Ricos Biology Journal, 2023, vol. 1 (1) 12-25

Considerations about Genomic and Proteomic of (SARS-CoV)-2 and Concern Variants

Soheir M. Syame* and Ashraf S. Hakim

Department of Microbiology and Immunology, National Research Centre, 33 Bohouth St., Dokki, Cairo, Egypt

Abstract

A novel coronavirus that related to previous SARS-CoV and Middle East respiratory syndrome coronaviruses, has been emerged in the end of 2019 in China and rapidly human to human transmitted causing what is termed pandemic COVID-19 illness. The causative agent was termed severe acute respiratory syndrome coronavirus (SARS-CoV)-2, and suggested to be of bat corona virus's origin. Later on, other virus variants were emerged and become as great concern. The investigated genomic and proteomic structures of the novel virus were rather mentioned in this article.

Keywords: Omicron, SARS-CoV-2 phylogeny,SARS-CoV2 genome, SARS-CoV-2 proteins.

Corresponding author: sohir_syam@yahoo.com
Submitted: December 16, 2022
Revised: February 25, 2023
Accepted: March 1st., 2023

Introduction

A short time ago at the end of 2019, one of the Chinese business and trade cities; Wuhan experienced a rapidly spread outbreak of an unusual respiratory distress [1]. In February 2020, WHO announced the disease as a global pandemic caused by an unprecedented coronavirus [2, 3]. The novel virus was termed as Wuhan coronavirus or 2019 novel coronavirus (2019-nCov) by the Chinese researchers and defined as a member of the β group of coronavirus. Later on, the International Committee on Taxonomy of Viruses (ICTV) named the virus as SARS- CoV-2 and the disease as Coronavirus disease 2019 (COVID-19) [4, 5]. It inducedunusual disruptions; social distancing, workingstoppage, cities lockdowns and travel restrictions resulted in remarkable weakness in world and consequently individual economy [6]. In 2021 exhibited many identified SARS-CoV-2 variant Gamma, Delta and Omicron. These variants of concern "VOC" alarming high number of alterations, particularly in the viral spike protein, increased transmission potency and evading the neutralizing antibodies [7].

This current essay provides a comprehensive overview on the origin, transmission, genomic structure, different structural and functional proteins of SARS-CoV-2variants.

1. Etiology

SARS-CoV-2 which is belongs to subgenus *Sarbeco virus*, the subfamily *Orthocoronavirinae* in the family of*Coronaviridae* of the order *Nidovirales* [8].

The family *Coronaviridae* is subdivided to four subgroups; alpha (α), beta (β), gamma (γ) and delta (δ). Corona viruses developed their name from presence of crown-like spikes on the outer surface of the virus. Coronaviruses are

considered as minute sized viruses; (65–125 nm in diameter) and their genome comprise a positive single-stranded RNA as a nucleic material, size ranging from26 to 32kbs in length (Fig. 1) [9].

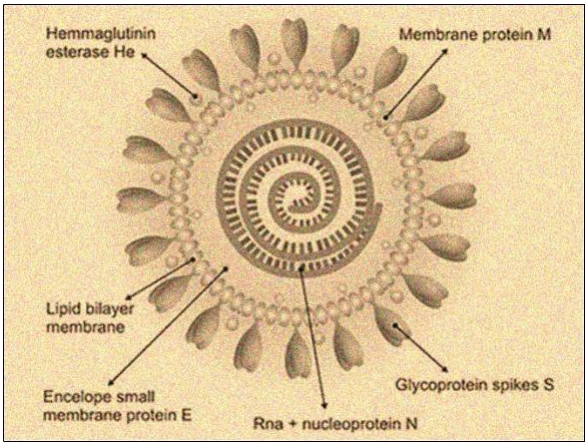

Fig. (1): diagrammatic structure of Corona viruses

2. Genome and phylogeny

Based on whole genomes phylogenetic analysis, ten Chinese and five USA current SARS-CoV-2 outbreak isolates weresequenced using the gamma distributionMEGA 7.0version. The obtained data exposed that the isolates are nearly identical S-gene based phylogeny suggesting a monophyletic clade [10]. The phylogenetic tree of family*Coronaviridae* falls into two clades. TheBeta-coronavirus genus constitute one clade, while the other clade comprises the Alpha, Gamma and Delta-coronaviruses[11]. The studies manifested that the SARS-CoV-2 is in the same Beta-coronavirus clade as SARS- CoV and MERS-CoV in the parallel manner to SARS-like bat CoVs indicating their close relation.

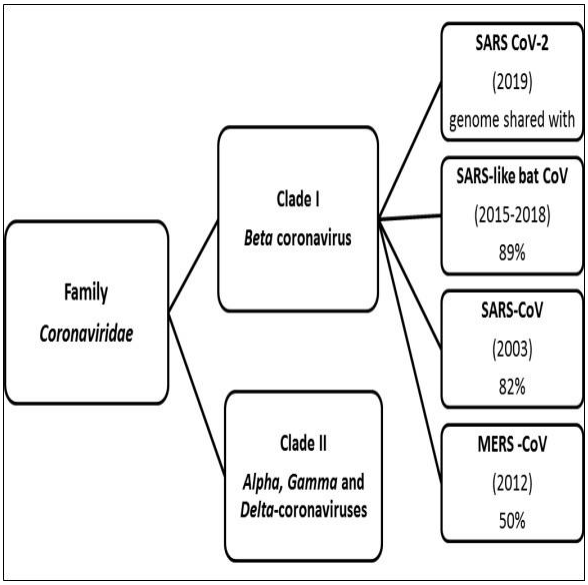

Fig. (2): Schematic diagram of phylo-genomicrelations of novel corona virus

Primer novel virus genome analysis demonstrated a close evolutionary association with the SARS like bat coronaviruses; through determining the first three genomes, namely Wuhan/IVDC-HB-01/2019 (HB01), Wuhan/IVDCHB-04/2019 (HB04), andWuhan/IVDC-HB-05/2019 (HB05), [12].

Later on, in-depth genome notation of the novel virus was carried out with a comparison to related coronaviruses; it was found that the genome of SARS-CoV- 2 is

almost identical to those three coronaviruses, with only five nucleotide differences in the genome of ~29.8 kb nucleotides. The genome of coronaviruses whose size domains approximately 26- 32 kb, is the largest among all known RNA viruses, with G + C contents varying from 32% to 43% and contains an inconstant number (at least 6) of ORFs [13].

To realize the greatest output from their courted genomes, viruses frequently avail what is called alternative open reading frames (ORF), in which translation is launched from a start codon within a presented gene and, getting out of frame, brings about a distinguished protein product [14].

The SARS-CoV-2genome was defined to possess 14 ORFs encoding 27 proteins. The first ORF "orf1ab" is the largest gene, looking over approximately 67% of the entire genome encodes pp1ab protein and other 15 non-structural proteins (NSPS), while the orf1a gene encodes for pp1a protein with 10 NSPS, both previous genes located at the 5''-terminus of the genome. However, the other ORFs resided at the 3'' -terminus of the genome encode structural as well as eight subsidiary proteins (3a, 3b, p6, 7a, 7b, 8b, 9b, and orf14), [9, 11, 15, 16].

"VOC" genomics

Available database of Omicron full-genome sequences exposed that this VOC alters by about 55 mutations (range 48 to 92; mean 54.6 ± 5.2 sequences) from the initial virus [17]. While in another dominating Delta VOC, the number is not that much higher; ranging from about 21 to 44 substitutions (34.4 ± 2.6; $n = 475$). Not surprise that the most mutations, being located in the viral spike gene that is essential for virus infectivity and considered the key goal of protective immune responses [18].

Interestingly, phylogenetic analysis revealed that the Omicron VOC obviously did not emerge from other VOCs, comprising the Delta VOC (figure 3). This unexpected observation may be explained via three different theories; the first is efficient replication in an immune-compromised individual over a long period of time [19]. The other two theories are either Omicron has been circulating in the human population for some time in parallel to COVID-19 or Omicron may have evolved in a nonhuman species from which it spilled back to humans [20].

On other hand, the fundamental number of alterations which found in the Delta VOC also have been formerly detected in other SARS-CoV-2 strains suggesting convergent evolution [18].

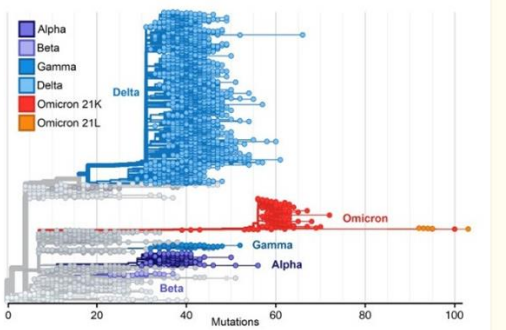

Fig. 3. Phylogenetic analysis of both initial SARS-CoV-2 and VOC isolates scaled according to their divergence compared to the Wuhan Hu-1 sequence.

Retrieved from Nextstrain on 18 January 2022 (https://nextstrain.org/ncov/gisaid/global?m=div).

3. SARS-CoV-2 Proteomes

At the proteome scale analysis, the SARS- CoV-2 is extremely similar to the related Beta-coronaviruses, unless there are some notable variances. For instance, the 3a, 3c and 8b accessory proteins are both closest to the SARS CoVs but varied in the amino acids numbers with the absence of 8a. On the other hand, the encoded structural proteins of pp1ab, pp1a, envelope, matrix, nucleocapsid as well as accessory protein 7a, showed a close relation to SARS-like bat CoVs, but regarding the spike S glycoprotein, the SARS-CoV-2 is closest to the bat CoVs [11].

Given the side of SARS-CoV-2 discrimination, the protein differences may represent in the amino acids chain length or amino acid substitutions, these variations may result in structural and functional segregation from other SARS-CoVs. Apart from identity presents in nonstructural protein 7 (nsp7), nsp13, envelope, matrix, or accessory proteins p6 and 8b, in total, there were 380 amino acid replacements between sequences of SARS- CoV-2and the parallel assent sequences of SARS and SARS-like viruses. For instance; 102, 61 and 27amino acid substitutions are located in nsp3, nsp2 and spike protein respectively. Moreover, four replacements inthe C-terminal of the receptor-binding subunit S1 domain are located in two peptides previously reported to be antigens for SARS- CoV, [21-23].

The researchers cannot give sensible captions about the reasons of presence or absence of those amino acid substitutions. The presence or absence of the substitutions could affect the host tropism and transmission property of the SARS-CoV-2compared to other parallel CoVs. The four major structural proteins of coronaviruses are the spike surface glycoprotein (S), small envelope protein (E), matrix protein (M), and nucleocapsid protein (N), fig (4:a,b,c,d), [24- 27].

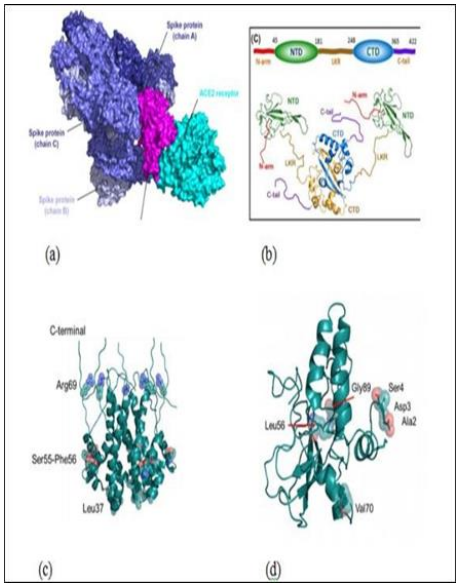

Fig. (4): the four structural proteins of SARS-CoV-2, (a): spike S protein; (b): nucleoprotein N protein; (c):envelope E protein; (d): matrix M protein;

3.1. Spike glycoprotein

Spike glycoprotein (S protein) is type 1membrane trimeric protein inserted in envelopprotein forming the spikes on the virus surfacewhich give the characteristic crown-likeappearance of corona viruses, [28].The receptor-binding domain (RBD) is a partof a protein sequence, life independent tertiary structure which binds to a specific atom or molecule. It is substantial because they help splicing, assembling, conformational changing and translating proteins.

Generally, the coronavirus S protein comprises two prime domains: the S1 subdomain at the N-terminus of the protein interposes binding tothe target receptor of the host cell and the C- terminus S2 domain enhances fusion of the virus membrane with cellular membrane of the host cell, [25]. It is reported that S protein of the novel coronavirus is modified via homologous recombination; a mixture of bat SARS-CoV and a not known Beta-CoV. The S1 subdomain of SARS-CoV2 includes 424–494 amino acids (AA), S1 C-terminal contain core structure of 5 antiparallel B-sheet (B1, B2, B3, B4, and B7) and short concave outer surface. It is declared that SARS-CoV-2 S protein contains 1273 AA; near SARS -CoV '1255 AA', but less than MERS-CoV '1353 AA'. Otherwise, MERS-CoV RBM is flat surface that is 4 anti-parallel B-sheet. The S2 subunit contains a fusion peptide, 2 heptad repeat domains HR1 and HR2, a transmembrane (TM) domain, and an endodomain (E) [25, 29-31].

The spike surface glycoprotein plays an essential role in binding to receptors on the host cell and determines host tropism. The RBD comes into direct contact with the extracellular binding site on ACE2 known as the peptidase domain (PD) [32]. There are two cleavage sites in the S protein, arginine R667 and R797. The R667 site is at the division between S1 and S2 and cleavage at the R797 site results in the final S2 polypeptide [33]. It is also reported that Spike glycoprotein of thenovel coronavirus is modified via homologous recombination. The spike glycoprotein of SARS-CoV-2 is the mixture of bat SARS-CoVand a not known Beta-CoV. There are some variations in RBD amino acid contact with ACE2 of SARS-COV2 in apart from SARS- CoV, fig (5:a,b) ; these differences make SARS-COV2 is stronger binding to ACE2 Via spike glycoprotein [34, 35].

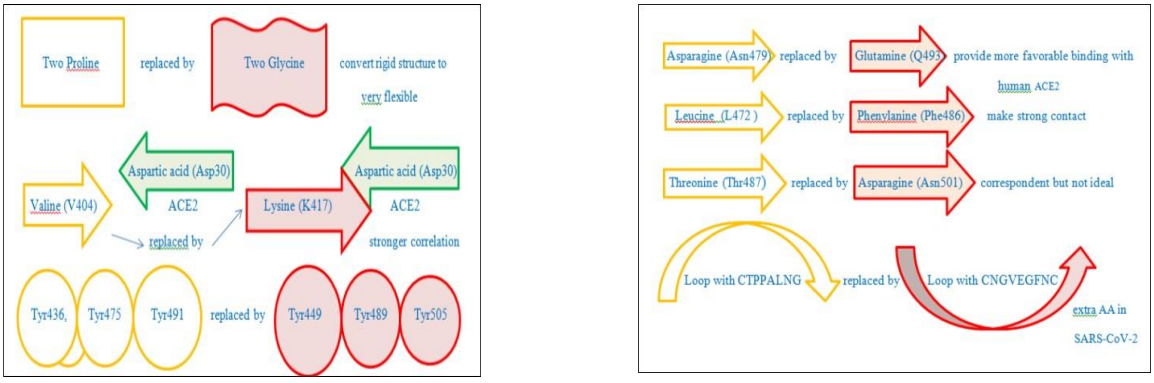

Fig. (5: a and b): Some variations in RBD (domains) amino acid contact with ACE2 between SARS-COV2 and SARS-CoV.

Finally, the C-terminal domain (CTD, 248– 365 AA) acts as a dimerization domain for oligomerization [36]. On the whole, the C-Terminal of N protein is identical in all coronavirus, so the variations are mainly focused on N-terminal domain. Unlike, the

common folded N-terminal nitrogenous base binding channel tail, SARS- CoV-2 has extended outward tail. This unique pattern leads to change charge distribution of N protein nucleotide surface making easier accessibility. Moreover, the phosphate groupbinding site in SARS-CoV-2 N-NTD has larger side chain amino acid compared to other coronaviruses. Finally, the edge of nitrogenous base in SARS-CoV-2 N-NTD has Arg 89 amino acid compared with Tyr 102 causing increase polar properties. As for the C-terminal domain fragment contains a short multiple hydrophobic interaction dimerization core besides positive charged regions, the N protein is able to bind to single-stranded RNA (ssRNA), single-stranded DNA (ssDNA) and double-stranded DNA (dsDNA) with greater affinities, [37].

Alterations in the VOC spike protein

Spike protein exhibits more than 60% of all mutations distinguishing the Omicron VOC from the original SARS-CoV-2 strain. The Omicron 21K and 21L Spike proteins share 20 amino acid alterations, with an additional 19 21K- and 14 21L-specific mutations, respectively, although even higher numbers of mutations might be found in some viral isolates [38]. In comparison, the Delta VOC differs by only 9 amino acid alterations from the initial viral isolate, two of which (E478K and D614G) are shared by the Omicron VOC. Twelve out of the 31 alterations in the S1 region of the Omicron Spike are located in the N-terminal domain (NTD), 15 in the ACE2 receptor-binding domain (RBD), and five at the C terminus. A total of 10 amino acid changes are located in the receptor-binding motif (RBM) that makes direct contact with the human ACE2 receptor [39].

3.2. Nucleocapsid protein (N)

Nucleocapsid protein "N" protein is a basic structural protein that protein binds to the virus genome forming ribonucleoprotein known as nucleocapsid. N protein is more stable than spike protein so it becomes a target for antiviral therapy, SARS-CoV-2 N protein sequence contains 419 amino acid and shares 89.74 % SARS-CoV N protein but only 48.59 % of MERS-CoV one, [37]. The three structural domains have characteristics common to all coronavirus N proteins, N-terminal domain (NTD, 45–181 AA) of the SARS-CoV- 2 N protein acts as RNA-binding domain, then Ser/Arg (SR)-rich linker is responsible for phosphorylation. N protein is considered a multifunctional; which binds with RNA for genomic protection, ensuring replication and transmission. Also associates with M protein during assembly. Moreover, N protein regulates host pathogens interaction such as actin reorganization, host cell cycle progression, and apoptosis. Correspondingly, it is able to induce immune response, so it is considered highly antigenic. Also, it can escape from immune system via prohibition of type I interferon and cytokines after virion infected the host cells, [40].

3.3. Envelope (E) protein

The SARS-CoV2 E protein is a small sized; 75 AA, coded by E gene and considered a critical component of purified virus particles; act as integral trans-membrane for ion channel activity and responsible for virion envelope morphogenesis, [41]. Also CoV E may have an anti-apoptotic function by suppressing the Unfolded Protein Response (UPR), probably as a survival mechanism essential for virus dissemination, [42]. Generally, the E protein in all coronaviruses comprised primary and secondary structures containing three domains. N-terminal domain is a short (8 amino acids) hydrophilic then followed by an unusually long

hydrophobic transmembrane domain (25–30 amino acids with 2–4 cysteine residues) that formed a helical hairpin. The ending hydrophilic carboxyl C-terminal domain is long (40 amino acid), ubiquitinated requiredfor proper virus assembly, [43].

In contrary of the common observation among *Coronaviridea* members that the E protein corresponding sequence in between is few; (e.g. MERS-CoV is 69, 5%) there is a 94.74% identity shared between SARS- CoV2 'E protein' sequence and that of SARS-CoV. SARS-CoV2 E protein is a pentameric and one unit of E protein consist of seven α- helices and eight loops , so that the ion channel activity of 'SARS-CoV2 E' proteins is modulated via pentameric ion channel, [44]. There are four differences between SARS-CoV and SARS- CoV-2 E proteins; two replacements in BH-3like helix and two in N-terminal, fig. (6),[41].

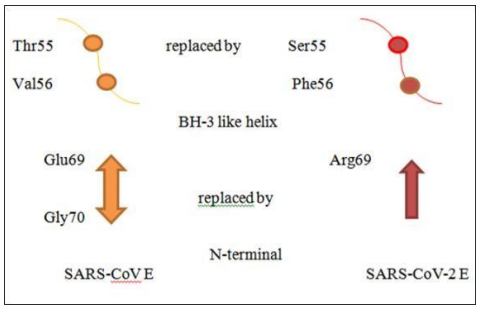

Fig. (6): The variations in amino acids of E proteinbetween SARS-COV2 and SARS-CoV.

3.4. Matrix M protein

It is essential in virus assembly, and plays an important role; turns cellular membranes into factories where virus and host factors join to make new virus particles. The M proteins fromSARS-CoV-2 as well as SARS-CoV, and MERS-CoV are targeted to the vicinity of the Golgi apparatus. It is suggested that M protein promotes assembly by interacting with the viral ribonucleoprotein (N protein) and S glycoproteins at the budding site, besides creation of a network of M-M interactions which have the ability to exclude some host membrane proteins from the viral envelope, [45, 46].

Differences in non- spike structural proteins:

Changes are found in all three non-Spike structural proteins: E (T9I), M (D3G, Q19E, A63T), and N (P13L, Δ31–33, R203K, G204R) associated to cell-mediated response evading. Three of the four changes in the N protein of Omicron (P13L, R203K, G204R) have been previously noticed in other VOC. [47]. Changes similar to R203K and G204R in N are also found in the Alpha and Gamma VOCs and may be associated with increased viral loads and sub genomic RNA expression [48]. However, the functional relevance of the T9I change in the E protein as well as the D3G, Q19E, and A63T substitutions in the M protein is currently unclear. Compared to other VOCs, the Omicron VOC has a surprisingly low number of changes in its accessory proteins. Only ORF9b contains changes of P10S and Δ27–29 but may just be changed as it overlaps the gene encoding the N protein [49].

3.5. Non- structural proteins of SARS-CoV-2 (Nsp):

The nonstructural proteins are coded by *ORF*1ab which encodes the *ORF*1ab poly-protein that contains from 1

– 7096 amino acid fig (7).

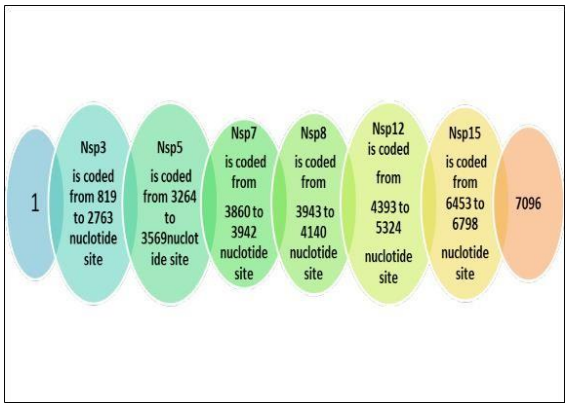

Fig. (7): *ORF*1ab sites of different non-structural proteins of SARS-CoV2.

3.6. Proteases (NSPS 3& NSPS 5)

Nsp 3 is a papain like protease (PLpro), while Nsp 5 constitutes 3-chymotrypsin-like protease (3CL pro), and Mpro which are encoded by *orf*1ab gene. These enzymes are important for virus replication and the translation of the polypeptide from the genomic RNA to protein component, [50]. The three- dimensional structure of SARS-COV-2Mpro is highly similar to SARS-CoV Mpro, sharing about 96%. The Nsp5 is about 305 amino acid, and the differences between them are only 12 amino acid that are at positions 33, 44, 63, 84, 86, 92, 132, 178, 200, 265, 283 and 284, and on the same line, 3C- like protease sequence is 100% identical, [51].

Nsp5 is asymmetric unit contains only one but two of these polypeptides associate to form a dimer designated (protomer A and B). Each protomer consist of 3 domain in which domains I and II (10-99 and 100-182 AA, respectively) are six-stranded antiparallel β-barrels that contain the substrate-binding site between them in the cleft. Domain III (198-303 AA) contains five α- helices that show globular cluster and shared in regulating dimerization of the Mpro. A Long loop (185– 200 AA) links between domain II and domain III that participate in the formation of the substrate binding pocket. There are twelve cysteine residues across the protein molecule with six are buried in the core and the other six are exposed to the surface, one of them (C145) located in the catalytic center that lies in a cleft between domain I and domain II, [52]. 3C like protease and papain like protease are important enzymes for Viral RNA translation to polyprotein process that they work on more than 11 cleavage sites on the large polyprotein 1ab (replicase 1ab), [51].

3.7. Nsp 9

Nsp9 in SARS-CoV2 is encoded from *orf*1ab polyprotein and share 97% sequence of Nsp9 in SARS-CoV. The apo-Nsp9 SARS-CoV2 structure is related closely to that belongs to SARS, and also like other Nsp9 homologues in which it exhibits an unusual fold that is not found outside of coronaviruses. The core of fold is a small 6-stranded enclosed β-barrel that showed series of extended loops projected outward. Two loops (β2-3- and β3-4) are projected from open face of barrel carry positive charge, rich glycine, and think to involve RNA binding, [53].

3.8. Nsp12 (RNA-dependent RNApolymerase RdRp)

Nsp12 in SARS-COV -2 is encoded from orf1ab polyprotein and is closely related to thatof picoronavirus (about 500 AA) apart to thatof The SARS-CoV nsp12 (932 AA). SARS - CoV2 differs from SARS-CoV in 31AA mutations; 22 AA are located in N-terminal of nsp 12 , the remaining nine amino acids are located in C-terminal of nsp12 and one of them(S783A) is a non-conservative mutation, (Robert and Ward, 2019) [54]. SARS-COV-2 NSP 12 consists of C-terminal domain, and right-hand polymerase domain, then Nidovirus unique N- terminal extension domain which contain nidovirus RdRp- associated nucleotide transferase (NiRAN) besides addition N-terminal β-hairpin. Finally, there is an interface domain connect between the right hand and NiRAN, [55].

Polymerase domain (right hand) in SARS-COV-2 structure is like structure of others inhuman associated coronaviruses that containspolymerase motifs A-G that form active site, template/primer entry, nucleoside triphosphate (NTP) entry, and nascent strand exit paths that is positively charged andsolvent-accessible, [56]. Onthe other hand, NiRAN in which a portion of the N- terminal extension domain (residues 4 to 28 and 51 to 249) consists of two (other Nidoviruses have eight) helices with a five-stranded β-sheet at the N terminus, and isconsidered a genetic marker for arrangementof Nidovirales that no viral or cellularhomologs is identified, [57]. Nsp12 alone is of little activity, so it requires co-factor such as Nsp 7 and Nsp 8 for stimulating its polymerase activity that catalyzes the synthesis of viral RNA and is necessary in replication and transcription [58].

3.9. Nsp 7 and Nsp 8

Nsp7 and Nsp 8 in SARS-COV-2 are encoded from orf1ab polyprotein, and found to be bind to Nsp12 to form an active polymerase "apo RdRp complex". This complex includes one nsp12, one nsp7 and two nsp8; the polymerase domain of nsp12 binds to nsp7 and nsp8, leaving the other nsp8 molecule sited on the top of the finger subdomain and interacting with the interface domain that make structure more flexible and stabilize, [55].

3.10. Nsp 10 and Nsp 14

Particularly, the nonstructural protein 10 (Nsp10) and its cofactor Nsp14 induce exonuclease activity that primarily minimize the accumulation of mutations in the viral genome [59].

3.11. Nsp 15 endoribonuclease

Nsp15 in SARS-COV-2 is encoded from orf1ab polyprotein and shares 88% sequence identity and 95% similarity with SARS-CoV Nsp15. SARS-CoV2 Nsp15 is two monomers as asymmetric unit. The structure of SARS-CoV-2 Nsp15 monomer is very similar to other Nsp15s from coronaviruses, [60].

Nsp15 consists of three domain that is N- terminal domain, middle terminal and C-terminal catalytic NendoU domain with several loops. N-terminal domain consists of three β strands as an antiparallel β-sheet (strands β1, β2, and β3) wrapped, around two α-helices (α1 and α2). Middle domain consist of 12 βstrands (4 β to 13 β strands) and three short helices and finally the C-terminal catalytic NendoU domain consist of 6 βstrands as two antiparallel β-sheets with their edges is site for catalytic substance. The concave surface of the β-sheets is flanked by five α-helices. Previous structure is a subunit of monomer, but latterly this

monomer arranges for forming hexamer that is important for enzymatic activity. C-terminal NendoU monomers assemble into a double- ring hexamer. The largest difference between SARS-CoV-2 and SARS-CoV seems to occur in the position of middle domains. The differences with H-CoV-229E are still more significant and show shifts in positions of α-helices, β-sheets and loops, [61].

Non-Spike alterations in the Omicron Proteome

A total of eighteen amino acid-changing mutations in the genome of Omicron VOC are located outside the Spike gene. A variety of alterations, such as K38R, Δ1265, and A1892T in Nsp3; P132H in Nsp5; I189V in Nsp6; P323L in Nsp12; and I42V in Nsp14, are present only in the Omicron VOC, and their potential functional consequences remain to be determined. Several mutations affect Nsp3, which deISGylates the RNA sensor MDA5 and signal transduction factors to antagonize innate immune responses [62, 63].

The other prime protease; Nsp5 and the signal adaptor mitochondrial antiviral-signaling protein (MAVS) known to enhance innate immune escape [64, 65], has a P132H mutation.

In support of a selective advantage, substitution of T492I in Nsp4 involved in viral replication is also present in the Lambda, Delta, and Gamma VOCs of SARS-CoV-2. A mutation similar to the 3-amino-acid deletion (Δ105–107) in a loop in the Omicron Nsp6 predicted by alpha-fold is also found in the Alpha, Beta, and Gamma VOCs and may be used for PCR-based differentiation from other SARS-CoV-2 variants [66].

In addition, it has been hypothesized that this mutation may promote innate immune evasion of the virus (58). Nsp7, Nsp8, and Nsp12 comprise the active RNA-dependent RNA polymerase complex of SARS-CoV-2. Notably, the P323L mutation in Nsp12 is at the interface with Nsp8 (Fig. 2C) [67].

Conclusion

A novel respiratory virus has been emerged at the end of 2019, the disease started in China then spread rapidly and widely all over the world result in until now over 15 million patient and over 600000 deaths. A novel virus belonged to *Coronaviridae* has been confirmed as the causative agent and termed SARS-CoV-2. The primary zoonotic source of the novel virus still not defined, may bat was suggested as the main reservoir. The phylogenic studies revealed that the SARS-CoV-2 is set in beta corona viruses' clade, and is closer to SARS- like bat CoVs. The genomic structure and constitutional proteins displayed some mutations and variances in comparison to other related human corona viruses.

REFERENCES

1. Wu D, Wu T, Liu Q, *et al*., (2020): The SARS- CoV-2 outbreak: what we know. *Int J InfectDis*, **94**: 44–8.
2. Mahase E, (2020): Covid-19: WHO declares pandemic because of "alarming levels" of spread, severity, and inaction. *BMJ* 2020; **368**: m1036.
3. World Health Organization, (2020):Laboratory testing for coronavirus disease 2019 (COVID-19) in suspected human cases:interim guidance, 2 March 2020, 2020.
4. Lai CC, Shih TP, Ko WC, *et al*., (2020): Severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) and corona virus disease-2019 (COVID19): the epidemic and the challenges. *Int J Antimicrob Agents* 2020:105924.

5. Wang C, Horby PW, Hayden FG, *et al*.,(2020): A novel coronavirus outbreak ofglobal health concern. *The Lancet* **395** (10223), DOI:10.1016/S0140- 6736(20)30185-9.
6. World Health Organization, (2020): Novel coronavirus situation dashboard. Geneva, Switzerland: WHO, 2020.
7. Jung C, Kmiec D, Koepke L, Zech F, Jacob T, Sparrer KMJ, Kirchhoff F. Omicron: What Makes the Latest SARS-CoV-2 Variant of Concern So Concerning? J Virol. 2022 Mar 23;96(6):e0207721.
8. Zhu N, Zhang D, Wang X, et al., (2020): A Novel Coronavirus from Patients with Pneumonia in China, 2019. N Engl J Med, 382 (2020): 727-733.
9. Cui J, Li F, and Shi ZL, (2019): Origin and evolution of pathogenic coronaviruses. *Nat. Rev. Microbiol*. 17: 181–192.
10. Malika YS , Shubhankar S , Sudipta B, *et al* (2020): Emerging novel coronavirus (2019-nCoV)—current scenario, evolutionary perspective based on genome analysis and recent developments. *Veterinary Quarterly* **40**: (1), 68–76.
11. Chen Y, Liu Q, Guo D, *et al*., (2020): Emerging coronaviruses: genome structure, replication, and pathogenesis. *J Med Virol*.
12. Zhou P, Yang XL, Wang XG, et al., (2020): Discovery of a novel coronavirus associated with the recent pneumonia outbreak in humans and its potential bat origin. bioRxiv.
13. Song Z, Xu Y, Bao L, *et al*., (2019): From SARS to MERS, Thrusting Coronaviruses into the Spotlight. *Viruses*, **11**: E59.
14. Meier C, Aricescu AR, Assenberg R, *et al*., (2006): The crystal structure of ORF-9b, a lipid binding protein from the SARS coronavirus. *Structure*, **14**(7):1157-65.
15. Lu R, Zhao X, Li J, *et al*., (2020): Genomic characterization and epidemiology of 2019 novel coronavirus: implications for virus origins and receptor binding. *Lancet*, **395**(10224):565–574.
16. Wu A, Peng Y, Huang B, *et al*., (2020): Genome composition and divergence of the novel coronavirus (2019-nCoV) originatingin China. *Cell Host Microbe* 2020.
17. Bansal K, Kumar S. 2021. Mutational cascade of SARS-CoV-2 leading to evolution and emergence of omicron variant. bioRxiv.
18. Planas D, Veyer D, Baidaliuk A, Staropoli I, Guivel-Benhassine F, Rajah MM, Planchais C, Porrot F, Robillard N, Puech J, Prot M, Gallais F, Gantner P, Velay A, Le Guen J, Kassis-Chikhani N, Edriss D, Belec L, Seve A, Courtellemont L, Péré H, Hocqueloux L, Fafi-Kremer S, Prazuck T, Mouquet H, Bruel T, Simon-Lorière E, Rey FA, Schwartz O. 2021. Reduced sensitivity of SARS-CoV-2 variant Delta to antibody neutralization. Nature 596:276–280.
19. Karim F, Moosa MYS, Gosnell BI, Cele S, Giandhari J, Pillay S, Tegally H, Wilkinson E, San JE, Msomi N, Mlisana K, Khan K, Bernstein M, Manickchund N, Singh L, Ramphal U, COMMIT-KZN Team, Hanekom W, Lessells RJ, Sigal A, de Oliveira T. 2021. Persistent SARS-CoV-2 infection and intra-host evolution in association with advanced HIV infection. medRxiv. https://www.medrxiv.org/content/10.1101/2021.06.03.21258228v1.
20. Wei C, Shan K-J, Wang W, Zhang S, Huan Q, Qian W. 2021. Evidence for a mouse origin of the SARS-CoV-2 Omicron variant. bioRxiv. https://www.biorxiv.org/content/10.1101/2021.12.14.472632v1.
21. Guo JP, Petric M, Campbell W, *et al*.,(2004): SARS corona virus peptides recognized by antibodies in the sera ofconvalescent cases. *Virology* **324**: 251–256.

22. De-Ming Y, Tai-Jay Ch, Mong-Lien W, *et al*., (2020): Hunting coronavirus severe acute respiratory syndrome coronavirus 2 (2019 novel coronavirus): From laboratory testing back to basic research. *J Chin Med Assoc*. **5**: 10.
23. Renhong Y, Yuanyuan Z, Yaning L, *et al*., (2020): Structural basis for the recognition of SARS-CoV-2 by full-length human ACE2. *Science*. 27; **367**(6485): 1444–1448.
24. Chang Ch, Ming-Hon H, Chi-Fon Ch, *et al*., (2014): The SARS coronavirus nucleocapsid protein – Forms and functions. *Antiviral Research*, **103**(3): 39-50.
25. Li F, (2016): Structure, Function, and Evolution of Coronavirus Spike Proteins. *Annu. Rev. Virol*. **3**: 237–261.
26. Bianchi M, Domenico B, Marta S, Pascarella A, *et al*., (2020): Sars-CoV-2 Envelope andMembrane Proteins: Differences from Closely Related Proteins Linked to Cross-species Transmission? Preprinted, April2020.
27. Max Perutz Labs, (2020): COVID-19 Q&A's- Max Perutz Labs.
28. Fang L, (2020): Structure, Function, and Evolution of Coronavirus Spike Proteins. *Annual Review of Virology*, **3**:237-261.
29. Rota PA, Oberste MS, Monroe SS, *et al*., (2003): Characterization of a novelcoronavirus associated with severe acute respiratory syndrome. *Science*. **300** (5624):1394-9.
30. Benvenuto D, Giovanetti A, Ciccozzi S, *et al*., (2020): The 2019-new coronavirus epidemic:evidence for virus evolution. *J Med Virol*,
31. Li B, Si H-R, Zhu Y, *et al*., (2020): Discovery of Bat Coronaviruses through Surveillance and Probe Capture-Based Next Generation Sequencing. *mSphere*, doi:10.1128/mSphere.00807-19.
32. Bosch BJ, van der Zee R, CornelisAM, *et al*., (2003): The Coronavirus SpikeProtein Is a Class I Virus Fusion Protein: Structural and Functional Characterization of the Fusion Core Complex. Journal ofVirology, 77(16): 8801–8811.
33. Millet JK and Whittaker GR, (2015): Host cell proteases: Critical determinants of coronavirus tropism and pathogenesis. *VirusRes*.**16** (202): 120–134.
34. Li Q, Guan P, Wu X, *et al*., (2020): Earlytransmission dynamics in Wuhan, China, of novel coronavirus-infected pneumonia. *N Engl J Med* (2020), doi:10.1056/NEJMoa2001316.
35. Xu X, Chen P, Wang J, *et al*., (2020): Evolution of the novel coronavirus from the ongoing Wuhan outbreak and modeling of its spike protein for risk of human transmission. *Sci China Life Sci*, **63**(3):457-460.
36. Dinesh DC, Chalupska D, Silhan J, *et al*., (2020): Structural basis of RNA recognition by the SARS-CoV-2 nucleocapsid phosphoprotein.
37. Kang S, Yang M, Hong Z, *et al*., (2020):Crystal structure of SARS-CoV-2 nucleocapsid protein RNA binding domain reveals potential unique drug targeting sites. *Acta Pharmaceutica Sinica B*. doi: 10.1016/j.apsb.2020.04.009.
38. Sarkar R, Lo M, Saha R, Dutta S, Chawla-Sarkar M. 2021. S glycoprotein diversity of the Omicron variant. medRxiv. https://www.medrxiv.org/content/10.1101/2021.12.04.21267284v2.
39. Escalera A, Gonzalez-Reiche AS, Aslam S, Mena I, Pearl RL, Laporte M, Fossati A, Rathnasinghe R, Alshammary H, van de Guchte A, Bouhaddou M, Kehrer T, Zuliani-Alvarez L, Meekins DA, Balaraman V, McDowell C, Richt JA, Bajic G, Sordillo EM, Krogan N, Simon V, Albrecht RA, van Bakel H, Garcia-Sastre A, Aydillo T. 2021. SARS-CoV-2 variants of concern have acquired mutations associated with an increased spike cleavage. bioRxiv. https://www.biorxiv.org/content/10.1101/2021.11.09.467693v1.

40. Lin Z, Gao Q, Qian F, *et al*., (2020): The nucleocapsid protein of SARS-CoV-2 abolished pluripotency in human induced pluripotent stem cells. *bioRxiv* doi.org/10.1101/2020.03.26.010694.
41. Navratil V, Lionnard L, Longhi S, *et al*., (2020): The severe acute respiratory syndrome coronavirus2 (SARS-CoV-2) envelope (E) protein harbors a conserved BH3-like sequence. *bioRxiv*.
42. DeDiego ML, *et al*., (2011): Severe acute respiratory syndrome coronavirus envelope protein regulates cell stress response and apoptosis. *PLoS pathogens*, **7**: e1002315.
43. Schoeman D, and Fielding BC, (2019): Coronavirus envelope protein: current knowledge. *Virology journal*, **16**(1): 69.
44. Gupta MK, Vemula S, Donde R, *et al*.,(2020): In-silico approaches to detect inhibitors of the human severe acuterespiratory syndrome coronavirus envelopeprotein ion channel. *Journal of BiomolecularStructure and Dynamics*, 1-17.
45. Neuman BW, Kiss G, Kunding AH, *et al*.,(2011): Structural analysis of M protein in coronavirus assembly andmorphology. Journal of StructuralBiology. **174**:11–22.
46. Hasöksüz M, Kiliç S, Saraç F, *et al*., (2020): Coronaviruses and SARS-COV-2. *Turk J Med Sci*. **50** (SI-1):549-556.
47. de Silva TI, Liu G, Lindsey BB, Dong D, Moore SC, Hsu NS, Shah D, Wellington D, Mentzer AJ, Angyal A, Brown R, Parker MD, Ying Z, Yao X, Turtle L, Dunachie S, COVID-19 Genomics UK (COG-UK) Consortium, Maini MK, Ogg G, Knight JC, ISARIC4C Investigators, Peng Y, Rowland-Jones SL, Dong T. 2021. The impact of viral mutations on recognition by SARS-CoV-2 specific T cells. iScience 24:103353. 10.1016/ j.isci. 2021. 103353.
48. Leary S, Gaudieri S, Parker MD, Chopra A, James I, Pakala S, Alves E, John M, Lindsey BB, Keeley AJ, Rowland-Jones SL, Swanson MS, Ostrov DA, Bubenik JL, Das S, Sidney J, Sette A, COVID-19 Genomics UK (COG-UK) Consortium, de Silva TI, Phillips E, Mallal S. 2021. Generation of a novel SARS-CoV-2 sub-genomic RNA due to the R203K/G204R variant in nucleocapsid: homologous recombination has potential to change SARS-CoV-2 at both protein and RNA level. bioRxiv. https://www.biorxiv.org/content/ 10.1101 /2020.04.10.029454v4.
49. Nchioua R, Schundner A, Klute S, Noettger S, Zech F, Koepke L, Graf A, Krebs S, Blum H, Kmiec D, Frick M, Kirchhoff F, Sparrer KMJ. 2021. The Delta variant of SARS-CoV-2 maintains high sensitivity to interferons in human lung cells. bioRxiv. https://www.biorxiv.org/content/10.1101/2021.11.16.468777v1.
50. Shankar AK, Bhanu D, Alluri A, *et al*.,(2020): Whole Genome Sequence Analysis and Homology Modelling of Main Protease and Non-Structural Protein 3 of the SARS-CoV-2 reveals an Aza-Peptide and a LeadInhibitor with Possible Antiviral Properties. *New Journal of Chemistry*. doi.org/10.1039/D0NJ00974A.
51. Zhang L, Lin D, Sun X, et al., (2020): Crystal structure of SARS-CoV-2main protease provides a basis for design of improved α- ketoamide inhibitors. Science, 368 (6489): 409-412.
52. Jin Z, Du X, Xu Y, *et al*., (2020): Structure of Mpro from COVID-19 virus and discovery of its inhibitors. *Nature* doi: 10.1038/s41586-020-2223-y.
53. Littler D, Gully B, Colson RN, *et al*., (2020):Crystal structure of the SARS-CoV-2 non-structural protein 9, Nsp9. *bioRxiv*.
54. Robert NK and Ward AB, (2019): Structure of the SARS-CoV nsp12 polymerase bound to nsp7 and nsp8 co-factors. *Nat Commun* **10**: 2342.

55. Peng Q, Peng R, Yuan B, *et al*., (2020): Structural and biochemical characterization of nsp12-nsp7-nsp8 core polymerase complex from COVID-19 virus. *bioRxiv.*
56. Gao Y, Yan L, Huang Y, *et al*., (2020):Structure of the RNA-dependent RNApolymerase from COVID-19 virus. *Science*
57. Shannon A, Le NTT, Selisko B, *et al*., (2020): Remdesivir and SARS-CoV-2: Structural requirements at both nsp12 RdRp and nsp14 Exonuclease active-sites. *AntiviralResearch*, 104793.
58. Robert NK, Ward AB, (2019). Structure of the SARS-CoV NSP12 polymerase bound to NSP7 and NSP8 co-factors. *bioRxiv* https://www.biorxiv.org/content/10.1101/551986v1.full
59. Saramago M, Bárria C, Costa VG, Souza CS, Viegas SC, Domingues S, Lousa D, Soares CM, Arraiano CM, Matos RG. 2021. New targets for drug design: importance of nsp14/nsp10 complex formation for the 3'-5' exoribonucleolytic activity on SARS-CoV-2. FEBS J 288:5130–5147. 10.1111/febs.15815.
60. Fu X, Li D, and Sun Y, (2020): orf1ab polyprotein [Severe acute respiratory syndrome coronavirus 2] GenBank: QIA20042.1.
61. Kim Y, Jedrzejczak R, Maltseva NI, *et al*., (2020): Crystal structure of Nsp15 endoribonuclease NendoU from SARS-CoV-*2. Protein Science*.
62. Cong Z, Evans JP, Qu P, Faraone J, Zheng Y-M, Carlin C, Bednash JS, Zhou T, Lozanski G, Mallampalli R, Saif LJ, Oltz EM, Mohler P, Xu K, Gumina RJ, Liu S-L. 2021. Neutralization and stability of SARS-CoV-2 Omicron variant. bioRxiv. https://www.biorxiv.org/content/10.1101/ 2021 .12 .16.472934v1.
63. Keeton R, Tincho MB, Ngomti A, Baguma R, Benede N, Suzuki A, Khan K, Cele S, Bernstein M, Karim F, Madzorera SV, Moyo-Gwete T, Mennen M, Skelem S, Adriaanse M, Mutithu D, Aremu O, Stek C, du Bruyn E, Mescht MAVD, de Beer Z, de Villiers TR, Bodenstein A, van den Berg G, Mendes A, Strydom A, Venter M, Grifoni A, Weiskopf D, Sette A, Wilkinson RJ, Bekker L-G, Gray G, Ueckermann V, Rossouw T, Boswell MT, Bihman J, Moore PL, Sigal A, Ntusi NAB, Burgers WA, Riou C. 2021. SARS-CoV-2 spike T cell responses induced upon vaccination or infection remain robust against Omicron. medRxiv. https:// www. medrxiv. org/ content/ 10.1101/ 2021.12.26.21268380v1.
64. Hayn M, Hirschenberger M, Koepke L, Nchioua R, Straub JH, Klute S, Hunszinger V, Zech F, Prelli Bozzo C, Aftab W, Christensen MH, Conzelmann C, Müller JA, Srinivasachar Badarinarayan S, Stürzel CM, Forne I, Stenger S, Conzelmann K-K, Münch J, Schmidt FI, Sauter D, Imhof A, Kirchhoff F, Sparrer KMJ. 2021. Systematic functional analysis of SARS-CoV-2 proteins uncovers viral innate immune antagonists and remaining vulnerabilities. Cell Rep 35:109126. 10.1016/j.celrep.2021.109126.
65. Liu Y, Qin C, Rao Y, Ngo C, Feng JJ, Zhao J, Zhang S, Wang T-Y, Carriere J, Savas AC, Zarinfar M, Rice S, Yang H, Yuan W, Camarero JA, Yu J, Chen XS, Zhang C, Feng P. 2021. SARS-CoV-2 Nsp5 demonstrates two distinct mechanisms targeting RIG-I and MAVS to evade the innate immune response. mBio 12:e02335-21. 10.1128/mBio.02335-21.
66. Nchioua R, Schundner A, Klute S, Noettger S, Zech F, Koepke L, Graf A, Krebs S, Blum H, Kmiec D, Frick M, Kirchhoff F, Sparrer KMJ. 2021. The Delta variant of SARS-CoV-2 maintains high sensitivity to interferons in human lung cells. bioRxiv. https:// www. biorxiv.org /content /10.1101 /2021.11.16.468777v1.
67. Hillen HS, Kokic G, Farnung L, Dienemann C, Tegunov D, Cramer P. 2020. Structure of replicating SARS-CoV-2 polymerase. Nature 584:154–156. 10.1038/s41586-020-2368-8.