International Journal for Multidisciplinary Research (IJFMR)
E-ISSN: 2582-2160 ● Website: www.ijfmr.com ● Email: editor@ijfmr.com

Genetic Mutations and Strain Dynamics of SARS-CoV-19: A Comprehensive Study

Dr Don J Scott Berin G[1], Nisha. R. I[2]

[1]Assistant Professor, Department Of Community Medicine, White Memorial Homoeopathic Medical College, Veeyanoor, Attoor, Kk Dist, Tamilnadu, India.
[2]Msc Biotech, (Phd), Research Faculty, Department Of Research, White Memorial Homoeopathic Medical College, Veeyanoor, Attoor, Kk Dist, Tamilnadu, India.

Abstract
Coronaviruses infect humans with varying degrees of severity and lethality. Four of these viruses(NL63, 229E, OC43, and HKU1) cause mild respiratory problems in humans and three others (MERS CoV, SARS CoV including the newly emerged SARS CoV 2) can cause severe respiratory syndromes. SARS CoV 2 infection was first reported from Wuhan, China, on 24th December 2019 and in less than three months, on 11th March 2020, WHO declared COVID 19 as a pandemic. By 3rd May, 2020, 3.3 million people worldwide (213 countries) were reportedly infected and 2,38,628 individuals died of COVID 19. We note that while the infectivity of SARSCoV 2 is much higher than SARS CoV or MERS CoV , its case fatality rate (0.9 3.3%) is substantially lower than that of SARS CoV (11%) and MERS CoV (34%). For SARS CoV 2, there are notable differences in case fatality rate and disease severity among geographical regions and among age groups of infected persons, with lower severity in infants and children than in adults (*1*:*9*). The case fatality rates increased in East Asia(China and neighbourhood) and Middle East (Iran and neighbourhood) have been substantially higher than in Europe (Italy, France and Spain). By March 11, 2020 the corona virus has been changing mutations like A, A2, A2a, B, B1, B1a, B4, B2, A1a, A3, A6, A7, A2a1, A2a1a, A2a2 or A2a2a it was assigned to, based on non-missing sites. The World Health Organization (WHO) made the assessment that COVID 19 should be characterized as a pandemic. At great economic cost, many countries have adopted unprecedented measures to curb the spread of the virus such as large scale use of isolation and quarantine, closing their borders, limits of public gathering and nation wide lockdowns.

KEYWORDS: Corona virus, strains, mutation, haplotypes.

INTRODUCTION:

In a paper published early in the pandemic,3 viral sequences collected from the earliest patients were assessed and compared to known viral sequences. Sequence analysis of 11 samples found that SARS-CoV-2 is in the same species as SARS-CoV; the 2 viruses are 94.6% similar in amino acid sequence (80% nucleotide sequence similarity) across the genome.3 However, other studies from early in the outbreak do not consider the viruses to be the same species, as they differ by more than 10% in the replicase genes.4 In February, the Coronavirus Study Group (CSG) of the International Committee on Taxonomy of Viruses officially named the novel coronavirus SARS-CoV-2. The CSG analyzed viral genomes from several patients and assessed phylogenetic (evolutionary) relationships between the new virus and known

coronaviruses. The committee found that the genome of viruses isolated from patients was similar enough to SARS genomes to be considered a variant of SARS, not an entirely novel virus. While the clinical presentation, epidemiologic patterns, and host range of SARS-CoV-2 may differ from the original SARS-CoV, it is the genetic similarity between the 2 viruses that is used to conclude they are the same species. For this reason, the CSG has named SARS-CoV and SARS-CoV-2 as variants of the species known as Severe acute respiratory syndrome– related coronaviruses. The name SARS-CoV-2 is distinct from the name of the disease, which the WHO has officially designated COVID-19.

SARS-CoV-2 Evolution

Phylogenetic analysis of 30 publicly available SARS-CoV-2 samples concluded that emergence of SARS-CoV-2 in the human population likely occurred in mid-November 2019. The sequences have limited variability in consensus sequences, suggesting the outbreak was initiated from either a single introduction into humans or from a very few animal-to-human transmission events.The mutation rate has been estimated in various groups, ranging from about 1.05x10–3 to1.26x10–3 substitutions per site per year, which is similar to some estimates of MERS-CoV mutation rates5,7-9. As more viral genomes are made publicly available, scientists will better be able to track viral evolution and mutation rates, so the exact estimates will vary.

Selection analysis of the genome suggests that 2 genes in the SARS-CoV-2, the S and N genes, are under episodic selection as the virus is transmitted between humans.This is normal for emerging viruses and means that parts of the genome are undergoing positive selection. Mutations and adaptation in the S and N genes could affect virus stability and pathogenicity.9 As more genomes are made publicly available, analysis of the genome sequence diversity across samples has revealed the highest diversity occurring in the structural genes, especially the S protein, ORF3a, and ORF8.

SARS-CoV-2 is evolving over the course of the pandemic. However, this evolution is not occurring faster than expected compared to other viruses during an outbreak. There are different clades of SARS-CoV-2 developing as COVID-19 spread across the globe.13 Different clades emerge as viruses evolve. This is entirely normal and does not mean there are new strains of SARS-CoV-2 that are more pathogenic than others circulating right now.

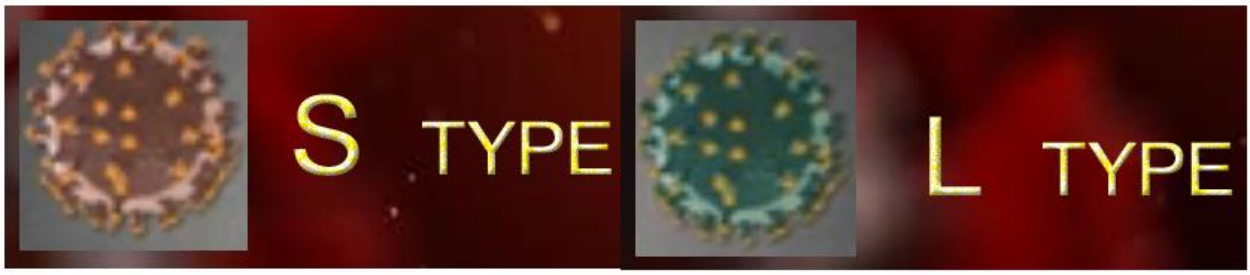
Scientists have done an incredible job sequencing samples of SARS-CoV-2 and sharing results during this pandemic. These sequences are allowing public health officials to estimate several important parameters of the epidemiology of COVID-19, such as the reproductive number and introduction of the virus into new regions.

Structure of SARS- CoV-2

Coronaviruses, including the pneumonia-causing novel coronavirus currently known as SARS-CoV-2, are enveloped, nonsegmented, positive-sense RNA viruses. Coronavirus genomes have some of the largest genomes among RNA viruses, with approximately 25-32 kilobases.1 The typical CoV genome includes a 5’-cap, 5’ untranslated region (UTR), open reading frames, a 3’-UTR, and 3’-poly(A) tail. The first two thirds of the genome typically codes for nonstructural proteins from 2 open reading frames that form the replicase complex. The last third of the genome encodes primarily structural proteins.2 There are 4 conserved structural proteins across CoVs: the spike (S) protein, membrane (M) protein, envelope (E) protein, and nucleocapsid (N) protein.1 The S protein is responsible for binding to host cell receptors and viral entry to host cells. The M, E, and N proteins are part of the nucleocapsid of viral particles.

Two Strains of COVID-2019 But One is Actually Deadlier Than the Other:

According to the study, which you can read here, the so-called S-Strain is the original, "ancestral" strain of the disease mutated from a coronavirus transmissible between bats. The second, L-Strain, by comparison, evolved from the S-Strain but appears to spread more quickly because it predominated in early cases of the disease identified in Wuhan. But in a twist, when he analyzed a later, wider sample set he found that the older S-Type appeared to be spreading more frequently than in the initial outbreak. In this article we'll delve into why that may be happening and also why experts warn it's unclear that the differences between the two strains are significant.



Mutating Viruses:

COVID-2019 is an RNA virus, which means it uses ribonucleic acid to encode its genetic material instead of DNA. RNA viruses, which include viruses like Ebola, influenza, rabies, and even the common cold, tend to mutate extremely frequently because their polymerase enzymes used for reproduction aren't as good at "proofreading" for errors when transcribing genetic code. The constant mutations make RNA viruses more of a moving target which annually require new vaccines to cure.

Still, that means it's not intrinsically remarkable that there are already multiple strains of COVID-2019 circulating according to scientists. Furthermore, mutations cut both ways— they can make viruses less deadly as well as more deadly, or may even have no significant effect. And a "successful" mutation—one that improves the virus's odds of propagation— isn't necessarily one that makes it more deadly.

The difference between the L-Strain and S-Strain amounts to two specific amino acids—sort of like a change of two lines of code in a computer program. However, the report's authors concede they're not sure what impact the changes have ("the concerned amino acid…plays a yet undefined role in the viral life cycle.")Scientists claim that for now the difference is small enough that a treatment against one strain will likely remain effective against the other.

Aggressive L-Strain and Ancestral S-Strain

Nonetheless, Xiaolu's study suggests there may be different traits in the two strains. The initial sample consisted of strains isolated Wuhan—the city where the COVID-2019 outbreak began—prior to January 7, 2020. He found 26 L-Type strains and just one S-Type. The paper characterizes the L-Strain as being "more aggressive and spread more quickly." In fact, it states the S-Type is the original, "ancestral" form of COVID-2019, while the LType is a mutation. Reportedly, the L-Strain genome diverges only 4 percent from the bat coronavirus designated RaTG-13.

By contrast, "our mutational load analysis indicated that the L-type had accumulated a significantly higher number of derived mutations than S-type."Here's where the narratives become more complicated. In a subsequent sample of cases identified after January 7 elsewhere in China (save for one) and in foreign

countries, he counted 72 L-Types and 29 S-Type viruses. That means the older S-Strain actually appeared to become more common, not less, than the supposedly more aggressive L-Strain. In his words: "…the S type, which is evolutionarily older and less aggressive, might have increased in relative frequency due to relatively weaker selective pressure."

What explains that paradox?

Xiaolu's hypothesis is that it all comes down to human intervention—namely that measures taken to contain COVID-2019 by humans may have impacted L-Strain infections more than S-Strain, selecting in favor of the latter. Xiaolu doesn't elaborate much on how intervention might have affected the two strains' relative success in propagating. Perhaps the L-Strain's symptoms manifested faster or more dramatically, allowing persons with the L-Strain to be identified and isolated faster while the less mutated S-Strain cases remained under the radar for longer.

How fast can the coronavirus mutate?

The new coronavirus, like all other viruses, mutates, or undergoes small changes in its genome. A recently published study suggested that the new coronavirus, SARS-CoV-2, had already mutated into one more and one less aggressive strain. But experts aren't convinced. In the study, a group of researchers in China analyzed the genomes of coronaviruses taken from 103 patients with COVID-19, the disease caused by SARS-CoV-2, in Wuhan, China, the epicenter of the outbreak. The team found differences in the genomes, which they said could be categorized into two "strains" of the coronavirus: the "L" type and the "S" type. The researchers found the "L" type, which they deemed the more aggressive type, in 70% of the virus samples. The more commonly found type today is the older, "S" type, because "human intervention" such as quarantines may have reduced the ability of the "L" type to spread.

SARS-COV-2 TESTING

SARS-CoV-2 infected individuals were identified in Iceland through both targeted testing of high-risk individuals and population screening. The targeted testing started on January 31 2020, and has focused on individuals at high risk of infection because they were symptomatic (cough, fever, aches, and shortness of breath) or coming from ski areas in Austria and northern Italy, their contacts and those who contacted the healthcare system. As of March 19, the high-risk areas designation was extended to the whole world outside Iceland. As of March 22, 4,551 individuals had been tested in this targeted screening. The population screening for SARS-CoV-2 was initiated by deCODE on March 14 when 323 individuals had tested positive for SARS-CoV-2 by the targeted screening. The screening was open to everyone symptom free or with mild symptoms of common cold that is highly prevalent in Iceland at this time of the year. The registration for the test was online and during sample collection recent travels, contacts with positive individuals, and symptoms compatible with COVID-19 were registered.

Over a six-day period 5,502 Icelanders were tested for SARS-CoV-2. All newly positives were placed in isolation and those who had been in contact with the positives, in quarantine for 2 weeks. All symptomatic individuals in quarantine were also tested. In addition to isolating positives and quarantining those at high risk of infection, on March 16 the Icelandic authorities initiated a ban on mass gathering above 100 people and stated that a social distancing of at least 2 meters should be maintained. On March 24, the ban on mass gathering was extended to 20 people. SARS-CoV-2 testing, it was recommended to take both nasopharyngeal and oropharyngeal samples. RNA from all samples was isolated within 24 hours.

TRACKING OF SARS-COV-2 INFECTIONS

All individuals who tested positive for SARS-CoV-2 were contacted by phone by a team designated by the authorities to track their infection. They were asked about their symptoms and when they started, recent travels and previous contacts with infected individuals. They were also asked to identify everyone whom they had been in contact with 24 hours before noticing their first symptom and for how long they interacted with each individual and how intimate the interaction was. All registered contacts were contacted by phone, requested to go into 2 weeks quarantine and asked about symptoms. Those with symptoms and those who developed them in quarantine were tested for SARS-CoV-2.

RNA EXTRACTION

Viral RNA samples were extracted either at the Department of Clinical Microbiology laboratory at Landspitali, the National University Hospital of Iceland (LUH) or at deCODE. Both extraction methods are based on an automated magnetic bead-purification procedure, which includes cell lysis and Proteinase K treatment. RNA from samples at LUH were extracted (32 samples per 60 min run) using the MagNA Pure LC 2.0 instrument from Roche LifeScience, with 200/100 µL input/output volume(s), respectively. Samples at deCODE were extracted from swabs (96 samples per 70 min run) using the Chemagic Viral RNA kit on the Chemagic360 instrument from Perkin Elmer, with 300/100 µL input/output volume(s), respectively. Each step in the workflow was monitored using an in-house LIMS (VirLab) with 2D barcoding (Greiner, 300 µL tubes) of all extracted samples.

TESTING OF SAMPLES FOR SARS-COV-2

Testing for SARS-CoV-2 was performed either at LUH or deCODE using similar quantitative real-time PCR (qRT-PCR) methods. TaqMan™ Fast Virus 1-step Master Mix, 2019-nCoV Assay kits v1 and 2019-nCov control kits were obtained from Thermo Fisher. Assay mix A, B and C were prepared containing FAM™ dye labelled probes for the SARS-CoV-2 specific genes ORF1ab, S-protein and N-protein, respectively. In addition, each assay mix contained VIC™ dye labelled probes for human RNase P as internal control. Samples from 96-well RNA sample plate(s) were dispensed into three wells each in a 384 plate layout, in addition to three negative (no template) and three positive controls.

Assay mix was added in a total reaction volume of 12.5 µL per sample. All sample aliquoting and mixing at deCODE was performed with an automated Hamilton STARlet 8-channel liquid handler and the assay plates were scanned in an ABI 7900 HT RT-PCR system following manufacturer´s instructions with a total of 40 cycles of amplification. Samples with FAM™ dye Ct values <37 in at least two of three assays were classified as positive. Samples with FAM™ dye Ct values between 37 and 40 were classified as inconclusive and their testing repeated. If repeated testing gave the same result the sample was classified as positive. Samples with undetected FAM™ dye Ct values or values equal to 40 in all three assays were classified as negative if the human RNaseP assay was positive (VIC™ dye Ct <40).

Validation of the RNA extraction and the qRT-PCR method(s) at deCODE was performed using 124 samples that had previously tested positive (n=104) or negative (n=20) with the qRT-PCR assay at LUH. All of the negative samples tested negative at deCODE and 102 of the 104 positive tested at LUH were also positive at deCODE. Two samples that tested positive at LUH were negative at deCODE. Upon subsequent sequencing viral genome could not be detected in these samples, probably because very few viral particles were present.

SAMPLE PREPARATION FOR SEQUENCING

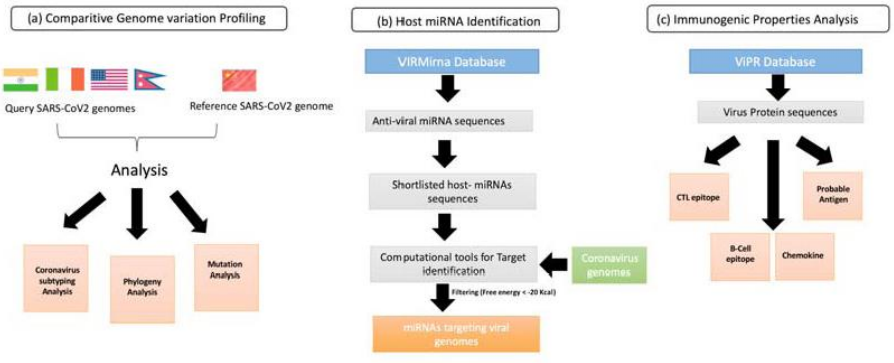
Reverse transcription (RT) and multiplex PCR was performed based on information provided by the Artic Network initiative (https://artic.network/) to generate cDNA. In short, extracted viral RNA was pre-incubated at 65 °C for 5 min in the presence of random hexamers (2.5 µM) and dNTP´s (500 µM). Sample cooling on ice was then followed by RT using SuperScript IV (ThermoFisher) in the presence of DTT (5 mM) and RNaseOUT inhibitor (Thermo Fisher) for 10 min at 42°C, followed by 10 min at 70 °C. Multiplex PCR of the resulting SARS-CoV-2 cDNA was performed using a tiling scheme of primers, designed to generate overlapping amplicons of approximately 800 bp (Table S1). The primers were generously provided by Dr. David Stoddard at Oxford Nanopore Technologies.

Two PCR reactions were done for each sample using primer pools A and B, respectively . PCR amplification was done using the Q5® Hot Start High-Fidelity polymerase (New England Biolabs) with primers at 1 µM concentration. The reactions were performed in an MJR thermal cycler with a heated lid at 105 °C, using 35 cycles of denaturation (15 sec at 98 °C) and annealing/extension (5 min at 65 °C). The resulting PCR amplicons were purified using Ampure XP magnetic beads (Beckman Coulter) and quantified using the Quant-iT™ PicoGreen dsDNA assay kit (Thermo Fisher). Amplified samples (20-500 ng) were randomly sheared using focused acoustics in 96-well AFA-TUBE-TPX plates (Covaris Inc.) on the Covaris LE220plus instrument with the following settings: Sample volume, 50 µL; temperature, 10 °C; peak incident power, 200W; duty factor, 25%; cycles per burst, 50; time, 350 sec. Sequencing libraries were prepared in the 96 well Covaris plates, using the NEBNext® Ultra II kit (New England Biolabs) following the manufacturer´s instructions. In short, end repair and A-tailing was performed in a combined reaction per sample (plate) for 30 min at 20 °C, followed by thermal enzyme inactivation at 65 °C for 30 min. Adaptor ligation was done using the NEBNext® ligation master mix plus enhancer and the TruSeq unique dual indexed IDT adaptors. Ligation reactions were incubated for 15 min at 20 °C. Ligated sequencing libraries were purified on a Hamilton STAR NGS liquid handler, using two rounds of magnetic SPRI bead purification (0.7X volume).

Viral Haplotypes

We sequenced SARS-CoV-2 RNA extracted from 643 samples; of these samples, we obtained coverage of more than 90% of the SARS-CoV-2 genome from 581 samples and more than 67% from 605 samples. We called 409 sequence variants, 291 of which were not found in the GISAID database. We usedclade-informative mutations to assign haplotypes to persons: 518 from the targeted testing group and 59 from the populationscreening group.

Geographic Viral Origin



To shed further light on the geographic origin of the SARS-CoV-2 infections in residents of Iceland, we generated a median-joining network of 1547 complete viral sequences (513 from complete viral genomes from Icelanders and 1034 from other populations around the world). Several viral lineages have emerged during the 3 to 4 months since the original outbreak in China, with an average of five mutations separating the lineages from the founding haplotype from Wuhan (the central haplotype of clade A). Although the sequencing efforts vary considerably among populations, it is clear that the geographical distribution of clades is highly structured. Thus, A and B haplotypes are common in East Asia, whereas the B1a haplotype appears to be at the center of the outbreak on the West Coast of the United States, and A2a and its descendants are almost exclusively found in European populations.

Composition of Haplotypes

The haplotypes of SARS-CoV-2 infections observed in Iceland cluster into several diverse clades. To estimate the number of introductions of the SARS-CoV-2 virus to Iceland, we searched for infected persons who had traveled
internationally or had an unknown source of infection. This led us to 363 persons for whom viral genomes had been sequenced. These genomes clustered into 42 distinct clades, which provided a lower boundary on the number of
individual introductions. Of the 157 sequenced virions obtained during the early targeted testing, 143 were in the A2 clade. By the time we initiated the population screening, all travelers who had returned from ski resorts in the Alps had been requested to self-quarantine and were not eligible for participation, which resulted in a substantially different composition of haplotypes.

For example, the A2a2 haplotype, which was most commonly seen in travelers coming from Austria in the early phase of targeted testing, was much less frequent in travelers in the population screening. The A1a haplotype was more common in the general population than in those who received targeted testing, with a total of 23 of 59 haplotypes among participants in the population-screening group, as compared with only 8 of 157 haplotypes in the early-targeted testing. The composition of haplotypes changed substantially from early targeted testing to later
targeted testing. The A2a1 and A2a2 haplotypes, which had collectively made up 103 of 157 haplotypes (65.6%) in the early-targeted testing, were reduced to 115 of 361 haplotypes (31.9%) in the later-targeted testing, mostly because of the increased frequency of the A1a and other A2a-derived haplotypes. This change probably
meant that population screening identified clusters of infected persons who seeded infection from areas that had not been designated as high risk, such as the United Kingdom.

The relatively high prevalence of A1a and A2a clades in the later-targeted testing group was unsurprising: the targeted testing had been extended to include those who had traveled to additional highrisk areas, and population screening had identified cases that could be used to inform tracking efforts. The A2a3a and A2a2a haplotypes were the two most common haplotypes in Iceland; of the 577 persons who provided samples that were sequenced, the A2a3a haplotype was found in 78 (13.5%) and the A2a2a haplotype was found in 45 (7.8%).

Table 2. Distribution of SARS-CoV-2 Haplotypes, According to Timing of Diagnosis and Internationally Imported or Local Transmission.*

Haplotype	Early Targeted Testing January 31–March 15			Population Screening March 13–April 1			Later Targeted Testing March 16–31		
	Imported	Country of Origin	Local	Imported	Country of Origin	Local	Imported	Country of Origin	Local
	no. (%)		no. (%)	no. (%)		no. (%)	no. (%)		no. (%)
All haplotypes	101 (100)	IT:42, AT:38	56 (100)	16 (100)	UK:9, US:3	43 (100)	83 (100)	AT:28, UK:23	278 (100)
A2a1	36 (35.6)	IT:29, AT:3	16 (28.6)	5 (31.2)	UK:4, DE:1	8 (18.6)	17 (20.5)	AT:6, UK:4	51 (18.3)
A2a2	33 (32.7)	AT:28, DK:2	18 (32.1)	2 (12.5)	US:2	1 (2.3)	19 (22.9)	AT:9, UK:5	28 (10.1)
A2a3	5 (5.0)	CH:5	11 (19.6)	0		11 (25.6)	5 (6.0)	ES:2	82 (29.5)
A2a	19 (18.8)	IT:11, AT:6	5 (8.9)	2 (12.5)	NL:1, DK:1	1 (2.3)	30 (36.1)	AT:13, UK:10	33 (11.9)
A1a	2 (2.0)	CH:1, IT:1	6 (10.7)	4 (25.0)	UK: 4	19 (44.2)	4 (4.8)	UK:2, ES:2	60 (21.6)
Other clade A	2 (2.0)	UK:1, AT:1	0	2 (12.5)	UK:2	3 (7.0)	2 (2.4)	US:1, UK:1	19 (6.8)
B1a1	4 (4.0)	US:4	0	1 (6.2)	US:1	0	6 (7.2)	US:6	5 (1.8)

* Among the imported haplotypes, each country is represented by a two-letter country code as follows: AT denotes Austria, CH Switzerland, DE Germany, DK Denmark, ES Spain, IT Italy, NL Netherlands, UK United Kingdom, and US United States. For each haplotype, the two most common countries of origin are indicated, followed by the number of participants with that haplotype.

Temporal and geographical spread of subtypes of SARS-CoV-2

The first set of RNA sequences collected from 17 infected individuals from Wuhan, China, had high sequence
identity, and was named the O subtype. The O subtype spread to other provinces (e.g., Guangdong, Jiangxi) of
China and also to nearby countries, e.g., Thailand (with the first submission to GISAID - Nonthaburi/61/2020
- on 8th January 2020). within two weeks. Our phylodynamic analysis showed that the virus evolved into B
and B2 in the first two weeks of January and later to B1, B4, A2a and A3. These subtypes rapidly spread worldwide to multiple East Asian countries (8 countries), Europe (5) and North America (2Analysis of data on 6181 sequences generated during the past three months revealed that 3789 mutation events distributed over 3772 nucleotide sites. Among these sites, variants at only 11 sites, of which 8 were coding (ORF8 -L84S, ORF1a - V378I, ORF1a - L3606F, ORF1a - A3220V, ORF3a - G251V, ORF1a - L3606F, S - D614G, ORF1b - P314L), were present in relatively high frequencies in multiple populations.

These 11 sites were the most useful in defining the phylogenetic clade structure of the viral sequences. The radial phylogenetic timetree is depicted in Supplementary Figure**,** with concentric circles showing the dates of sequence deposition/collection; earlier dates of deposition/collection are closer to the centre**.** We found that 5629 (91%) of 6181 of SARS-CoV-2 RNA sequences were submitted from East Asian, European and North American regions. The remaining 9% of the sequences were from Oceania, South America, Africa and South-West- Central Asian countries. During 15th January to 31st March 2020 (10 weeks), SARS-CoV-2 had evolved into 11 clades. Four of these 11 clades have attained frequencies higher than 5%; A2a=60.95%, O=13.3%, B1=9% and A1a=7.8%.

The A2a clade with the highest frequency is defined by nucleotide changes at two “highly informative” sites that are in complete non-random association (linkage disequilibrium): a non-synonymous D614G (Aspartate (D) -> Glycine (G)) mutation in the Spike glycoprotein and a P314L mutation in Orf1b protein of SARS-CoV-2. Of the 11 clades, only 2 clades (a minor A2 subtype and the major A2a subtype) have Glycine at the 614th amino acid position; the remaining 9 clades (ancestral O and evolved clades B, B1, B2, B4, A3, A6, A7 and A1a) have Aspartate Supplementary Figure. Analysis has shown that the most frequent A2a subtype arose in China in mid-January 2020 (Inferred date: January 15th, CI: 4th-20th January) and spread to multiple locations within 15 days. The initial A2a sequences were deposited from

China (Zhejiang/HZ103/2020 on 24th January), and then from another province Shanghai 300Kms away (Shanghai/SH0014/2020 on 28th January and Shanghai/SH0086/2020 on 31st January).
The earliest evidence of A2a in Europe was from Germany (Germany/BavPat1/2020 sequence deposited on 28th
January). By the end of January, 98.3% of all submitted sequences from East Asia, 90% of all in Europe and
100% of all in North America were of non-A2a subtypes with 614D. The present at a very low frequency in East Asia (3 sequences), Europe (1 sequence) and absent in North America. Dramatic changes in the viral landscape took place very rapidly. By the end of February, the frequency of A2a rose from 9% to 56.25% in Europe and from 0% to 9.7% in North America.
Contrastingly, the A2a landscape remained low in China (<1% in February) and other East Asian countries (only 1 A2a sequence out of 206 sequences deposited by the end of February); The A2a - 614G subtype continued to rise in frequency replacing all previously frequent non-A2a-614D subtype in Europe (69% in March comprising 2373 sequences) and in North-America (61% in March comprising 933 sequences) to become the most dominant subtype. Sequence submission from East Asia dropped drastically in March possibly indicating control of the COVID-19 disease; The RNA viruses mutate fast and accumulates changes during transmission process from one infected individual to another. Interestingly, during January 15th and March 31st, accumulation of new isolate-specific mutations, were substantially fewer for the A2a clade than for the other clades.
These facts are consistent with our negative estimates of Tajima's D that indicate rapid population expansion coupled with selective sweep of non-A2a-614D clades in East Asia and of A2a-614G clade in Europe and North America. Further, the impact of positive selection on A2a may have been higher in Europe than in East Asia, as evidenced by the A2a isolates in Europe and North America having acquired a fewer number of new mutations during this time period than those in East Asia.

The clade specific D614G mutation provides advantage to A2a subtype for host cell entry

The non-synonymous D614G mutation, which defines the A2a clade of SARS-CoV-2, is located between S1-RBD and S2 junctions of the SARS-CoV-2 spike (S) protein. The amino acid position 614 is monomorphic for D (Aspartate) residue in SARS-CoVs obtained from bat, civet, pangolin and human. It is intriguing that a mutated allele (614G, present in the A2a subtype) at this conserved amino acid site should rapidly rise to a high frequency in some, but not all, regions of the world. *In vitro* introduction of new proteolytic sites at and around the S1-S2 junction is known to substantially increase SARS-CoV fusion with cell membrane. We have predicted, by proteolytic cleavage prediction analysis, that a novel serine protease (elastase) cleavage site has been introduced in SARS-CoV-2 at 615-616 residues on S protein .
The position 614 on the S protein is the nearest substrate site for serine protease to cleave at 615-616. The S glycoprotein must be cleaved by host proteases to enable fusion of the viral envelope with the host cell membrane. This fusion is essential for viral entry. The introduction of the new cleavage site for the elastase protease in A2a possibly provides this subtype an advantage over the other subtypes for entry into the host cell. Elastase is mainly expressed by host neutrophils that is elevated in COVID-19 . By gaining an additional elastase cut site, the A2a - 614G subtype is likely to obtain a substantial advantage to enter host cells more efficiently due to simultaneous processing of exogenous elastase proteases and membrane

bound TMPRSS2. Of relevance is the fact that SARS-CoV infection was shown to be enhanced by exogenous proteases, including elastase.

SEQUENCING DATA ANALYSIS

Amplicon sequences were aligned to the reference genome of the SARS-CoV-2 (NC_045512.2)2 using bwa mem14, possible PCR duplicates were marked with markduplicates from Picard tools15 and reads with less than 50 bases aligned were omitted from the alignment. The resulting aligned filtered reads were used for variant calling with bcftools14. For consensus sequence generation only variants reported as homozygous were used. In regions targeted with primers we allowed variants to have allele frequency below one in individual. The consensus sequence was masked with ambiguous nucleotides (N) at positions if the depth of coverage was strictly less than 5 reads after restricting to bases of quality 20 or higher. Consensus sequences with more than 10,000 ambiguous nucleotides were discarded from analysis. The mutations in Table S3 were used to define haplogroups/clades. For network analysis of haplogroups a median-joining network8 of SARS-CoV-2 sequences was generated using data from our sequencing effort in Iceland and from GISAID available on March 22. Only sequences with start positions <=200 and stop positions >=29750 were included in the analysis.

For the GISAID sequences, only those with <=1% missing nucleotides were used, whereas for the Icelandic sequences a more permissive threshold of <=5% was imposed. To reduce noise in the network, an imputation step was implemented for sequences with missing nucleotides at sites where other sequences varied, whereby the missing nucelotide was imputed to the consensus variant for the clade (A, A2, A2a, B, B1, B1a, B4, B2, A1a, A3, A6, A7, A2a1, A2a1a, A2a2 or A2a2a) it was assigned to, based on non-missing sites.They have also identified a non-silent damaging variant in a highly conserved SRCR domain (rs12329760; V160M) of *TMPRSS2* that is present in East Asians at double the frequency (MAF=0.4) than in Europeans (MAF=0.2). The SRCR domain interacts with external pathogens and helps in positioning of the protease head towards the substrate (*64*).

The SARS-CoV-2 A2a subtype specific D614G mutation is in a region of the S protein that is highly conserved across the coronavirus family. We predict that the V160M mutation in the host and D614G mutation in the A2a subtype will result in reduced interaction of the SRCR domain of TMPRSS2 and the S protein of the coronavirus leading to a reduction in viral entry in the host cell. High prevalence of the V160M mutation in populations of East Asia is a likely reason for the reduced spread of A2a subtype in East Asia. In this study, we have provided strong evidence that host genetic variation regulating the expression of the *TMPRSS2* gene has shaped the global spread of the A2a subtype of SARS-CoV-2.

Dominant SARSCoV2 strain A2a binds more easily to ACE2 receptors

The COVID19 pandemic has affected almost all countries across the world, and India is no exception. Researchers have found that severe acute respiratory syndrome coronavirus 2 (SARSCoV2) has undergone several mutations with time. The virus was first detected in Wuhan, Hubei province of China, in December 2019. Drs. Nidhan Biswas and Partha Majumder from the National Institute of Biomedical Genomics in Kalyani, West Bengal, India, are ready with their findings on the mutated version of the virus.

What was this new study about?

Since its first detection in December 2019, the novel coronavirus or the SARS CoV2 virus has been found to have undergone mutations and has at present 10 different strains. One of these strains is A2a. This strain of the virus has infected the most number of people around the world. This dominant strain, say researchers is extremely virulent and can infect human lung cells in huge numbers and quickly overwhelm the patient, especially those with underlying comorbidities or other illnesses such as high blood pressure, diabetes, heart disease, respiratory disease, and renal disease.

What did the researchers do?

The researchers compared this new virus with its predecessor, the SARS virus, that led to the SARS outbreak in 2003. This outbreak had killed over 800 and infected 8,000 individuals and was also quite capable of infecting the
lungs. The A2a strain of SARS CoV2, however, is more virulent the researchers said. They explained that the transmission rate of this new virus is an indication of how fast this new strain can move from one individual to another. For this study, the team looked at the shared data from RNA sequencing from different nations. This data is available at the public database, GISAID freely shared by researchers across the world for the benefit of other researchers. They collected 3636 RNA sequences of novel coronaviruses from 55 countries between December 2019 and April 6, 2020.

What did they find?

The researchers believe that their work will help the development of an efficient vaccine against this strain of the virus. Wuhan strain of the virus, they wrote was the "O" type of the virus. This new A2a strain is one of the ten
mutations that originated from this original strain. From the samples of the infection, the team looked at strains from 55 countries and India. From 55 nations they looked at 3,636 coronavirus RNA sequences. From these, they noted there were 50.8 percent or 1,848 samples containing the A2a strain. 580 sequences were showing the O strain and 505 sequences showing the B1 strain. From the 35 Indian RNA sequences checked, 47.5 percent or 16 were A2a strain and 13, 5, and 1 were A3, O, and B, respectively.
It (A2a) has become the dominant type of SARSCoV2." Majumder, professor and founding director of NIBMG, said, "The coronavirus can be classified into many types O, A2, A2a, A3, B, B1, and so on. Currently, there are 11 types, including type O, which is the ancestral type' that originated in Wuhan. In Italy, 80 percent of the samples contain A2a strains, say reports. There is a predominance of this train in other nations such as the United Kingdom, the United States, Spain, Iceland, Congo, and Brazil. Since China has not submitted sequencing reporting after February 2020, the strains from China at present are not known.To live, a virus must propagate by infecting other animals. A mutation usually disables the virus from transmitting itself." However, not always does a mutation render a virus incapable of transmission, as has been seen here. The A2a strain originating from the O strain is more efficient in infecting humans. Such mutant viruses increase the frequency (of transmission) and sometimes completely replace the original type of the virus.

CONCLUSION:

The ongoing pandemic of the coronavirus disease 2019 (COVID-19) is an infectious disease caused by severe acute respiratory syndrome coronavirus 2 (SARS-CoV2). We have performed an integrated

sequence-based analysis of SARS-CoV2 genomes from different geographical locations in order to identify its unique features absent in SARS-CoV and other related coronavirus family genomes, conferring unique infection, facilitation of transmission, virulence and immunogenic features to the virus. The phylogeny of the genomes yields some interesting results. Systematic gene level mutational analysis of the genomes has enabled us to identify several unique features of the SARS-CoV2 genome, which includes a unique mutation in the spike surface glycoprotein (A930V (24351C>T)) in the Indian SARS-CoV2, absent in other strains studied here. I have also predicted the impact of the mutations in the spike glycoprotein function and stability, using computational approach. To gain further insights into host responses to viral infection, we predict that antiviral host-miRNAs may be controlling the viral pathogenesis. My analysis reveals nine host miRNAs which can potentially target SARS-CoV2 genes. Interestingly, the nine miRNAs do not have targets in SARS and MERS genomes. Also, hsa-miR-27b is the only unique miRNA which has a target gene in the Indian SARS-CoV2 genome. I also studied the immune epitopes in the genomes.

Bibliography

1. Huang C, Wang Y, Li X, et al. Clinical features of patients infected with 2019 novel coronavirus in Wuhan, China. Lancet 2020;395(10223):497–506.
2. Zhou P, Yang X-L, Wang X-G, et al. A pneumonia outbreak associated with a new coronavirus of probable bat origin. Nature 2020;579(7798):270–3.
3. Zhu N, Zhang D, Wang W, et al. A novel coronavirus from patients with pneumonia in China, 2019. N Engl J Med 2020;382(8):727–33.
4. WHO [Internet]. Available from: https://www.who.int/emergencies/diseases/novel-coronavirus-2019
5. WHO [Internet]. Available from: https://www.who.int/docs/default-source/coronaviruse/situation-reports/20200323-sitrep-63-covid-19.pdf
6. WHO [Internet]. Available from: https://www.who.int/docs/default-source/coronaviruse/situation-reports/20200311-sitrep-51-covid-19.pdf
7. Guan W-J, Ni Z-Y, Hu Y, et al. Clinical Characteristics of Coronavirus Disease 2019 in China. N Engl J Med [Internet] 2020 [cited 2020 Mar 25];Available from: http://www.ncbi.nlm.nih.gov/pubmed/32109013
8. Onder G, Rezza G, Brusaferro S. Case-Fatality Rate and Characteristics of Patients Dying in Relation to COVID-19 in Italy. JAMA [Internet] 2020 [cited 2020 Mar 25];Available from: http://www.ncbi.nlm.nih.gov/pubmed/32203977
9. No Title [Internet]. [cited 2020 Feb 28];Available from: https://www.landlaeknir.is/um-embaettid/frettir/frett/item39279/Frettatilkynning-vegna-koronaveirunnar-COVID-19-28-02-2020
10. Drosten C, Günther S, Preiser W, et al. Identification of a Novel Coronavirus in Patients with Severe Acute Respiratory Syndrome. N Engl J Med [Internet] 2003 [cited 2020 Mar 25];348(20):1967–76. Available from: http://www.nejm.org/doi/abs/10.1056/NEJMoa030747
11. Ksiazek TG, Erdman D, Goldsmith CS, et al. A Novel Coronavirus Associated with Severe Acute Respiratory Syndrome. N Engl J Med [Internet] 2003 [cited 2020 Mar 25];348(20):1953–66. Available from: http://www.nejm.org/doi/abs/10.1056/NEJMoa030781
12. Zaki AM, van Boheemen S, Bestebroer TM, Osterhaus ADME, Fouchier RAM. Isolation of a Novel Coronavirus from a Man with Pneumonia in Saudi Arabia. N Engl J Med [Internet] 2012 [cited 2020 Mar 25];367(19):1814–20. Available from: http://www.nejm.org/doi/abs/10.1056/NEJMoa1211721

13. Shu Y, McCauley J. GISAID: Global initiative on sharing all influenza data – from vision to reality. Eurosurveillance [Internet] 2017 [cited 2020 Mar 25];22(13):30494. Available from: http://www.eurosurveillance.org/ViewArticle.aspx?ArticleId=22750
14. Li H, Durbin R. Fast and accurate short read alignment with Burrows-Wheeler transform. Bioinformatics 2009;25(14):1754–60.
15. Broad Institute [Internet]. Available from: http://broadinstitute.github.io/picard/;Lo C-Y, Tsai T-L, Lin C-N, Lin C-H, Wu H-Y. Interaction of coronavirus nucleocapsid protein with the 5'- and 3'-ends of the coronavirus genome is involved in genome circularization and negative-strand RNA synthesis. *FEBS J* 2019;286(16):3222-3239. doi:10.1111/febs.14863
16. Zhou P, Yang X-L, Wang X-G, et al. Discovery of a novel coronavirus associated with the recent pneumonia outbreak in humans and its potential bat origin. *Microbiology* 2020. doi:10.1101/2020.01.22.914952
17. Zhu N, Zhang D, Wang W, et al. A novel coronavirus from patients with pneumonia in China, 2019. *N Engl J Med* 2020. doi:10.1056/ NEJMoa2001017
18. Rambaut A. Phylodynamic analysis of SARS-CoV-2 genomes- 27- Jan-2020. *Virological* January 27, 2020. http://virological.org/c/ novel-2019-coronavirus/33/l/latest. Accessed January 28, 2020.
19. Bedford T, Richard N, Hadfield J, Hodcroft E, Muller N, Llcisin M. Narrative: Genomic analysis of nCoV spread. Situation report 2020-01-23. nextstrain. https://nextstrain.org/narratives/ncov/ sit-rep/2020-01-23?n=1. Published January 23, 2020. Accessed January 24, 2020
20. Cotten M, Watson SJ, Zumla AI, et al. Spread, circulation, and evolution of the Middle East respiratory syndrome coronavirus. *mBio* 2014;5(1). doi:10.1128/mBio.01062-13
21. Dudas G, Carvalho LM, Rambaut A, Bedford T. MERS-CoV spillover at the camel-human interface. *eLife* 2018;7:e31257. doi:10.7554/eLife.31257
22. Baric RS, Yount B, Hensley L, Peel SA, Chen W. Episodic evolution mediates interspecies transfer of a murine coronavirus. *J Virol* 1997;71(3):1946-1955. doi:10.1128/ JVI.71.3.1946-1955.1997
23. Benvenuto D, Giovanetti M, Ciccozzi A, Spoto S, Angeletti S, Ciccozzi M. The 2019-new Coronavirus epidemic: evidence for virus evolution. *bioRxiv* January 2020. doi:https://doi.org/10.1101/2020.01.24.915157
24. Sironi M, Cagliani R, Forni D, Clerici M. Evolutionary insights into host–pathogen interactions from mammalian sequence data. *Nat Rev Genet* 2015;16(4):224-236. doi:10.1038/nrg3905
25. Nextstrain / ncov (2/3/2020). nextstrain. https://nextstrain.org/ ncov?m=num_date. Accessed February 3, 2020. February 3, 2020.
26. Bell SM, Müller N, Wagner C, et al. Narrative: genomic analysis of COVID-19 spread. Situation report 2020-04-10. nextstrain April 10, 2020. https://nextstrain.org/narratives/ncov/sit-rep/2020-04-10. Accessed April 11, 2020.
27. Masters PS. Coronavirus genomic RNA packaging. *Virology* 2019;537:198-207. doi:10.1016/j.virol.2019.08.031.
28. 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 Chandrika Bhattacharyya1#, Chitrarpita Das1#, Arnab Ghosh1#, Animesh K. Singh1, Souvik Mukherjee1, Partha P. Majumder1,2, Analabha Basu1, and Nidhan K. Biswas1* These authors contributed equally.National Institute of Biomedical Genomics, Kalyani, India – 741251, 2 Indian Statistical Institute, Kolkata, India – 700108.

29. www.thehindu.com Date : 2020-04-30 ONE MUTATED CORONAVIRUS TYPE DOMINATES GLOBALLY Relevant for: Developmental Issues | Topic: Health & Sanitation and related issues.
30. The New England Journal of medicine Spread of SARS-CoV-2 in the Icelandic Population.
31. Virus has mutated into 10 types, one now dominant across regions: Study, https://timesofindia.indiatimes.com/city/mumbai/virushasmutatedinto10typesonenowdominantacrossregionsstudy/ Articleshow/75417399.cms.
32. THE NATIONAL INTEREST There Are Two Strains of COVID-2019—But Is One Actually Deadlier Than the Other? March 21, 2020, by Sebastien Roblin.