ANNQUEST JULY 2020
ISSN:2321-3043

www.stannsannquest.com

STRUCTURAL DESIGN OF CORONA VIRUS AND ITS DIVERSITY

Ajjan Vaishnavi

St. Ann's college for women

Hyderabad

Abstract

Coronaviruses (CoVs) have caused outbreaks of deadly pneumonia in humans since the beginning of the 21st century. The severe acute respiratory syndrome coronavirus (SARS-CoV) emerged in 2002 and was responsible for an epidemic that spread to five continents with a fatality rate of 10% before being contained in 2003 (with additional cases reported in 2004). The Middle-East respiratory syndrome coronavirus (MERS-CoV) emerged in the Arabian Peninsula in 2012 and has caused recurrent outbreaks in humans with a fatality rate of 35%. SARS-CoV and MERS-CoV are zoonotic viruses that crossed the species barrier using bats/palm civets and dromedary camels, respectively. No specific treatments or vaccines have been approved against any of the six human coronaviruses, highlighting the need to investigate the principles governing viral entry and cross-species transmission as well as to prepare for zoonotic outbreaks which are likely to occur due to the large reservoir of CoVs found in mammals and birds. Here, I review the understanding of the infection mechanism used by coronaviruses derived from recent structural and biochemical studies.

Key words:SARS, CoV-2, COVID-19, intrahost variant, microbiota, transmission, bats.

Introduction

Coronaviruses (CoVs) are enveloped positive-sense RNA viruses. The club-like spikes projecting out from their surface gave them the name. Coronaviruses possess an unusual large RNA genome as well as a unique replication strategy. Coronaviruses cause a variety of diseases in animals ranging from cows, pigs to chicken, and other birds. In humans, coronaviruses can cause potentially lethal respiratory infections.

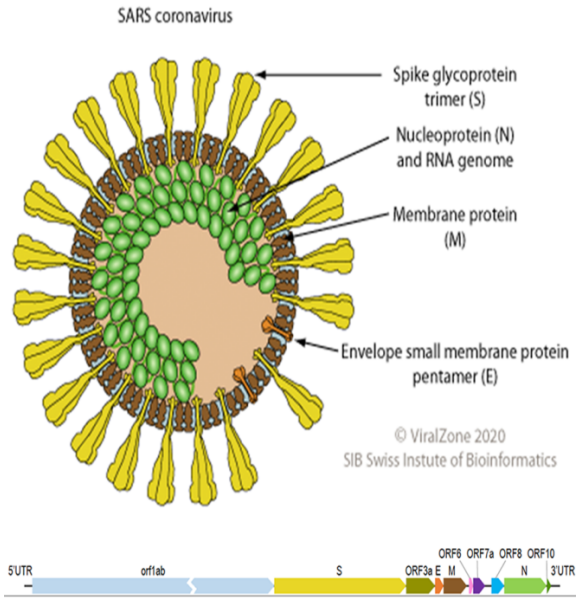
In an extraordinary effort, the full-length genome sequence of the SARS-coronavirus (SARS-CoV) was elucidated within weeks after the identification of this novel pathogen and published by the Michael Smith Genome Sciences Center (Vancouver, Canada, Entrez Genomes accession number NC_004718 (AY274119)), the Centers for Disease Control and Prevention (Atlanta, USA,5 GenBank accession number AY278741), and others. The SARS-CoV genome is ~29.7 kb long and contains 14 open reading frames (ORFs) flanked by 5' and 3'-untranslated regions of 265 and 342 nucleotides, respectively. Homologs of proteins conserved in all coronaviruses are encoded by the overlapping ORFs 1a and 1b, and by ORFs 2, 4, 5, 6 and 9a.

The mechanisms of some coronaviruses entry into cells have already been studied, such as severe acute respiratory syndrome coronavirus (SARS-CoV), murine hepatitis virus (MHV), and human coronavirus (HCoVs). Entry of SARS-CoV into HepG2 and COS7 cells is

Edit with WPS Office

clathrin-dependent while entry into Vero E6 cells is clathrin- and caveolae-independent, but the lipid raft plays an important role in the process. MHV entry into cells needs clathrin, the same as HCoV-NL63. For HCoV-229E, caveolae-mediated endocytosis is utilized to enter human fibroblast cells. SARS-CoV and MHV-CoV can induce continuous micropinocytosis, but this occurs in the later phase during infection and is not associated with virus entry. Coronaviruses enter host cells via various endocytic pathways after viral spike glycoprotein (S) interacts with receptors and then initiates the endocytic process. Internalized viruses are trafficked like cargoes to membrane fusion sites through specific transport routes. Different CoVs have varying fusion sites. The fusion site of the middle east respiratory syndrome coronavirus (MERS-CoV) takes place in the early endosome, while MHV and the feline infectious peritonitis virus (FIPV) are transported to the lysosome to fuse.

Structure of corona virus



Common features of coronaviruses include

(i) a highly **conserved genomic organization** with a large replicase gene preceding structural and accessory genes,

(ii) expression of many **non-structural genes** by ribosomal frameshifting,

(iii) several **unique of unusual enzymatic activities** encoded within the large replicase-transcriptase polyprotein.

(iv) expression of downstream genes by synthesis of 3'-**nested sub-genomic mRNAs**.

The typical organization of the genome is5'-leader-UTR-replicase-**S**(Spike)-**E**(Envelope)-**M**(Membrane)-**N**(Nucleocapsid)-3'-UTR-poly(A) tail. M Accessory genes are interspersed within the structural genes at the 3'-end of the genome.Accessory proteins are not needed for replication in tissue culture but appear to be important in viral pathogenesis. The synthesis of polypeptide ab (pp1ab) involves programmed ribosomal frame shifting during translation of open-reading frame 1a (orf1a). Frame shifting results in a new reading frame that produces a trans-frame protein product. In coronaviruses, a fixed portion of the ribosomes translating orf1a change reading frame at a specific location now decoding information contained in orf1b.

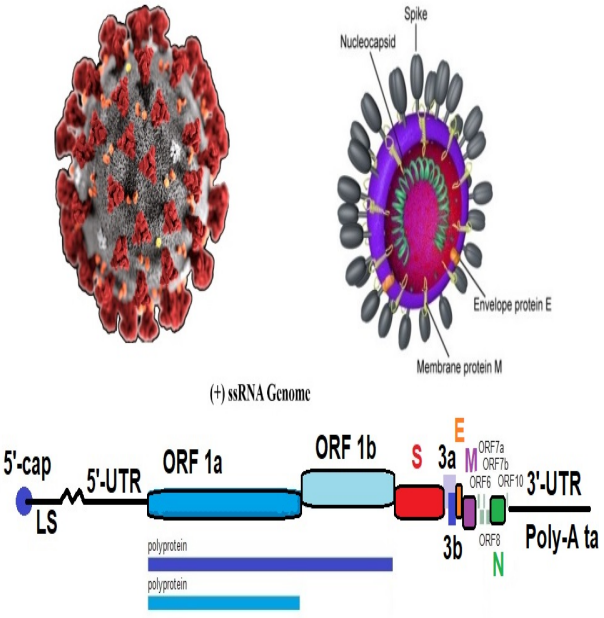
U_UUA_AAC is a universal frame-shifting site

Coronaviruses contain a frameshifting stimulation element as a conserved RNA sequence forming a stem-loop that promotes ribosomal frameshifting. Ribosomal frameshifting is a mechanism in which open-reading frame 1b (orf1b) is expressed. Replicase-transcriptase proteins are

Edit with WPS Office

encoded in open-reading frame 1a and 1b (orf1a and orf1b) and are synthesized initially as two large polyproteins termed pp1a and pp1b. A comparative analysis performed by Baranov et al. in 2004 revealed the sequence U_UUA_AAC as a universal shift site. Frameshifting was characterized in SARS-CoV cultured in mammalian cells using a dual luciferse reporter system and mass spectrometry. Tandem tRNA slippage on the sequence U_UUA_AAC was confirmed by mutagenic analysis of the shift site. Mass spectrometry was used for the analysis of affinity tagged frameshift products. Further analysis of the frameshifting site showed that a proposed RNA secondary structure in loop II and two unpaired nucleotides at the stem I-stem II junction in SARS-CoV are important for frameshift stimulation.

SARS-CoV-2 (COVID-19)



Taxonomy: Group IV ((+)ssRNA); Coronaviridae; Coronavirinae; Betacoronavirus; Sarbecovirus; 2019-nCoV; Severe acute respiratory syndrome coronavirus 2

Virion: enveloped, spherical, 60-140 nm in diameter with spikes of about 9-12 nm

Genome: ~30 kb positive-sense, single -stranded RNA

RNA Transcript: 5' cap, 3' poly-A tail

Proteome: 10 proteins

Transmission: Some early cases had some link to a seafood and animal market, suggesting animal-to-human transmission. Later cases indicate sustained human-to-human transmission.

Phylogeny: More closely related to bat-SL-CoVZC45 and bat-SL-CoVZXC21 than SARS

Epidemiology: First reported in Wuhan, China, in December 2019.

Clinical: Reported 2019-nCoV cases have ranged from no symptoms to severe pneumonia and death. Symptoms can include fever, cough, shortness of breath.

Internal features of the virus

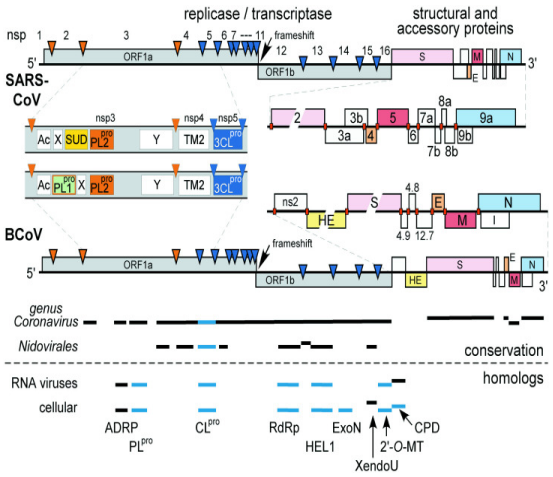
Coronaviruses are enveloped, positive-stranded RNA (+RNA) viruses, with a single-stranded genome of between 27 kb and 31.5 kb, the largest among known RNA viruses. The genomes of coronaviruses and related viruses in the order Nidovirales are polycistronic and are expressed through a sophisticated combination of poorly understood regulatory mechanisms. Coronavirus genome expression starts with the translation of two large replicase ORFs, whose coding capacity is about twice that of the average complete +RNA virus genome. *Via* a −1 ribosomal frameshift, the ORF1a polyprotein (pp1a; >4000 amino acid residues) can be extended with ORF1b-encoded sequences to yield a >7000 amino acid residue pp1ab polyprotein. Replicase

Edit with WPS Office

polyprotein processing is carried out by two or three ORF1a-encoded viral proteinases. The processing products are a group of largely uncharacterized (putative) replicative enzymes, including an RNA-dependent RNA polymerase, an RNA helicase that is fused to a complex N-terminal Zn-finger, and a Zn-ribbon-containing papain-like proteinase. The replicase subunits are thought to assemble into a viral replication complex that is targeted to cytoplasmic membranes by various membrane-associated viral proteins. In addition to genome replication, the coronavirus replicase complex mediates the synthesis of an extensive nested set of subgenomic (sg) mRNAs (transcription) to express all ORFs downstream of ORF1b, which encode a variety of structural and accessory proteins. The number and composition of these 3'-proximal ORFs vary greatly among coronaviruses, but they always include genes for the structural proteins S, M, E and N, which drive cytoplasmic virus assembly. The mechanisms underlying the synthesis of genomic and subgenomic RNAs are poorly understood. To explain the composite structure of the sg mRNAs, which are both 5' and 3'-coterminal with the viral genome, several models have been put forward, of which the one postulating the discontinuous synthesis of negative-stranded sg templates for sg mRNA synthesis has received wide support recently.

On the basis of antigenic cross-reactivity, coronaviruses were originally classified into three groups (termed groups 1, 2, and 3). Subsequently, the phylogeny-based clustering of coronaviruses proved at first (almost) identical with that based on antigenic cross-reactivity. The same three clusters were evident upon analysis of the replicase region which does not contribute to virion antigenicity. This indicated that different regions of the coronavirus genome have indeed co-evolved and that intergroup recombination has not played a prominent role in coronavirus evolution. However, the agreement between the two classifications is not perfect, as some coronaviruses are sufficiently different to not have antigenic cross-reactivity with the established groups, but close enough to cluster with one of them (group 1) on the basis of sequence comparisons. Consequently, these viruses were placed into (the expanded) group 1. Here, we refer to coronavirus groups as evolutionary clusters that unite viruses not necessarily having antigenic cross-reactivity.

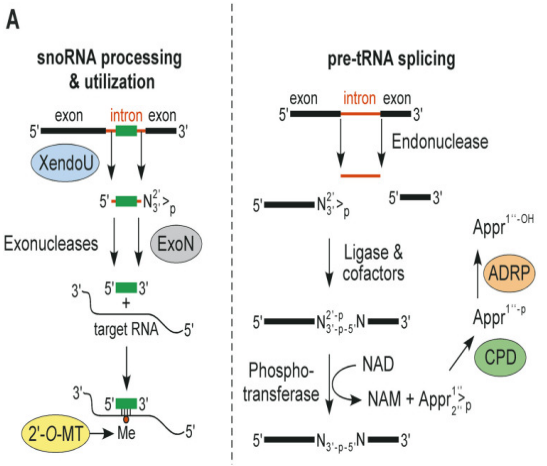
Using the recently published SARS-CoV genome sequences, the evolution, organization and expression of SARS-CoV. The SARS-CoV genome and proteome were compared with those of other coronaviruses, distantly related nidoviruses, and databases, and several of our predictions were verified experimentally.



Overview of the SARS-CoV genome organization and expression. Comparison of the genome organizations of SARS-CoV and bovine coronavirus (BCoV). The replicase genes are depicted, with ORF1a,

Edit with WPS Office

ORF1b, and ribosomal frameshift site indicated. Arrows represent sites in the corresponding replicase polyproteins that are cleaved by papain-like proteinases (orange) or the 3C-like cysteine proteinase (blue). Cleavage products are provisionally numbered nsp1–nsp16. In the 3'-terminal part of the genomes, homologous structural protein genes are indicated in matching colors. Close-ups of two regions with major differences are shown (and see the text). In the N-terminal half of replicase ORF1a, SARS -CoV lacks one of the PLpro domains (indicated in orange/green in BCoV) and contains a unique insertion (SUD). In the region with structural and accessory protein genes, the location of the body TRSs involved in subgenomic RNA synthesis are indicated with red boxes (see Figure 3 and Hofmann et al.76). The bottom part of the Figure illustrates which parts of the genome are conserved in the genus Coronavirus and in the order Nidovirales (the ORF1a sequence of toroviruses, which largely remains to be sequenced, could not be included). Furthermore, it is indicated for which domains homologs have been identified in other RNA viruses and the cellular world. Enzymes for which structural data are available are shown in blue. SUD, SARS-CoV unique domain; PLpro, papainlike cysteine proteinase; 3CLpro, 3C-like cysteine proteinase; TM, transmembrane domain; ADRP, adenosine diphosphate-ribose 1"-phosphatase; ExoN, 3'-to-5' exonuclease; CLpro, chymotrypsin-like proteinase; RdRp, RNA-dependent RNA polymerase; HEL1, superfamily 1 helicase; XendoU, (homolog of) poly(U) -specific endoribonuclease; 2'-O-MT, S-adenosylmethionine-dependent ribose 2'-O-methyltransferase; CPD, cyclic phosphodiesterase.



Homologs	XendoU	ExoN	2'-O-MT	CPD	ADRP
SARS-CoV	+	+	+	-	+
Group 2 Coronaviruses	+	+	+	+	+
genus *Coronavirus*	+	+	+	+/-	+
genus *Torovirus*	+	+	+	+	?
Roniviridae	+	+	+	-	-
Arteriviridae	+	-	-	-	-

Nidoviruses encode homologs of cellular enzymes involved in RNA processing. The cellular pathways for processing of pre-U16 snoRNA and pre -tRNA splicing are summarized, with relevant enzymatic activities indicated. For details, see the text. Homologs of the highlighted enzymes have been identified in nidoviruses. summarizing the conservation of homologs of the cellular enzymes presumably involved in RNA processing in SARS-CoV and different nidovirus groups.

In the other pathway, which involves tRNA-processing, the utilization of a 2'-phosphate group of a splicing intermediate involves the conversion of adenosine diphosphate ribose 1"-2" cyclic phosphate (Appr>p) by CPD into adenosine diphosphate ribose 1"-phosphate (Appr-1"-p), of which the phosphate group may be further processed by an ADRP. Both these activities may drive the production of mature tRNA. Although the nidovirus homologs of CPD and ADRP remain to

Edit with WPS Office

be characterized, they are not under the control of the ORF1a/ORF1b ribosomal frameshift signal and may thus, unlike the ORF1b-encoded enzymes, be produced in larger quantities.

The nidovirus homologs of the five RNA processing enzymes discussed above may interfere with these or similar cellular RNA processing pathways to reprogram the cell for the benefit of virus reproduction. It seems even more conceivable that they, alone or in concert with other enzymes like the RdRp or helicase, are involved directly in viral RNA synthesis, particularly in transcription, which, in an apparent parallel with snoRNA-driven processes, is guided by conserved oligonucleotide base-pairing interactions. The viral enzymes, like their cellular counterparts, might be part of separate pathways or, alternatively, cooperate in a single pathway in which the XendoU, ExoN and 2'-*O*-MT homologs provide RNA specificity, and the CPD and ADRP homologs modulate the pace through processing of compound(s) containing 2'-phosphate groups. In this respect, we note that both the XendoU/ExoN/2'-*O*-MT and CPD/ADRP cellular pathways start with an endoribonuclease-mediated cleavage to produce molecule(s) with 2'-3'-cyclic phosphate termini, indicating the structural basis for possible cooperation of the coronavirus homologs of these enzymes in a single pathway. The expected functional hierarchy of the five putative nidovirus enzymesis supported by their corresponding evolutionary conservation, with the XendoU homolog being absolutely conserved and the CPD homolog being least conserved among nidoviruses

Diversity of corona virus

Coronaviruses originate in animals like camels, civets and bats and are usually not transmissible to humans. But occasionally a coronavirus mutates and can pass from animals to humans and then from human to human.

Coronaviruses belong to the largest group of viruses called the Nidovirales order. Members of this order include the Coronaviridae, Arteriviridae, and Roniviridae families. The Coronvirinae are one of two subfamilies in the Coronaviridae family. Coronavirinae are further subdivided into for groups, the alpha, beta, gamma, and delta coronaviruses. Nowadays, these viruses are divided using phylogenetic clustering. These virus families have animal and human hosts. The Middle Eastern Respiratory Syndrome Coronavirus (MERS-CoV) and Severe Acute Respiratory Coronavirus (SARS-CoV) are examples.

Nidoviruses contain an infectious, linear, positive-sense RNA genome that is capped and polyadenylated. Based on their genomesize,nidoviruses are divided into two groups large and small nidoviruses.

All Nidovirales viruses are enveloped, non-segmented positive-sense RNA viruses containing very huge genomes.

The most recent common ancestor (MRCA) of all coronaviruses is estimated to have existed as recently as 8000 BCE, although some models place the common ancestor as far back as 55 million years or more, implying long term coevolution with bat and avian species. The most recent common ancestor of the alphacoronavirus line has been placed at about 2400 BCE, the betacoronavirus line at 3300 BCE, the

Edit with WPS Office

gammacoronavirus line at 2800 BCE, and the deltacoronavirus line at about 3000 BCE. Bats and birds, as warm-blooded flying vertebrates, are an ideal natural reservoir for the coronavirus gene pool (bats the reservoir for alphacoronavirus and betacoronavirus – and Avian species are the reservoir for gammacoronavirus and deltacoronavirus).

Many human coronavirus have their origin in bats. The most closely related bat coronavirus and SARS-CoV diverged in 1986. A possible path of evolution, of SARS coronavirus and keen bat coronaviruses, suggests that SARS related coronaviruses coevolved in bats for a long time. The ancestors of SARS-CoV first infected leaf-nose bats of the genus Hipposideridae; subsequently, they spread to horseshoe bats in the species Rhinolophidae, and then to civets, and finally to humans.

The human coronavirus discovered in 2003, SARS-CoV, which causes severe acute respiratory syndrome (SARS), has a unique pathogenesis because it causes both upper and lower respiratory tract infections.

In December 2019, a pneumonia outbreak was reported in Wuhan, China. On 31 December 2019, the outbreak was traced to a novel strain of coronavirus, which was given the interim name 2019-nCoV by the (WHO), later renamed as SARS-CoV-2, Severe acute respiratory syndrome (SARS).

The Wuhan strain has been identified as a new strain of Betacoronavirus from group 2B with approximately 70% genetic similarity to the SARS-CoV. The virus has a 96% similarity to a bat coronavirus, so it is widely suspected to originate from bats as well.

Most, but not all, of the first known cases in December 2019 were traced to an animal market in the Chinese city of Wuhan and are believed to have come from contact with live animals that were infected.

COVID-19 symptoms may begin like the flu but go on to develop fever, cough, and shortness of breath that is severe enough to warrant hospitalization in many patients. Unfortunately, to date, there is no antiviral drug or vaccine to treat this infection. Symptom relief and supportive care (many requiring hospital care) are the current treatment methods. However, early supportive care may reduce the disease severity.

Discussion

RNA viruses have a high mutation rate owing to the lack of proofreading activity of polymerases. Consequently, RNA viruses are prone to evolve resistance to drugs and escape from immune surveillance. The mutation rate of SARS–CoV-2 is still unclear. However, considering that the median number of pairwise sequence differences was 4 (interquartile range, 3–6) for 110 sequences collected between 24 December 2019 and 9 February 2020, the mutation rate should be at the same order of magnitude in SARS-CoV ($0.80–2.38 \times 10^{-3}$ nucleotide substitution per site per year). The high mutation rate also results in a high level of intrahost variants in RNA viruses. The median number of intrahost variant in patients with COVID-19 was 4 for variants with frequency ≥5%, and this incidence did not differ significantly from that reported in a study on Ebola (655 variants with frequency ≥5% in 134 samples; $P > .05$).

An exoribonuclease has been

Edit with WPS Office

proposed to provide proofreading activity in SARS–CoVand noted that all 3 key motifs in the gene were identical in SARS–CoV and SARS–CoV-2. In addition, neither polymorphism nor intrahost variant was detected in these motifs, suggesting that the gene is highly conserved. Although we did not find any mutation hot-spot genes in either polymorphism or intrahost variants, the observation of shared intrahost variants among different individuals implied the possibility of convergent or adaptive evolution of the virus in patients, which could potentially affect the antigenicity, virulence, and infectivity of the virus.

It is worth noting that the SARS–CoV-2 genome in patients could be highly diverse, which was also observed in other viruses. The high diversity could potentially increase the fitness of the viral population, making it hard to be eliminated. Further studies are needed to explore how this may influence the immune response toward the virus and whether there is a selection acting on different strains in the human body or during the transmission. In a single possible transmission event investigated in this study, we found no evidence for the transmission of multiple strains. However, it is unclear whether these intrahost variants occurred before the transmission or after the transmission, which would result in different conclusions. In addition, a bottleneck may be involved in the transmission, which could also decrease diversity. Nevertheless, the observation of high mutation burden in some patients emphasized the possible rapid evolution of this virus, despite that its biological significance is largely unknown.

Recent studies have shown that the microbiota in the lung contributed to the immunological homeostasis and potentially altered susceptibility to viral infection. Meanwhile, the lung microbiota could also be regulated by invading viruses. However, besides finding that microbial diversity was significantly lower in patients with pneumonia than that in healthy controls, we did not identify any specific microbiota pattern shared among patients with COVID-19, nor among patients with CAP. A possible reason for this could be the use of antibiotics in patients with pneumonia. However, this was not true for all pneumonia samples, because some samples had a substantial proportion of bacteria, including 2 samples from patients with COVID-19.

It is well known that secondary bacterial infection—a common complication of viral infection, especially for respiratory viruses—often significantly increases morbidity rates. Thus, the elevated level of bacteria in the BALF samples from some patients with COVID-19 might increase the risk of secondary infection. In the clinical data, the secondary infection rate for COVID-19 was between 1% and 10%. However, the quantitative relationship between bacterial relative abundancetiter and infection is unclear.

Overall, our study has revealed the evolution of SARS–CoV-2 in the patient, a common feature shared by most RNA viruses. How these variants influence the fitness of viruses and genetic diversity in the population awaits further investigation. Hence, there is an urgent need to accumulate more sequences to trace the evolution of the viral genome and associate the changes with clinical symptoms and outcomes.

Results

The median number of intrahost variants was 1–4 in SARS–CoV-2–infected patients, ranged from 0 to

Edit with WPS Office

51 in different samples. The distribution of variants on genes was similar to those observed in the population data. However, very few intrahost variants were observed in the population as polymorphisms, implying either a bottleneck or purifying selection involved in the transmission of the virus, or a consequence of the limited diversity represented in the current polymorphism data. Although current evidence did not support the transmission of intrahost variants in a possible person-to-person spread, the risk should not be overlooked. Microbiotas in SARS–CoV-2–infected patients were similar to those in CAP, either dominated by the pathogens or with elevated levels of oral and upper respiratory commensal bacteria.

Conclusion

SARS–CoV-2 evolves in vivo after infection, which may affect its virulence, infectivity, and transmissibility. Although how the intrahost variant spreads in the population is still elusive, it is necessary to strengthen the surveillance of the viral evolution in the population and associated clinical changes.

References

1.https://www.gisaid.org/

2.https://uk.news.yahoo.com/coronavirus-antibody-test-99-per-120300172.html?guce_referrer=aHR0cHM6Ly9yLnNlYXJjaC55YWhvby5jb20vX3lsdD1Bd3JYbkNHNmVxbGVTVjBBaGhYQUhBeC47X3lsdT1YM29ETVRCeWIybHZiWFZ1QkdOdmJHOERaM0V4QkhCdmN3TXhCSFowYVdRREJITmxZd056Y2ctLS9SVj0yL1JFPTE1ODgxOTQxMDYvUk89MTAvUlU9aHR0cHMlM2ElMmYlMmZ1ay5uZXdzLnlhaG9vLmNvbSUyZmNvcm9uYXZpcnVzLWFudGlib2R5LXRlc3QtOTktcGVyLTEyMDMwMDE3Mi5odG1sL1JLPTIvUlM9T0Rha1I3UFU3UkdWZnRQeXkwa21GU2FzRGVNLQ&guce_referrer_sig=AQAAAGRoGluq_19Yo7OUSUx3yShq43oeFGZ9TMn1wnwoY0OJaWkQaJ_QRhMrJFxYid68WFIqX70e4fo3ywJJxyXUxDqx02KqzdjcvQYpD7jTNbkVl8raJayUCymV8Zki38gW2JNIOYs-R1DL3x2yoeHEPMvtAmN1Bduto5q2QEaR9DvQ&guccounter=2

3.https://www.npr.org/sections/goatsandsoda/2020/01/24/798661901/wuhan-coronavirus-101-what-we-do-and-dont-know-about-a-newly-identified-disease

Edit with WPS Office