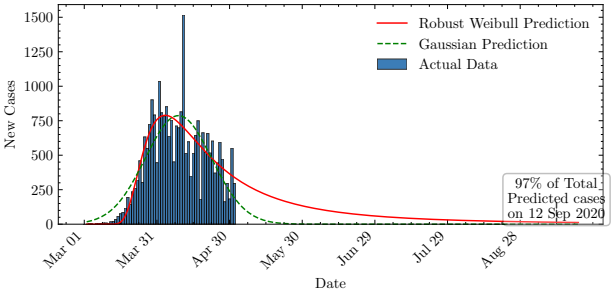
(xiii) Netherlands



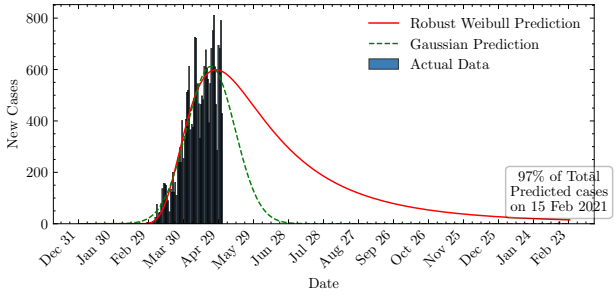
(xiv) India



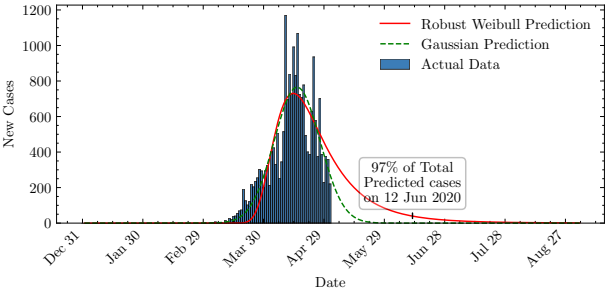
(xv) Switzerland



(xvi) Portugal

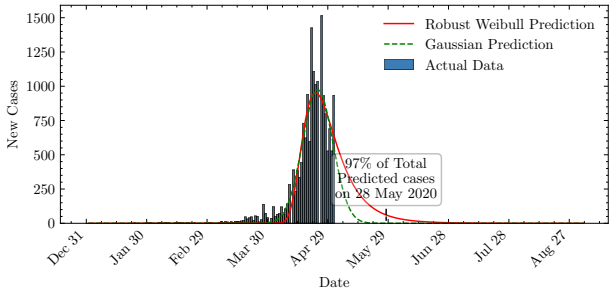


(xvii) Sweden

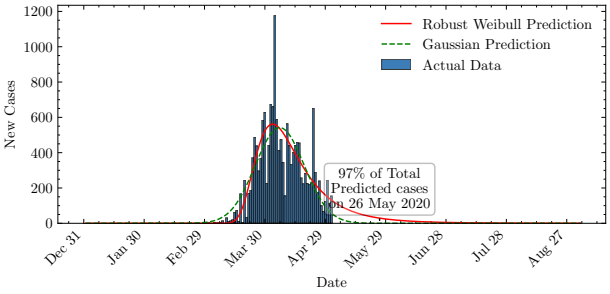


(xviii) Ireland

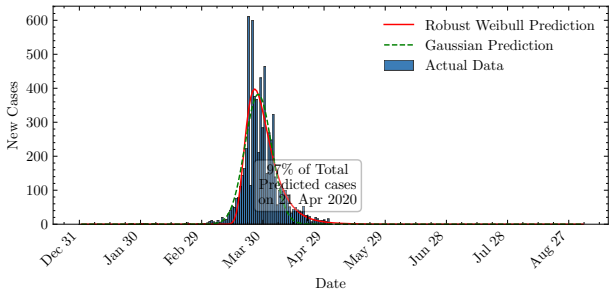
Figure 7: New cases for different countries (continued on next page)



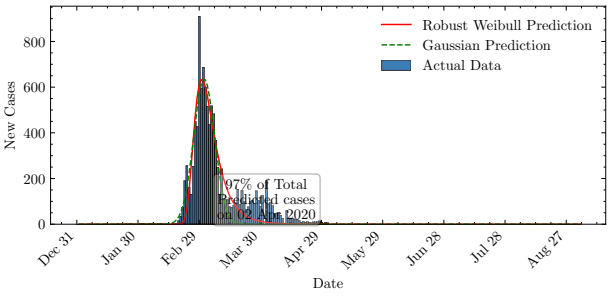
(xix) Singapore



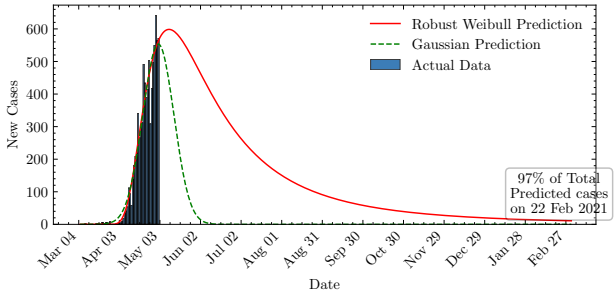
(xx) Israel



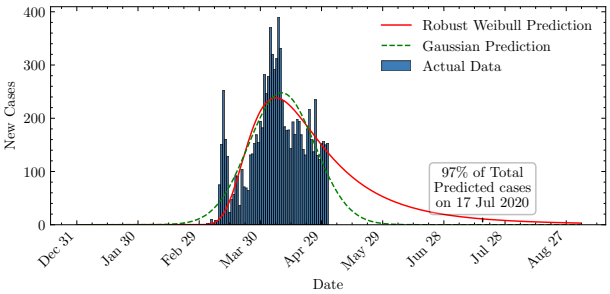
(xxi) Australia



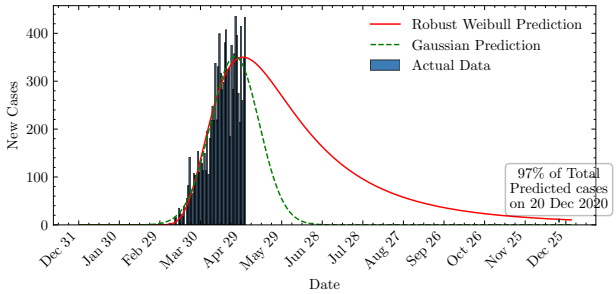
(xxii) South Korea



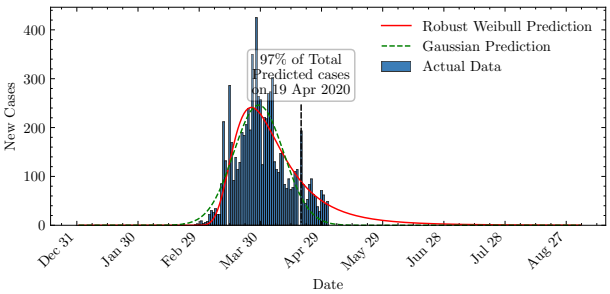
(xxiii) Bangladesh



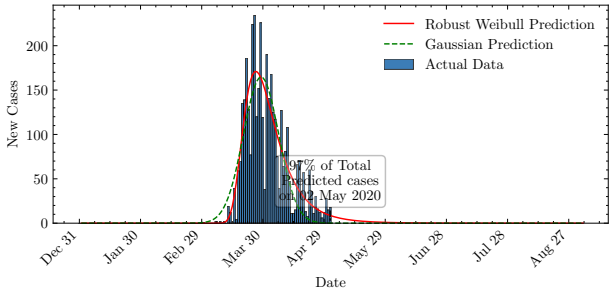
(xxiv) Denmark



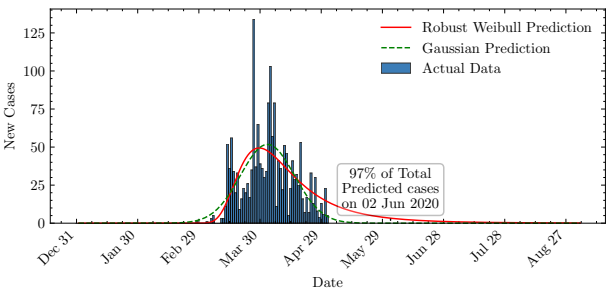
(xxv) Indonesia



(xxvi) Norway



(xxvii) Luxembourg



(xxviii) Estonia

Figure 7: New cases for different countries

Author's Biography

Shreshth Tuli is an undergraduate student at the Department of Computer Science and Engineering at Indian Institute of Technology - Delhi, India. He is also a co-founder of Qubit Inc. company which works on providing next generation solutions for industrial problems. He is a national level Kishore Vaigyanic Protsahan Yojana (KVPY) scholarship holder for excellence in science and innovation. He has worked as a visiting researcher at the Cloud Computing and Distributed Systems (CLOUDS) Laboratory, Department of Computing and Information Systems, the University of Melbourne, Australia. His research interests include Internet of Things (IoT), Fog Computing, Network Design, and Artificial Intelligence.

Shikhar Tuli is an undergraduate student at the Department of Electrical Engineering at Indian Institute of Technology - Delhi, India. He is the founder and CEO of Qubit Inc. He has worked remotely with the Cloud Computing and Distributed Systems (CLOUDS) Laboratory, Department of Computing and Information Systems, the University of Melbourne, Australia in the realization of the FogBus framework. He has also worked at the Embedded Systems Laboratory, EPFL, Switzerland in the design of low-power and physics-optimized Edge devices made from emerging Non-Volatile Memories. His research interests include Internet of Things (IoT), In-memory and Neuromorphic computing architectures and Nanoelectronics. He specializes in designing novel hardware technologies that are valuable to both industry and academia.

Rakesh Tuli is Senior Research Advisor; J C Bose National Fellow, UIET, Panjab University, Chandigarh. Before this, he was executive Director, National Agri-Food Biotech Institute, Mohali; and Director, National Botanical Research Institute, Lucknow. He has many publications in reputed journals and conferences including Nature Biotechnology. His research includes Genomics and Transgenic Approaches to Improving Plants for Agricultural and Health/Medicinal Applications.

Sukhpal Singh Gill is a Lecturer (Assistant Professor) in Cloud Computing at School of Electronic Engineering and Computer Science, Queen Mary University of London, UK. Prior to this, Dr. Gill has held positions as a Research Associate at the School of Computing and Communications, Lancaster University, UK and also as a Postdoctoral Research Fellow at CLOUDS Laboratory, The University of Melbourne, Australia. His research interests include Cloud Computing, Fog Computing, Software Engineering, Internet of Things and Big Data.

Hydroxychloroquine: A relatively obscure antimalarial takes centre stage in COVID-19

Govindarajan Padmanaban and Viswanathan Arun Nagaraj

Hydroxychloroquine (HCQ) is an old antimalarial known for its use to treat rheumatoid arthritis. Recently, it has been shown to be useful in COVID-19 treatment, although its efficacy is being debated and the data from large-scale clinical trials are yet to be available. Overall, it appears that HCQ-Azithromycin-Remdesevir may be a viable treatment option on a case to case basis. There is also a case to evaluate curcumin as an adjunct food supplement to prevent and treat COVID-19.

Spanish Flu was the deadliest virus infection in the 20th century. It emerged in March 1918 during the First World War and came in three waves infecting 500 million people worldwide and claiming the lives of 20–50 million people. In 21st century, we are witnessing a similar pandemic affecting almost every country in the world. At the time of writing this article, COVID-19 virus (SARS-CoV-2) has infected close to 2.6 million people, killing more than 170,000 people. The figures keep changing by the hour. This virus emerged in December 2019 from Wuhan, China, although one does not really know whether it incubated in China much longer. It has taken the globe by storm, and one hopes that with advanced medical care and clampdowns, COVID-19 would be contained sooner than later. There have been constant updates through print media and TV channels (24×7). An unprecedented effort has also gone into the development of medical equipment, diagnostics, drugs and vaccines. At the last count, 115 vaccine trials are underway. In terms of drug development, we should think of testing every kind of chemical with therapeutic potential including antiretrovirals, anticancer drugs, natural molecules, interferon-α/β, combination therapies, etc. Plasma from recovered patients and monoclonal antibodies against the viral proteins are additional therapeutic options. We will confine to the hydroxychloroquine (HCQ) story of how a relatively obscure antimalarial has taken centre stage to treat COVID-19 and provide our views on the potential of nutraceuticals like curcumin as adjunct therapy for COVID-19, as it falls within our area of research interest.

History of chloroquine use

The story of HCQ (also known as Plaquenil) has come full circle. Chloroquine (CQ, Aralen) was introduced to treat the US army at home and combat areas with severe malaria problems during World War II in the 1940s. HCQ, considered less toxic than chloroquine, was introduced in the mid-1950s. These two drugs belong to the class of 4-aminoquinolines (Figure 1). The two main *Plasmodia* species causing malaria are *Plasmodium falciparum* and *Plasmodium vivax*. *P. falciparum* is more deadly and accounts for 90% of the deaths, especially in Africa. *P. vivax* causes morbidity and incapacitation for prolonged periods and is reported to become more virulent in recent years. In the 1970s, *P. falciparum* became resistant to chloroquine and artemisinin derivative-based combination therapies (ACTs) have become the choice to treat malaria patients, although CQ is still effective against *P. vivax* malaria[1].

HCQ and COVID-19

It was shown that CQ has strong antiviral effect against SARS-CoV-1 infected primate cells in *in vitro* cultures[2]. In a recent study, Chinese authors have shown that Remdesevir (a nucleoside analogue) and CQ could inhibit SARS-CoV-2 at EC_{90} concentrations of 1.76 and 6.90 μM respectively, in vero cell culture[3]. This is followed by another publication from Chinese authors, who claimed the following: 'Chloroquine phosphate, an old drug for treatment of malaria, is shown to have apparent efficacy and acceptable safety against COVID-19 associated pneumonia in multicenter clinical trials conducted in China. The drug is recommended to be included in the next version of the Guidelines for the Prevention, Diagnosis, and Treatment of Pneumonia Caused by COVID-19 issued by the National Health Commission of the People's Republic of China for treatment of COVID-19 infection in larger populations in the future'. The authors indicated that the study was performed in 10 hospitals in Wuhan with more than 100 patients[4]. In another study with French patients, it has been shown that HCQ was effective in clearing virus load in 50% of the patients (7/14 on day 5 post therapy), HCQ-Azithromycin was effective in 100% of the patients (6/6), with untreated controls showing clearance in 18.8% of the patients (3/16)[5]. Azithromycin is a macrolide antibiotic, known to inhibit bacterial protein synthesis.

Given this background, CQ/HCQ has been added to the treatment protocols in some countries including China, South Korea, Belgium and India with varying levels of conditionality, although large-scale clinical data from randomized trials are not yet available. Reports have indicated that UK has decided to wait for the outcome of the trials and Swedish hospitals have abandoned trials on CQ

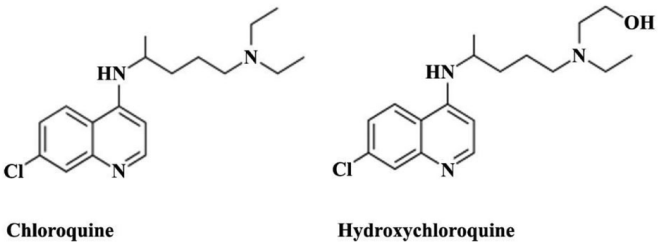


Figure 1. Structures of chloroquine and hydroxychloroquine.

because of side effects. In a manuscript submitted to the *New England Journal of Medicine* from a study carried out with 32 patients admitted to the hospital in Detroit, HCQ administration was found to increase the need for ventilator support. There were no benefits on mortality, lymphopenia orneurtophil–lumphocyte ratio improvement. However, no data was provided on clearance of viral load[6]. In a very recent study, the same French group mentioned earlier (led by Didier Raoult), has released data with the following findings: 'From 3 March to 9 April 2020, 59,655 specimens from 38,617 patients were tested for COVID-19 by PCR. Of the 3,165 positive patients placed in the care of our institute, 1061 previously unpublished patients met our inclusion criteria. Their mean age was 43.6 years old and 492 were male (46.4%). No cardiac toxicity was observed. A good clinical outcome and virological cure was obtained in 973 patients within 10 days (91.7%). Prolonged viral carriage at completion of treatment was observed in 47 patients (4.4%) and was associated to a higher viral load at diagnosis ($p < 10^{-2}$), but viral culture was negative at day 10 and all but one were PCR-cleared at day 15. A poor outcome was observed for 46 patients (4.3%); 10 were transferred to intensive care units, 5 patients died (0.47%) (74–95 years old) and 31 required 10 days of hospitalization or more. Among this group, 25 patients are now cured and 16 are still hospitalized (98% of patients cured so far). Poor clinical outcome was significantly associated to older age (OR 1.11), initial higher severity (OR 10.05) and low hydroxychloroquine serum concentration. In addition, both poor clinical and virological outcomes were associated to the use of selective beta-blocking agents and angiotensin II receptor blockers ($P < 0.05$). Mortality was significantly lower in patients who had received >3 days of HCQ-AZ than in patients treated with other regimens both at IHU and in all Marseille public hospitals ($p < 10^{-2}$). Interpretation: The HCQ-AZ combination, when started immediately after diagnosis, is a safe and efficient treatment for COVID-19, with a mortality rate of 0.5%, in elderly patients. It avoids worsening and clears virus persistence and contagiosity in most cases.' In another recent study, Shangai Jiao Tong University School of Medicine has shown in a randomized trial that HCQ limits the inflammatory response and hastens alleviation of symptoms. Over all, based on published and unpublished results, it looks HCQ-Azithromycin could be a potential treatment option in the clinical management of COVID-19 disease.

Unfortunately, a potential therapy to save lives has become a political hotbed with US President claiming it as a 'Game Changer' and the latest results by the French group being released to the French President! It is not clear whether the drumbeat is just to claim priority or there is something deeper in terms of Chinese versus US/Europe conspiracy theories.

Mechanism of action of CQ/HCQ

CQ and HCQ, although generally known as antimalarial and antiamoebic drugs, have been reported to have a wide range of clinically relevant effects. The non-protonated form of chloroquine can cross the endosomal membrane within a cell and can get protonated in the acidic compartment. The protonated CQ is held within the endosome, raising the pH. Some viruses, including flaviviruses, retroviruses and coronaviruses enter the target cells by endocytosis. The acidic pH in the endosomal (lysosomal) compartment leads to disruption of the viral particle with the release of the nucleic acid and enzymes necessary for viral replication (Figure 2). CQ prevents the pH-dependent entry of the virus and the viral replication. CQ was proposed as a therapy for SARS virus and some preliminary results were obtained using *in vitro* cultures[7]. However, CQ did not seem to have undergone any clinical trial for SARS-CoV-1. Perhaps, SARS did not spread like COVID-19 and reports indicate that SARS-CoV infected around 8000 people with almost 700 deaths. Further, these drugs have immunomodulatory effects and they can suppress the release of inflammatory cytokines (TNF-α and IL-6), with a potential to inhibit the inflammatory consequences of viral infections and other chronic diseases. Thus, HCQ and CQ have been used in the treatment of rheumatoid arthritis, systemic lupus erythematosis and other inflammatory diseases, although the mechanisms involved are still being analysed[8].

In terms of the anti-malarial action of CQ, the effects have been understood based on the toxicity of free heme to the parasite. The equivalent of the endosome/lysosome compartment is referred to as 'food vacuole' in the parasite, in which the haemoglobin taken up by the parasite from the host red cell is degraded to give rise to amino acids and heme. While the amino acids are utilized

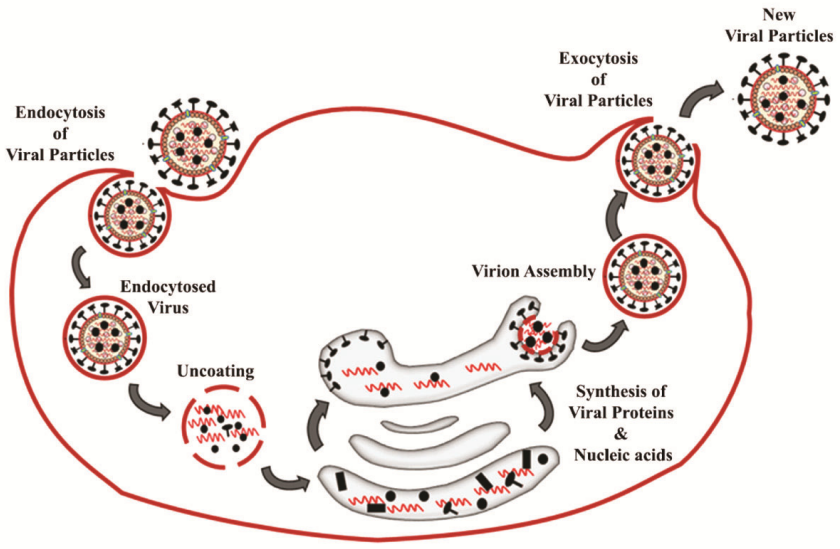


Figure 2. Steps involved in the entry and assembly of corona and related viruses. SARS-CoV-2 has a positive stranded RNA genome and is about 30 kb in size. It codes for many proteins needed for viral replication and viral assembly, including the characteristic spike protein that binds to the ACE-2 receptor of the host and facilitates the virus entry into the host cell through endocytosis. Figure based on ref. 7.

by the parasite for its own protein synthesis, free heme is highly toxic to the parasite. Therefore, the parasite converts free heme into a polymer called hemozoin – a brown pigment that accumulates in the parasite. CQ binds to free heme and prevents the conversion of free heme into hemozoin, leading to the death of the parasite[9]. As an extension of this mechanism, an interesting argument has been made for the treatment of COVID-19 with vitamin C in the perspective of acute respiratory distress syndrome (ARDS). A detailed review records the basis for this proposition[10]. ARDS patients have a 30–50% mortality rate. This is due to an uncontrolled cascading event starting with increased permeability of pulmonary capillary endothelial cell and leakage of fluid into the pulmonary parenchyma. This is followed by cytokine storms marked by acute inflammatory responses[11]. It has been proposed that the ARDS seen in critically ill COVID-19 patients is due to cell-free haemoglobin in the lung air space[12]. Hemolysis leads to free heme generation that contains iron in ferric (Fe^{3+}) state and this needs to be reduced and kept in the ferrous (Fe^{2+}) state. Ascorbic acid (not made by the human body) is involved in this process and is regenerated through a cyclic mechanism involving cytochrome b561. It has been suggested that oral ascorbic acid can save the lives of critically ill COVID-19 patients[10].

Safety of CQ/HCQ

Despite the use of CQ as an anti-malarial for decades and HCQ for rheumatoid arthritis, it can cause side effects and overdose can even be fatal. These drugs have common as well as independent side effects as described in the medical literature[13,14]. Therefore, these are prescription drugs and self-medication needs to be strictly avoided.

Potential of curcumin as an adjunct food supplement to prevent and treat COVID-19

Several studies in our laboratory have demonstrated the beneficial effects of curcumin from turmeric as an adjunct nutraceutical in preventing malaria parasite recrudescence and cerebral malaria in the animal model along with the antimalarial, artemisinin derivative[15–17]. Preliminary studies were initially carried out with chloroquine–curcumin combination, but detailed studies were performed later on with artemisinin derivative, as it is a more effective antimalarial in the mouse model. Curcumin has immunomodulatory effects and has anti-oxidative and anti-inflammatory properties. Formulations to increase the bioavailability of curcumin are already available. A clinical trial to test the efficacy of curcumin in malaria has been approved by DCGI as well. Here, we would like to propose the evaluation of curcumin as a simple food supplement to prevent as well as treat COVID-19 along with approved therapy such as HCQ, Azithromycin and Remdesevir. Finally, while vaccines and drugs would be very useful to prevent and treat the infection, a pandemic of this dimension can only be controlled through public health measures such as social distancing, quarantine and personal hygiene. This has been the experience in earlier corona virus infections as well.

1. Yeung, S., Pongtavornpinyo, W., Hastings, I. M., Mills, A. J. and White. N. J., *Am. J. Trop. Med. Hyg*., 2004, **71**, 179–186.
2. Vincent, M. J. *et al*., *Virology J*., 2005, **2**, 69; doi:10.1186/1743-422X-2-69.
3. Wang, M. *et al*., *Cell Res*., 2020, **30**, 269–271.
4. Gao, J., Tian, Z. and Yang, X., *Biosci. Trends*, 2020, **14**, 72–73; doi:10.5582/bst.2020.01047.
5. Gautret, P. *et al*., *Int. J. Antimicrobial Agents,* 2020, **20**, 105949; doi: 10.1016/j.ijantimicag.2020.105949.
6. Barbosa, J. *et al*., *N. Eng. J. Med*., 4 April 2020 (submitted).
7. Savarino, A., Boelaert, J. R., Cassone, A., Majori, G. and Cauda, R., *Lancet Infect. Dis*., 2003, **3**, 722–727; doi: 10.1016/s1473-3099(03)00806-5.
8. Schrezenmeier, E. and Dörner, T., *Nat. Rev.*, 2020, **16**, 155–166; doi:10.1038/s41584-020-0372-x.
9. Fitch, C. D., *Trans. Am. Clin. Climatol. Assoc*., 1998, **109**, 97–106.
10. Loh, D., 24 March 2020; https://www.evolutamente.it/covid-19-ards-cell-free-hemoglobin-the-ascorbic-acid-connection/
11. Gonzales, J. N., Lucas, R. and Verin, A. D., *Austin J. Vasc. Med*., 2015, **2**, pii; 1009.
12. Shaver, C. M. *et al*., *Am. J. Physiol. Lung Cell Mol. Physiol*., 2016, **310**, L532–L541; doi:10.1152/ajplung.00155.2015.
13. University of Illinois-Chicago, Drug Information Group. Hydroxychloroquine, oral tablet. Healthline. Medically reviewed by Alex Brewer on 23 March 2020; https://www.healthline.com/health/hydroxychloroquine-oral-tablet
14. Cuhna, J. P., https://www.rxlist.com/aralen-side-effects-drug-center.htm
15. Vathsala, P. G. *et al*., *PLoS ONE*, 2012, **7**, e29442; doi:10.1371/journal.pone.0029442.
16. Padmanaban, G. and Rangarajan, P. N., *Trends Pharmacol. Sci*., 2016, **37**, 1–3.
17. Dende, C. *et al*., *Sci. Rep*., 2017, **7**, 10062; doi:10.1038/s41598-017-10672-9.

ACKNOWLEDGEMENT. Studies on curcumin in the malaria parasite were carried out in a Centre of Excellence project supported by the Department of Biotechnology, Government of India.

Govindarajan Padmanaban is in the Department of Biochemistry, Indian Institute of Science, Bengaluru 560 012, India; Viswanathan Arun Nagaraj is in the Infectious Disease Biology, Institute of Life Sciences, Nalco Square, Bhubaneswar 751 023, India.*
**e-mail: geepee@alumni.iisc.ac.in*

India's Role in the Race to Deliver COVID-19 Vaccines

By Prof. Virander Singh Chauhan
Former Director of the International Centre
for Genetic Engineering and Biotechnology
Currently holds the Arturo Falaschi Chair

As one of the world's largest vaccine manufacturers, India is poised to be a key player in distributing these products.

The extensive international efforts to develop vaccines to prevent COVID-19, and the speed with which these efforts are progressing, is truly unprecedented. There are now more than 130 candidates in development, according to the World Health Organization (WHO). Remarkably, 10 of these candidates are already in clinical trials, while the others are in various stages of preclinical development.

Three vaccine candidates, one each from the US, UK, and China, have already completed Phase I trials. While the efficacy of these vaccines in preventing COVID-19 still needs to be evaluated, the good news is that there have been few or no adverse reactions in humans, and this paves the way for larger trials. Every vaccine platform available to scientists, novel or well-established, is being explored. Some of these platforms, such as DNA and RNA, have not yet been used in licensed vaccines, which underscores the innovative approaches being applied to developing vaccines against this deadly virus.

Given India's enormous progress in producing and delivering vaccines, the country is prepared to play a major role in the development and manufacturing of a COVID-19 vaccine. There are now almost 30 groups in India – from large vaccine manufacturersto individual academic research groups – working to develop and manufacture vaccines against this novel coronavirus.

Indian vaccine developers are currently exploring several vaccine platforms including a novel DNA vaccine, inactivated rabies virus vector platforms, and protein-based vaccine candidates, as well as repurposing existing vaccines such as the recombinant BCG vaccine. Government funding has played a pivotal role in these efforts, alongside the investment by major vaccine manufacturing companies. Simultaneously, efforts are underway to accelerate regulatory and manufacturing processes to ensure that when a vaccine is available, it reaches the population as quickly as possible.

Experience in Vaccine Development and Delivery

India is currently one of the largest manufacturers of vaccines in the world in terms of volume. In 2016, under the Universal Immunization Program, it became the first country in the region to introduce an indigenously developed and produced vaccine against rotavirus, one of the leading causes of death in children younger than five. In 2018, this vaccine, and another rotavirus vaccine made in India, received WHO prequalification. By 2019, rotavirus immunizations were scaled up across the country, protecting thousands of children against this potentially fatal disease. That same year another vaccine made in India for pneumococcal pneumonia was also prequalified by WHO.

Many of these indigenous vaccines are considerably less expensive than those already on the market. In fact, one of the indigenous rotavirus vaccines costs less than a dollar a dose, making it more affordable for low- and middle-income countries to access this product.

India's vaccine delivery infrastructure has also improved considerably over the last decade. This is in part due to polio-elimination efforts, which succeeded in 2014 in a monumental public health victory. Nationwide surveillance networks and mass vaccination campaigns established in an effort to eliminate polio have resulted in improved cold-chain systems and intensive on-site and door-to-door outreach programs that connect hard-to-reach areas and/or migrant populations to immunizations programs. This helped strengthen India's National Immunization Program, which now delivers life-saving vaccines to the majority of the nation's more than 25 million children born every year, resulting in significantly reduced child mortality.

Lessons from these efforts – especially how to deliver vaccines widely despite infrastructural and socio-cultural challenges – can be vital today in our bid to deliver COVID-19 vaccines to millions of people across the globe.

The Way Forward

Of course, there is still a long way to go in the development of safe and efficacious vaccines against this coronavirus. We still do not know the exact nature of the immune responses required to afford protection. And although the timelines for vaccine development have been greatly accelerated due to the pandemic, extensive trials in populations with ongoing outbreaks in multiple locations are still necessary. This will require a coordinated effort and international collaboration.

If there is one thing that this pandemic has taught us, it is that no one is protected unless everyone is protected. Whenever and wherever COVID-19 vaccines are developed, India will have a key role in distributing these products. As one of the world's largest vaccine manufacturers, India is uniquely positioned to be able to supply vaccines to low- and middle-income countries.

Now, more than ever before, it is critical that countries like India actively participate and play a lead role in international collaborations, working together with governments and multilateral agencies to amplify and support efforts to ensure that vaccines reach those who need them the most. The future of our world depends on the success of these collaborations.

Global COVID Lab Meeting

(Photo: University Health Network)

Join us for the next Global COVID Lab Meeting with Dr. Eleanor Fish, Professor at the Department of Immunology at the University of Toronto. She will discuss the role interferons play in viral infections and how their broad-spectrum antiviral properties can be applied to COVID-19.

Type I interferons can directly inhibit viral replication and also support immune responses that help clear a viral infection. Interferon (IFN)-α was used previously in the treatment of SARS. In May, Dr. Fish and colleagues published results in *Frontiers in Immunology* from a study involving 77 hospitalized COVID-19 patients from Wuhan, China. The study showed that treatment with IFN-α2b reduced levels of detectable SARS-CoV-2 in the upper respiratory tract and also appeared to reduce some of the inflammatory responses that are a hallmark of the more serious cases of COVID-19 disease. The authors conclude that IFN-α2b should be further evaluated in a randomized, controlled clinical trial as a therapeutic option.

“Rather than developing a virus-specific antiviral for each new virus outbreak, I would argue that we should consider interferons as the 'first responders' in terms of treatment. Interferons have been approved for clinical use for many years, so the strategy would be to 'repurpose' them for severe acute virus infections,” says Dr. Fish.

Find out more about her research on June 18th at 10 AM EDT. Register here for the next Global COVID Lab Meeting.

Must Read

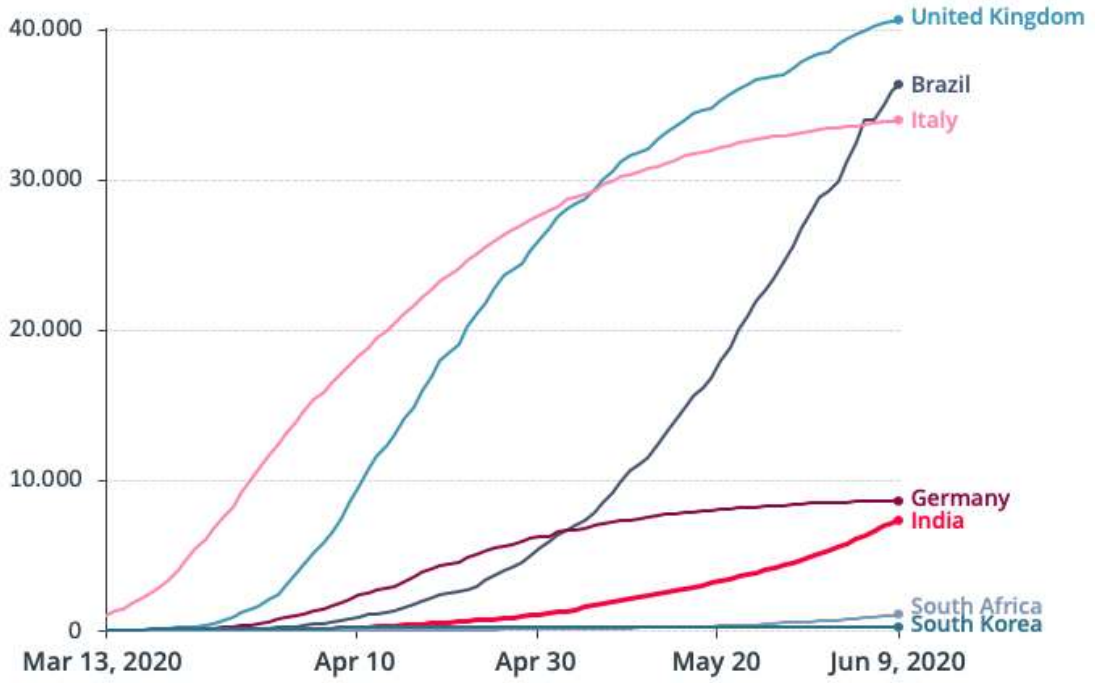
A series of preprint publications related to how antibody responses are measured and to what degree they appear following SARS-CoV-2 infection were recently made available, including a review article that summarizes the immunogenicity of the leading COVID-19 vaccine candidates in development. Other articles focus on the importance of testing and deploying vaccines in diverse settings and populations.

- In this preprint publication, scientists from Rockefeller University report on the neutralizing antibody titers detected in convalescent sera.
- Researchers report on antibody responses to SARS-CoV-2 in patients from the New York City metropolitan area in this preprint publication.

- In a related preprint publication, authors suggest that sera from mild COVID-19 cases should be included in validation panels to obtain more accurate seroprevalence rates of antibodies to SARS-CoV-2 at the population level.
- This preprint review article analyzes the immunogenicity data of leading vaccine candidates and the antibody responses triggered by SARS-CoV-2 infection.
- In this paper from *The Journal of Infectious Diseases*, Singh et al. present the case for testing COVID-19 vaccine candidates in Africa.
- Authors Kao, Orenstein, and Anderson discuss the importance of testing SARS-CoV-2 vaccine candidates in children in this article, published in *Clinical Infectious Diseases*.

COVID-19 in Numbers

Total Confirmed COVID-19 Deaths



Source: European Centre for Disease Prevention and Control - Situation Update Worldwide – Last updated 9th June, 10:45 (London time)

WORLD ENVIRONMENT DAY

AKS UNIVERSITY
THE UNIVERSITY WITH DIFFERENCE

Department of Environmental Science

INTERNATIONAL WEBINAR

on

COVID-19: Challenges & Opportunities for Sustainable Environmental Development

05 to 7 June 2020, Timing: 11.00 am to 02.00 pm, Platform Google meet

In Association with

The National Academy of Sciences, India Bhopal Chapter

The Society of Life Sciences, Satna, India

The Academy of Environmental Biology, India

INVITED SPEAKERS

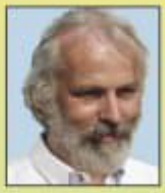
Frof. S.D. Tripathi
Hon. General Secretary NASI

Dr. Krishna Gopal
Director IIDM Lucknow
General Secreutry AEB

Dr. Prashanth N Suravajhala
Senior scientist, System Biology
BBSR Jaipur

Dr. Sajjan Khan
Senior Environment Expert
Qatar Petrolium Doha

Dr. Rakesh Dubey
Former Director
DMI Govt of MP



Dr. Ulrich Berk
President German Association of Homa Therapy

Prof. Rana Pratap Singh
Dean Academics, BBAU Lucknow

Prof. Akhilesh K Pandey
Ex. Chairman, MPPURC Bhopal

Prof. Amar P. Garg
Vice-Chancellor
Shobhit University, Meerut UP

Dr. M. Suresh Kumar
Chief Scientist and Head
EISD, NEERI Nagpur

Dr. Niraj Kumar
Executive Secretary
National Academy Of Science

Prof. Rashmi singh
Chairman, BOS & HOD(Botany)
Govt. PG College Satna

Dr. Anil K Gupta
Prof. & Head NIDM
MHA Govt. of India

Dr. Rajesh Saxena
Principal Scientist
MPCST, Bhopal

Dr. Devendra. K. Soni
Addl. Director Regional
PCB Lucknow

Dr. A.K. Shrivastava
Regional Officer
MP Pollution Control Board Satna

Prof. I.K. Pai
Former Head Zoology
Goa University Goa

Mr. Rajeev Mishra, IFS
Conservator of Forest I/C DFO

Er. Anant Soni
Pro Chancellor and Chairman
AKS University, Satna

Prof. Harshwardhan
Pro. Vice Chancellor
AKS University Satna

Dr. G.P. Richhariya
Dean, FLST
AKS University Satna

Dr. M. Kali Durai, IFS
Director Horticulture and Food Processing Govt. of M.P.

ORGANIZER

Dr. Shivesh Pratap Singh
Prof & Head Zoology
Secretary, Bhopal Chapter, NASI

Dr. Mahendra K Tiwari
Organizing Secretary
Prof. & Head Environment Sci.

- Kindly Register on the given link :- https://forms.gle/XfkqfNrWf491WS9H9
- The E-Certificate of the participation will be given to all the registered & attendee Participants

For live Streaming kindly Follo link :- https://www.aksuniversity.ac.in/live & meet.google.com/hvg-huvw-thw

National Webinar on

"Issues and Solutions for Saving Biodiversity in Himalayan Region"

On the Occasion of World Environment Day-2020 (Theme: Biodiversity)

5 June, 2020 (Friday), Duration: 12.00 noon to 2.00 pm

Organized Jointly by: Uttarakhand State Council for Science and Technology (UCOST), Dehradun and The National Academy of Sciences, India (NASI), Allahabad

Programme Schedule

S.N.	Programme	Detail/ Title	Duration
1.	Inaugural Session	Lightening of Lamp and Presentation of Bouquet to Guests	12.00 noon to 12.05 pm
2.	Welcome Address by Co-Organizing Secretary	Dr. Ashutosh Mishra, Senior Scientific Officer UCOST, Dehradun	12.05 pm to 12.10 pm
3.	About the Webinar by Organizing Secretary	Dr. D.P. Uniyal, Joint Director, UCOST, Dehradun	12.10 pm to 12.15 pm
4.	Address by Webinar Chairman; Topic: *Issues and Solutions for Saving Biodiversity in Himalayan Region*	Dr. Rajendra Dobhal, Director General UCOST, Dehradun	12.15 pm to 12.35 pm
5.	Key Note Address; Topic: *Evolution and Biodiversity*	Dr. A. N. Purohit, Former Vice-Chancellor, HNB Garhwal (Central) University, Srinagar	12.35 pm to 12.55 pm
6.	Invited Lecture; Topic: *Biodiversity and Human Health*	Dr. Mahesh Bhatt (Surgeon, Author, Public Health Consultant) MD, MMBSHS Trust; President, VIBHA Uttarakhand	12.55 pm to 1.10 pm
7.	Invited Lecture; *Topic: Reconciling Mountain Biodiversity and Post Covid - 19 Pandemics: Challenges and Way Forward*	Dr. G.S. Rawat, Former Dean, Wild Life Institute of India (WII), Dehradun	1.10 pm to 1.25 pm
8.	Vote of Thanks by Webinar Coordinator	Dr. Prashant Singh, Coordinator UCOST, DAV (PG) College, Dehradun	1.25 pm to 1.30 pm

World Environment Day was celebrated on 5th June 2020 at Patna Chapter of NASI

Photo: Plantation by Dr. Pradeep Das, Chairman of Patna Chapter NASI, and also by the senior scientists in Rajendra vatika campus of ICMR-Rajendra Memorial Research Institute of Medical Sciences, Patna. Scientists, Members of NASI, Research Scholars, Technical officers and staff were present on this occasion with the massage to save the environment.

Century of Quantum Mechanics and Still Going Strong

A Webinar Jointly organized by

Applied Physics Department, Faculty of Technology and Engineering
The Maharaja Sayajirao University of Baroda,
Indian Physics Association (Baroda Chapter),
Indian Association of Physics Teachers (IAPT-RC7)
and Gujarat Science Academy
in Association with National Academy of Sciences (Delhi Chapter)

Programme

Day 1	15th June 2020	Monday

10:45 am	Setting up of link in Zoom Cloud Webinar
11:00 am	Start of the Session
11:00 am – 11:10 am	Overview of Webinar by Prof. Arun Pratap, Dean, FTE, MSUB.
11:10 am – 11:15 am	Introduction of Hon'ble Vice Chancellor, MSUB
11:15 am – 11:25 am	Address by Prof. Parimal H Vyas, Hon'ble Vice Chancellor, MSUB
11:25 am – 11:30 am	Presidential Address by Prof. Pankaj Joshi, Provost, CHARUSAT.
11:30 am – 11:35 am	Comments by Prof. A K Singhvi, Former President, GSA.
11:35 am – 11:40 am	Introduction of Prof. S Lokanathan by Dr. Madhusudan
11:40 am – 12:10 pm	**Birth of Modern Physics** Prof. S. Lokanathan, Former Professor, Rajasthan University.
12:10 pm – 12:25 pm	Moderated Q & A Session
12:25 pm – 12:30 pm	Introduction of Prof. Ajoy Ghatak by Prof. K N Joshipura
12:30 pm – 01:15 pm	**Evolution of Quantum Theory** Prof. Ajoy Ghatak, Former Professor, IIT Delhi.
01:15 pm – 01:30 pm	Moderated Q & A Session
01:30 pm – 01:35 pm	Vote of Thanks and Closing of Day 1 (Dr J N Pandya, MSUB)

Day 2	16th June 2020	Tuesday
10:45 am	Setting up of link in Zoom Cloud Webinar	
11:00 am	Start of the Session	
11:00 am – 11:05 am	Introduction of Prof. Pankaj Joshi by Dr. T C Pandya	
11:05 am – 11:50 am	**Quantum Cosmology** Prof. Pankaj Joshi, Provost, CHARUSAT.	
11:50 am – 12:05 pm	Moderated Q & A Session	
12:05 pm – 12:10 pm	Introduction of Prof. V N Potbhare by Dr. B S Chakrabarty	
12:30 pm – 01:15 pm	**Uncertainty Principle** Prof. V N Potbhare, Former Professor, FTE, MSUB.	
01:15 pm – 1:30 pm	Moderated Q & A Session	
1:30 pm – 1:35 pm	Vote of Thanks and Closing of Day 2 (Prof. N K Jain, GSA)	

Day 3	17th June 2020	Wednesday
10:45 am	Setting up of link in Zoom Cloud Webinar	
11:00 am	Start of the Session	
11:00 am – 11:15 am	Blessings from Padma Shri Prof. H C Verma and Introduction of Dr. Anand Kumar Jha by Prof. H C Verma	
11:15 am – 12:00 noon	**Quantum Entanglement** Dr. Anand Kumar Jha, Associate Professor, IIT Kanpur.	
12:00 noon – 12:15 pm	Moderated Q & A Session	
12:15 pm – 12:20 pm	Introduction of Prof. Pafulla K Jha by Prof. Arun Pratap	
12:30 pm – 01:15 pm	**Application of Quantum Mechanics and Machine Learning Methods to Material Design** Prof. Pafulla K Jha, Professor, FoS, MSUB.	
01:15 pm – 1:30 pm	Moderated Q & A Session	
1:30 pm – 1:35 pm	Vote of Thanks and Concluding Remarks of the Webinar by Prof. Arun Pratap, Dean, FTE, MSUB.	

Webinar Series by Distinguished Experts, jointly organized by **The National Academy of Sciences India (NASI)-Delhi Chapter** and MHRD-Institution Innovation Council Deen Dayal Upadhyaya College (University of Delhi)
(Started from June 24 2020)

- registration on
  : https://docs.google.com/forms/d/e/1FAIpQLSeMMpMi2ABiHoZ6p1Suntqn Lkl-9k2E-dmmOXb5quFUkpDWuw/viewform

Schedule of the Webinar Series

June 24, 2020
Professor Katepalli Sreenivasan
Dean Emeritus of NYU Tandon School of Engineering;
The Eugene Kleiner Professor for Innovation in Mechanical Engineering;
Professor of Physics (Faculty of Arts and Science);
Mathematics (Courant Institute of Mathematical Sciences)

June 25, 2020 @ 10 am Indian Standard Time
Printed and Flexible Electronics and Devices
Dr. Jin-Woo Han
(Recipient of Presidential Early Career Award for Scientists and Engineers (PECASE)
(Recipient of Early Career Awards from the IEEE Electron Devices Society and the IEEE Nanotechnology Council)
Research Scientist, Center for Nanotechnology, NASA Ames Research Center, Moffett Field, California, USA

July 03, 2020
New chemistry and physics in magnetic oxides
Prof. J. Paul Attfield, FRS FRSE FRSC
Professor of Materials Science at Extreme Conditions
School of Chemistry, Centre for Science at Extreme Conditions,
The University of Edinburgh, Edinburgh

July 10, 2020 @ 01:30 India Standard Time
Can Future Energy Needs be met Sustainably?
Professor Sir Chris Llewellyn Smith, FRS, FAPS (USA), Honorary Fellow, IOP (UK), Foreign Fellow INSA(India)
Rudolf Peierls Centre for Theoretical Physics
Parks Road, Oxford OX1 3PU

July 11, 2020

Professor Philip K. Maini, FRS, FIMA, FRSB, FMedSci, Foreign Fellow INSA (India)
Wolfson Centre for Mathematical Biology
Mathematical Institute, Andrew Wiles Building, Radcliffe Observatory Quarter
Woodstock Road, Oxford

July 14, 2020 @ 04:30 pm Indian Standard Time

The influence of infection on Society before Covid19

Professor Sir Peter Julius Lachmann, FRS, FRCP, FRCPath, FMedSci, Foreign Fellow INSA(India), Fellow,
Emeritus Sheila Joan Smith Professor of Immunology
Christ College, University of Cambridge

Last Week of July/August 2020 - Date and Title TBD

Professor Eli Yablonovitch, FRS
The James & Katherine Lau Engineering Chair Professor, Electrical Engineering & Computer Sciences Dept., Director of the NSF Center for Energy Efficient Electronics Science, E3S Member, Kavli Energy Nano-Sciences Institute at Berkeley Senior Faculty Scientist, Lawrence Berkeley National Laboratory
University of California, Berkeley

Professor Graham R Fleming, FRS
Professor of Chemistry, University of California Berkeley
Senior Faculty Scientist, Lawrence Berkeley National Laboratory

Engineering the future of Medicine

Professor Molly S. Shoichet, FRS
Tier 1 Canada Research Chair in Tissue Engineering
Professor of Chemical Engineering & Applied Chemistry and Biomaterials & Biomedical Engineering
Donnelly Centre for Cellular & Biomolecular Research
University of Toronto,Toronto, Ontario M5S 3E1 Canada

Coordinator:
Dr. Manoj Saxena, *MNASc* (Member-NASI)
Associate Professor
Department of Electronics
Deen Dayal Upadhyaya College
University of Delhi
Dwarka Sector-3, New Delhi-110078
India