Review

Genome composition and genetic characterization of SARS-CoV-2

Ayman M. Al-Qaaneh, Thamer Alshammari, Razan Aldahhan, Hanan Aldossary, Zahra Abduljaleel Alkhali, J. Francis Borgio

A Department of Genetic Research, Institute for Research and Medical Consultations (IRMC), Imam Abdulrahman Bin Faisal University, Dammam 31441, Saudi Arabia

B Drug Information Center, Pharmacy Services Department, Johns Hopkins Aramco Healthcare (JHAH), Dhahran 31311, Saudi Arabia

C Department of Epidemic Diseases Research, Institute for Research and Medical Consultations (IRMC), Imam Abdulrahman Bin Faisal University, Dammam 31441, Saudi Arabia

ABSTRACT

SARS-CoV-2 is a type of Betacoronaviruses responsible for COVID-19 pandemic disease, with more than 1.745 million fatalities globally as of December-2020. Genetically, it is considered the second largest genome of all RNA viruses with a 5' cap and 3' poly-A tail. Phylogenetic analyses of coronaviruses reveal that SARS-CoV-2 is genetically closely related to the Bat-SARS Like-Corona virus (Bat-SL-Cov) with 96% whole-genome identity. SARS-CoV-2 genome consists of 15 ORFs coded into 29 proteins. At the 5' terminal of the genome, we have ORF1ab and ORF1a, which encode the 1ab and 1a polypeptides that are proteolytically cleaved into 16 different nonstructural proteins (NSPs). The 3' terminal of the genome represents four structural (spike, envelope, matrix, and nucleocapsid) and nine accessory (3a, 3b, 6, 7a, 7b, 8b, 9a, 9b, and orf10) proteins. As the number of COVID-19 patients increases dramatically worldwide, there is an urgent need to find a quick and sensitive diagnostic tool for controlling the outbreak of SARS-CoV-2 in the community. Today, molecular testing methods utilizing viral genetic material (e.g., PCR) represent the crucial diagnostic tool for the SARS-CoV-2 virus despite its low sensitivity in the early stage of viral infection. This review summarizes the genome composition and genetic characterization of the SARS-CoV-2.

© 2021 The Author(s). Published by Elsevier B.V. on behalf of King Saud University. This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/4.0/).
1. Introduction

As a human pathogen, coronaviruses (CoVs) have been drawing a great consideration as a global health threat. They are types of viruses that can attack many systems in humans and vertebrates, such as the respiratory system, digestive system, nervous system, and liver (Chen et al., 2020). It belongs to the Coronavirinae subfamily, a subset of the Coronaviridae family and Nidovirales order (Chan et al., 2020). Genetically they are categorized into four distinct genera; Alpha, Beta, Gamma, and Deltacoronavirus (Wu et al., 2020). While bats and rodents consider the primary sources of alpha and Betacoronaviruses, the avian species are the primary sources of Gamma and Deltacoronaviruses (Z. W. Ye et al., 2020).

Among four betacoronaviruses that can infect humans, severe acute respiratory syndrome coronavirus (SARS-CoV) and Middle East respiratory syndrome coronavirus (MERS-CoV) are considered the most serious pathogens able to cause severe lower respiratory tract infection. They are responsible for both the 2003 Severe acute respiratory syndrome (SARS) endemic and 2012 Middle East respiratory syndrome (MERS) endemic, respectively (Ahmed et al., 2020; Cui et al., 2019; Shereen et al., 2020). Recently we had a new addition to the betacoronavirus family, a novel coronavirus-2019 (2019-nCoV) that has been named on February 12, 2019 by World Health Organization (WHO) as SARS-CoV-2, which is responsible for the pneumonia-like disease since December 2019 (Guo et al., 2020; Gurwitz, 2020). It is well documented and approved that these three betacoronaviruses have shown the possibility of animal to human and human to human transmission (J. Y. Li et al., 2020). As of today, December 25, 2020, around 79 million patients have been diagnosed with SARS-CoV-2 and 1.745 million deaths all over the world (“COVID-19 Map - Johns Hopkins Coronavirus Resource Center,” n.d.). In addition, the reported cases geographically vary and are related to local responses divergence of hosts. Yet, if the biological consequences of hosts are different due to alterations in the SARS-CoV-2 genome, this could lead to serious complications for combating the pandemic (Garvin et al., 2020). Like any other Betacoronaviruses, Bats represent the primary host of SARS-CoV-2, which then transfers to a secondary amplification host (e.g., palm civets and Raccoon dogs) before being able to infect humans (C. Li et al., 2020; Shereen et al., 2020). This review will discuss the genome composition and genetic characteristics of SARS-CoV-2, the mechanism of SARS-CoV-2 genome replication and translation, the function of different proteins expressed by these genes, and using molecular genetic techniques in diagnosis SARS-CoV-2.

2. Different types of Coronaviruses, including human coronaviruses

Since 1960s (the date of the first recognition of human coronaviruses), seven types of human coronaviruses have been identified (Seah and Agrawal, 2020). They are classified based on their protein sequence into; Alphacoronavirus, Betacoronavirus, Gammaronavirus, and Delta coronavirus. The alpha and betacoronaviruses are the two genera that infect mammals, whereas Gamma and Delta coronaviruses infect birds (Wu et al., 2020).

Generally, RNA viruses encode for proteins that are required to control the host cells. For instance, they encode for the protein types needed to assemble the new viral particles. Further, RNA viruses are well-known for their high mutation rate, which supports them to rapidly adapt in host cells (Manfredonia et al., 2020).

Human coronaviruses, in general, share core genomic features, e.g., I. They are vast and highly conserved RNA genomes viruses. II. Genome structure is organized as a 5’-untranslated region (UTR)-replicase-S-M-N-UTR-3’ with the genome function as a replicase gene. III. All family members have two enveloped (E) protein species, helical Nucleocapsid (N), spike (S) and membrane (M) proteins. They all share the expression of various nonstructural genes (Snijder et al., 2016). In addition, they all are required to achieve an infectious cycle of replication, starting from entering, assembling, packaging, and releasing the new virus particle inside of the host (Hassan et al., 2020). The two overlapping open reading frames (ORFs) ORF1a&1ab represent the two-third of the genome’s upstream region (Fig. 1). The general genetic characteristic that distinguishes Alphacoronavirus from other coronaviruses is that they have a unique type of nonstructural protein NSP1 that differ in sequence and size from other Betacoronavirus NSP1 (~9 kDa for Alphacoronavirus and ~20 kDa for Betacoronavirus). In contrast, Gamma and Deltacoronaviruses lack NSP1 moiety (Jaimes et al., 2020) (“Virus Taxonomy – 1st Edition,” n.d.). NSP1 protein plays a crucial role in inhibits host gene expression, and its location is essential for virus virulence (Shen et al., 2019). [Table 1] shows different genomic structures for human coronavirus genera (Fehr and Perlman, 2015; Jansson, 2013; Payne, 2017; Snijder et al., 2016).

Scientists explored seven Coronavirus strains able to infect humans (HCoV-229E, HCoV-HKU1, HCoV-NL63, HCoV-OC43, SARS-CoV, MERS-CoV, and SARS-CoV-2) (Arabi et al., 2020; Gaunt et al., 2010; Z.-W. Ye et al., 2020). While HCoV-229E and HCoV-NL63 belong to alphacoronaviruses family and use different host proteins as a target receptor (HCoV-229E binds with host aminopeptidase N (hAPN) as a receptor, and HCoV-NL63 binds angiotensin-converting enzyme II (ACE 2) in addition to host hAPN as a receptor), the other five beta coronaviruses bind to different host receptors (HCoV-OC43 and HCoV-HKU1 use 9-O-acetylsialic acids as a receptor, SARS-CoV and SARS-CoV-2 bind ACE2 receptors, and MERS-CoV uses Dipeptidyl peptidase-4 (DPP4) receptor (Li et al., 2019; Z. W. Ye et al., 2020).

Clinically, HCoV-229E, HCoV-NL63, HCoV-OC43, and HCoV-HKU1 cause common cold-like symptoms (e.g., sneezing, fever, dry cough, sore throat, Dyspnea, Myalgia, and diarrhea), which may be progressed to pneumonia in cardiopulmonary and immunocompromised patients. However, SARS-CoV, MERS-CoV, and SARS-CoV-2 are highly pathogenic viruses and can cause respiratory system failure in immunocompromised or cardiopulmonary patients (Seah and Agrawal, 2020). [Table 2] represents the classification of different human coronaviruses with their animal reservoir, intermediate host, cellular receptors, and their main sign and symptoms.

3. Common genomic features and characterizations of coronaviruses

Coronaviruses (CoVs) genome represents the second largest genome of all RNA viruses. It is a single-stranded, positive-sense RNA, with genome size ranging from 26 Kb to 32 Kb in length and has a cap and poly-A at its 5’ and 3’ tail, respectively (Chan et al., 2020; Wu et al., 2020). In general, the number of ORFs in the CoVs genome ranging from 6 to 15 ORFs (Song et al., 2019). The 1a & 1ab ORFs represent the biggest gene in the coronavirus’s genome and cover almost two-thirds of its entire genome. These two genes encode different NSPs like replication-transcription complex (RTC), which responsible for synthesizing and transcrib-
ing the subgenomic RNA (sgRNA) [Gorbalenya et al., 2000; Hussain et al., 2005; Snijder et al., 2006].

Transcription regulatory sequence that mediates the transcription process is located between ORFs in the sgRNA, and serves as a template for the mRNA synthesis [Sawicki et al., 2007]. The Frame-shift mutation between ORF1a and ORF1b results in two polypeptide synthesis (pp1a and pp1ab) that will then be processed into 16 NSPs by the aid of either chymotrypsin-like proteases (3CLpro) or papain-like proteases with the main protease (Mpro) [Masters, 2006]. The remaining one-third of the coronaviruses genome is responsible for encoding at least four structural proteins like spike protein (S), nucleocapsid protein (N), envelope protein (E), and the membrane (M) proteins, besides some accessory proteins such as 3a/b, 4a/b, and Hemagglutinin-Esterase (HE) proteins [Hussain et al., 2005]. Sequence alignment of all CoVs genome illustrates an identity of 43% for structural proteins coding regions and 58% for nonstructural proteins coding regions. In comparison, the identity in the entire genome among all CoVs is about 54%. These results suggest that the structural proteins have more diversity than the other nonstructural proteins [Chen et al., 2020]. Nonstructural protein 14 (NSP-14) is a 3′ to 5′ exoribonuclease enzyme distinctive for all CoVs. Its function is related to maintaining the whole RNA genome of CoVs and proofreading the replication-transcription complex [Eckerle et al., 2010; Smith et al., 2013].

![Sequence-region of SARS-CoV-2. GenBank accession: NC_045512.2. Resources: The 2019 Novel Coronavirus Resource (Zhou et al., 2020) and National Center for Biotechnology Information (www.ncbi.nlm.nih.gov). Table 1 Different genomic structures for human coronavirus genera.](image)

| Genera          | Alphacoronavirus | Betacoronavirus | Gammacoronavirus | Deltacoronavirus |
|-----------------|------------------|-----------------|------------------|------------------|
| Spike Glycoproteins | Presented | Presented | Presented | Presented |
| ORFs            | 7–10 additional ORFs | 7–10 additional ORFs | 7–10 additional ORFs | 7–10 additional ORFs |
| Papain-like protease (PLpro) | two PLpro domains | Absent | Single PLpro domains | Absent |
| Hemagglutinin Esterase (HE) | Absent | Present | Absent | Absent |
| Envelope proteins | Present | Present | Present | Present |
| Nucleocapsid protein | Present | Present | Present | Present |
| Membrane Glycoproteins | Present | Present | Present | Present |

![Table 1](image)
Amino acid sequences of both N and E proteins in SARS-CoV-2 were found to have a ~92% sequence identity (Azeez et al., 2020).

4. Genome composition of SARS-CoV-2

SARS-CoV-2 genome consists of 15 ORFs coded into 29 proteins (Srinivasan et al., 2020); at the 5’ terminal of the genome, we have ORF1ab and ORF1a that encode 1ab and 1a Polypeptide, respectively [Fig. 1]. The 3’ terminal of the genome represents four structural (spike, envelope, matrix, and nucleocapsid) and nine accessory (3a, 3b, 6, 7a, 7b, 8b, 9a, 9b, and orf10) proteins. [Table 3] represents the coding region, Base length, amino acid length, and functions of different ORFs in SARS-CoV-2, and [Table 4] represents the coding region, amino acid length, and functions of different NSPs in the SARS-CoV-2 virus genome.

5. Origin of SARS-CoV-2 based on gene sequencing

Phylogenetic analyses of CoVs [Fig. 2] reveal that the SARS-CoV-2 is related to SARS-CoV and very closely related to both bat-SL-CoV ZC45 and bat-SL-CoV ZXC21. As previously reported, SARS-CoV and MERS-CoV are bat-originated (Cui et al., 2019); bats continue to be the prime suspected origin of SARS-CoV-2 since SARS-CoV-2 has a 96% resemblance of the whole genome identity with the bat coronavirus RaTG13 isolated from Rhinolophus affinis (Zhou et al., 2020). However, there are some important differences in genomic factors between SARS-CoV-2 and RaTG13, including the furin cleavage near the junction of the spike (S) protein subunits; S1 and S2, induced by the insertion of amino acids residues (PRRA) (Coutard et al., 2020). This insertion, which is absent in other related betacoronaviruses, may be the reason for the increased infectivity of the virus (Zhang and Holmes, 2020). Also, there is a decreased similarity between SARS-CoV-2 and RaTG13 in the receptor-binding domain (RBD) protein encoded by the (S) gene (Wrapp et al., 2020).

This divergence region within the RBD corresponds to the receptor-binding motif (RBM), which determines the receptor binding specificity. All these variabilities, along with the fact that animal species like civets and camels generally host SARS-CoV and MERS-CoV before infecting humans, suggest the existence of an intermediate host between bats and humans for SARS-CoV-2 (Cui et al., 2019; Wong et al., 2020). This suggestion opposes the early indication that SARS-CoV-2 can leap directly from bats to humans (Zhou et al., 2020). Full genome sequencing of coronaviruses isolated from Malayan pangolins (Liu et al., 2019) revealed ~90% identity to RaTG13 and ~91% to SARS-CoV-2 (Zhang and Holmes, 2020). Although the percentage identity between SARS-CoV-2 and Pangolin-CoV is lower than that between SARS-CoV-2 and RaTG13, the S1 protein in Pangolin-CoV showed high similarity to the one in SARS-CoV-2 compared to that of RaTG13. Furthermore, the RBM segment is highly conserved in both SARS-CoV-2 and RaTG13, sharing all five key residues. Unlike RaTG13, where only one of the five residues is shared (Zhang and Holmes, 2020; Zhou et al., 2020). Accordingly, the RBM introduced in SARS-CoV-2 might be resulting from a recombination event between Pangolin-CoV and RaTG13, and it is therefore unexpected.

---

Table 2
Classification of different type of human coronavirus.

| Genus          | Coronavirus Type | Animal reservoir | Intermediary host | First identified | Symptoms                                                                 | Cellular Receptor | Epidemiology     | References                  |
|----------------|------------------|------------------|-------------------|-----------------|---------------------------------------------------------------------------|-------------------|------------------|---------------------------|
| Alpha-coronavirus | HCoV-229E        | Bats             | Bats              | 1966            | Common cold, sneezing, fever, dry cough, sore throat, Myalgia, diarrhea, and pneumonia | APM receptor      | Endemic          | (Pene et al., 2003; Vijgen et al., 2005; Weiss and Leibowitz, 2011) |
|                | HCoV-NL63        | Bats             | palm civets       | 2004            | Sneezing, fever, dry cough, sore throat, Myalgia, diarrhea, fever, and convulsion | ACE2 receptor     | Endemic          | (Gaunt et al., 2010; Lim et al., 2016; Weiss and Leibowitz, 2011) |
| Beta- coronaviruses | HCoV-OC43        | Mice             | Cattle            | 1967            | Common cold, sneezing, fever, dry cough, sore throat, Myalgia diarrhea, and pneumonia | 9-O-acetylsialic acids as receptor | Endemic          | (Li et al., 2019; Lim et al., 2016; Weiss and Leibowitz, 2011) |
|                | HCoV-HKU1        | Mice             | Mice              | 2005            | Common cold, sneezing, fever, dry cough, sore throat, Myalgia, fever, and convulsion | 9-O-acetylsialic acids as receptor | Endemic          | (Gaunt et al., 2010; Li et al., 2019; Lim et al., 2016; Seah and Agrawal, 2020; Weiss and Leibowitz, 2011) |
|                | SARS-CoV         | Bats             | civet cat         | 2003            | Common respiratory symptoms, fever, cough, shortness of breath, breathing difficulties, and pneumonia | ACE2              | Epidemic         | (Lee and Teanor, 2016; Lim et al., 2016; Weiss and Leibowitz, 2011) |
|                | MERS-CoV         | Bats             | Dromedary camels  | 2012            | Cough, shortness of breath, and, on occasion, pneumonia, Gastrointestinal symptoms (diarrhea). | Dipeptidyl peptidase-4 (DPP4) receptor | Epidemic         | (Weiss and Leibowitz, 2011; Z.-W. Ye et al., 2020) |
|                | SARS-CoV-2       | Bats             | Bats              | 2019            | Common respiratory symptoms, fever, cough, shortness of breath, and breathing Myalgia, Dyspnea, Vomiting pneumonia, respiratory system failure and even death | ACE2              | Pandemic         | (Z.-W. Ye et al., 2020) |
that RaTG13 has obtained the same mutations found in the genome of Pangolin-CoV by random chance. The low nucleic acid identity between Pangolin-CoV and SARS-CoV-2 compared to the high nucleic acid identity between Pangolin-CoV and SARS-like coronaviruses. Also, it was discovered that the SARS-CoV-2 helicase amino acid sequence is similar to bat and SARS-like coronaviruses. Enabled genetic drift to take place as the time passes (Wong et al., 2020). Moreover, the sequence of SARS-CoV-2 has evolved into two types: L and S. Type L is derived from type S, and it is suggested that it is more aggressive and contagious than type S. However, the frequency of L type was high at the beginning of the outbreak in Wuhan, China but has decreased later (Tang et al., 2020).

6. Mutations in SARS-CoV-2 genome

Mutations drive viral evolution and genome variability, facilitating virus escape from different immune systems and promoting antivirus drug resistance. In general, RNA viruses are characterized by a high mutation rate, which can sometimes reach up to a million times than their hosts (Pachetti et al., 2020). For the SARS-CoV-2 virus, more than 10 thousand single nucleotide polymorphism (SNP) have been recognized and reported globally, with expected mutation rates ranging between (0.0001–0.01) substitutions per nucleotide site per cell infection (Cotten et al., 2014; Peck and Lauring, 2018). These mutations are distributed between different genomic structure of the virus. e.g., mutations in Spike protein, nucleocapsid, ORF (7b, 8, 9b, and 14), and RdRP (Badua et al., 2020). Among discovered mutations, two of them located in spike protein have been reported to have a higher rate of transmissibility; D614G and N501Y (Discovered in UK) (Daniloski et al., 2020; Shi et al., 2020). The African variant ‘501.V2’ is characterized by N501Y, E484K, and K417N mutations in the S protein – so it shares the N501Y mutation with the UK variant but the other two mutations are not found in the UK variant– but the impact of this variant is still under investigation (Tegally et al., 2020).
| NSP       | Coding region (aa) | Amino acid length (aa) | Function                                                                                                           | Reference                                      |
|-----------|--------------------|------------------------|--------------------------------------------------------------------------------------------------------------------|------------------------------------------------|
| NSP1 (leader protein) | 1–180             | 180                    | mRNA degradation, interferon (IFN) signaling inhibition                                                           | (Huang et al., 2011; Tanaka et al., 2012)     |
| NSP2      | 181–818            | 638                    | still unknown                                                                                                     | (Gadlage et al., 2008; Graham et al., 2005)   |
| NSP3      | 819–2763           | 1945                   | ADRP, cleaving Polypeptide, blocking the innate immunity of the host, enhancing expression of cytokines             | (Lei et al., 2018; Serrano et al., 2009)      |
| NSP4      | 2764–3263          | 500                    | still unknown                                                                                                     | (Beachboard et al., 2015; Gadlage et al., 2010; Lei et al., 2018) |
| NSP5 (3C-like proteinase) | 3264–3569      | 306                    | Mpro, 3CLpro, polypeptides cleaving, and IFN signaling inhibiting                                                 | (Beachboard et al., 2015; Stobart et al., 2013) |
| NSP6      | 3570–3859          | 290                    | formation of DMV, and restricting autophagosome expansion                                                          | (Angelini et al., 2013; Cotten et al., 2014)   |
| NSP7      | 3860–3942          | 83                     | cofactor with NSP8 and NSP12                                                                                       | (Kirchdoerfer and Ward, 2019; Zhai et al., 2005) |
| NSP8      | 3943–4140          | 198                    | primase, cofactor with NSP7 and NSP12                                                                             | (Kirchdoerfer and Ward, 2019; Te Velthuis et al., 2012) |
| NSP9      | 4141–4253          | 113                    | RNA binding, dimerization                                                                                        | (Egloff et al., 2004; Zeng et al., 2018)       |
| NSP10     | 4254–4392          | 139                    | Scaffold protein for NSP14 and NSP16                                                                             | (Bouvet et al., 2014; Fang et al., 2008; Ma et al., 2015) |
| NSP11     | 4393–4405          | 13                     | still unknown                                                                                                     | (Bouvet et al., 2014; Fang et al., 2008)       |
| NSP12     | 4393–4932          | 932                    | Primer dependent RdRp                                                                                                | (Ahn et al., 2012; Kirchdoerfer and Ward, 2019; te Velthuis et al., 2009) |
| NSP13     | 4932–5234          | 601                    | RNA helicase, and 5′ triphosphatase                                                                               | (Adedeji and Lazarus, 2016; Jia et al., 2019)  |
| NSP14     | 5235–5925          | 601                    | 3′-to-5′ triphosphatase                                                                                           | (Bouvet et al., 2012; Eckerle et al., 2010; Minskaia et al., 2006) |
| NSP15     | 5926–6452          | 527                    | 3′-to-5′ exonuclease                                                                                                | (Bouvet et al., 2012; Eckerle et al., 2010; Minskaia et al., 2006) |
| NSP16     | 6453–6798          | 346                    | dsRNA sensors evasion, and NendoU                                                                                  | (Bhardwaj et al., 2006; Zhang et al., 2018)    |
| NSP16     | 6799–7096          | 298                    | OMT, negatively regulates the innate immunity                                                                    | (Chen et al., 2011; Decroly et al., 2011; Shi et al., 2019) |

ADRDP: adenosine diphosphate-ribose 1′-phosphatase; 3CLpro: 3C-like cysteine proteinase; RdRp: RNA-dependent RNA polymerase; NendoU: nidoviral endoribonuclease specific for U; OMT: S-adenosylmethionine-dependent ribose 2′-O-methyltransferase.
to find a quick and sensitive diagnostic tool as it is crucial for COVID-19 patients increased worldwide, there was an urgent need regarding morbidity is scary (Simmonds, 2020). As the number of technique translation Genome sequence similarity among beta-coronaviruses.

A.M. Al-Qaaneh, T. Alshammari, R. Aldahhan et al. Saudi Journal of Biological Sciences 28 (2021) 1978–1989

7. Mechanism of SARS-CoV-2 genome replication and translation

Since the SARS-CoV-2 genome has the 5’ methylated cap (where the nonstructural proteins are placed) and 3’ polyadenylated side (where structural proteins reside), this allows direct translation after infection without needing an intermediate transcription stage. Also, the SARS-CoV-2 genome includes multiple ORFs that can be transcribed by several transcription regulating sequences (TRS). The viral life cycle begins when the virion enters the human respiratory tract and depend on the viral and epithelial cell membrane interaction. The infection initiates at the cell surface when the S1 subunit recognizes angiotensin-converting enzyme 2 (ACE2) as a targeted receptor or recognize exopeptidases as a key receptor for entry to the host cell. The entry mechanism depends on cellular proteases such as cathepsins and transmembrane protease serine 2 (TMPRSS2), and human airway trypsin protease (HAT) (Hoffmann et al., 2020). The S1 subunit consists of two subdomains that can operate as receptor binding domains proteins (RBD); N-terminal domain (NTD) as a mediated sugar-binding and a C-terminal domain (CTD) as receptors recognition. (Xia et al., 2020). Mutated RBD of the S1 subunit is needed for cross-species of SARS-CoV-2 transmitting. Once the binding occurs, conformational changes in the S protein, like pH reducing and/or S protein proteolysis, promote splitting of S1 from S2. Consequently, membrane fusion is placed by the S2 region (Han et al., 2020). S2 subunit has two heptad repeat (HR) domains, (HR1) and (HR2), that interact together to form a six-helix bundle (6-HB) fusion core to approximate the viral and cellular interaction (Xia et al., 2020). The replicase/transcriptase genes encoded by ORF1a/1ab translate two polyproteins (pp1a and pp1ab) that are cleaved into individual 16 nonstructural proteins (nsp1 to 16) by the action of chymotrypsin-like cysteine proteases enzyme (3CLpro, nsP3); that control viral replication and it is crucial for SARS-CoV-2 life cycle and papain-like proteases enzyme (PLpro, nsP5 proteases) which has a crucial role in suppressing immune system by deubiquitylating the host cell proteins (Morse et al., 2020; Tahir ul Qamar et al., 2020). RNA-dependent RNA polymerase (RdRp) that encoded by NSP12 plays a central role in the synthesis of a new genome molecule (RNA) that interact with each other. The other method is the two-steps assay, where the reaction occurs in two different tubes. This assay method is more precise than the one-step assay, yet it also consumes a longer time and needs the optimization of extra parameters (Bustin, 2002; Wong and Medrano, 2005). Detection of SARS-CoV-2 through reverse transcription and amplification are combined into one reaction. This method can yield fast and reproducible results; however, it is challenging because it requires the optimization of both the reverse transcription and amplification, so they do not compete with each other. The other method is the two-steps assay, where the reaction occurs in two different tubes. This assay method is more precise than the one-step assay, yet it also consumes a longer time and needs the optimization of extra parameters (Bustin, 2002; Wong and Medrano, 2005). Detection of SARS-CoV-2 through rRT-PCR assays requires collecting respiratory samples. The upper respiratory tract specimens, which involve nasopharyngeal swabs, oropharyngeal swabs, and nasal aspirates, are recommended for the initial diagnostic testing. However, the lower respiratory tract specimens, which involve sputum, Bronchoalveolar Lavage Fluid (BALF), and tracheal aspirates, are recommended when the patient is having a productive cough (“Interim Guidelines for Clinical Specimens for COVID-19 | CDC,” n.d.). Yang et al. have reported that sputum, along with nasal swabs, are more accurate for SARS-CoV-2 diagnosis.

8. Diagnosis of SARS-CoV-2 based on molecular genetic technique

There is no doubt that the presented outcome of the pandemic regarding morbidity is scary (Simmonds, 2020). As the number of COVID-19 patients increased worldwide, there was an urgent need to find a quick and sensitive diagnostic tool as it is crucial for controlling the outbreak of SARS-CoV-2 in the community. World Health Organization (WHO) has recommended molecular testing methods like polymerase chain reaction (PCR) of respiratory tract samples for the identification and laboratory confirmation of COVID-19 cases (World Health Organization, 2020). By comparison with syndromic testing and the computed tomography (CT), molecular techniques are more suitable because of their ability to target and identify particular infectious agents. Developing the molecular techniques requires recognizing the pathogen genomics and proteomics composition and understanding gene expression changes within the host during and post-infection (Udugama et al., 2020). Real-time reverse transcription-polymerase chain reaction (rRT-PCR), which is the most used diagnostic method for COVID-19 (“Real-time RT-PCR Primers and Probes for COVID-19 | CDC,” n.d.), allows the genetic detection of SARS-CoV-2. Thus, various kits have been designed to reverse-transcribe the viral RNA genome to complementary DNA (cDNA) and amplify specific regions of the cDNA (Freeman et al., 1999; Kageyama et al., 2003). Corman et al. have found three regions of the SARA-CoV-2 genome with conserved sequences, the RNA-dependent RNA polymerase gene in the ORF1ab region, the E and the N gene. Analytically, both RdRp and E genes provided high detection sensitivity, while the N gene showed poorer sensitivity (Corman et al., 2020). [Table 6] represents some probes and primers used for SARS-CoV-2 detection via real-time PCR tests. Besides designing the reaction primers and probes, other factors must be optimized to achieve a proper and accurate RT-PCR assay, including reagent conditions, incubation time, and temperature (Wong and Medrano, 2005). Also, selecting the appropriate controls is essential to ensure the results’ reliability (Yan et al., 2020). In general, these controls can help identify the detrimental factors such as contamination and ensure pathogen recognition in the addressed samples. Thus, allowing for the detection of any false-negative or false-positive amplification. There are two main methods to perform RT-PCR, the one-step assay format where reverse transcription and amplification are combined into one reaction. This method can yield fast and reproducible results; however, it is challenging because it requires the optimization of both the reverse transcription and amplification, so they do not compete with each other. The other method is the two-steps assay, where the reaction occurs in two different tubes. This assay method is more precise than the one-step assay, yet it also consumes a longer time and needs the optimization of extra parameters (Bustin, 2002; Wong and Medrano, 2005). Detection of SARS-CoV-2 through rRT-PCR assays requires collecting respiratory samples. The upper respiratory tract specimens, which involve nasopharyngeal swabs, oropharyngeal swabs, and nasal aspirates, are recommended for the initial diagnostic testing. However, the lower respiratory tract specimens, which involve sputum, Bronchoalveolar Lavage Fluid (BALF), and tracheal aspirates, are recommended when the patient is having a productive cough (“Interim Guidelines for Clinical Specimens for COVID-19 | CDC,” n.d.). Yang et al. have reported that sputum, along with nasal swabs, are more accurate for SARS-CoV-2 diagnosis.

Further, in severe cases, BALF is necessary for detecting the viral RNAs to diagnose and monitor the viruses (Yang et al., 2020). Respiratory samples negative results do not eliminate the possibility of having the disease and may arise as results of poorly-handled sampling techniques or small viral amount in the sampled area (Ai et al., 2020; Winichakoon et al., 2020). Like any other diagnostic tool, RT-PCR for SARS-CoV-2 diagnosis has some drawbacks, such as the shortage of the kits with the increased demand, its inability to detect prior infection in recovered patients who showed no symptoms of the disease, and requiring advanced equipment that are often not available in resource-limited regions. However, other alternative nucleic acid tests may overcome these limitations (X. Li...
et al., 2020; Udugama et al., 2020), like the Loop-mediated isothermal amplification test (LAMP), which is used for DNAs and RNAs amplification. It contains five primers designed for aiming at the 5’ region of the open reading frame 1a gene and Gene N in the SARS-CoV-2 RNA. Each gene has been checked with both synthetic DNA substrates and RNA transcripts. There are different ways to detect amplifying DNA either by turbidity (induce by-products of the reaction), fluorescence (a fluorescent dye binds to double-stranded DNA), or color (pH-sensitive dye), and observed either on colorimetric detection or agarose gel electrophoresis. However, when comparing the two genes, it showed that the synthetic DNA substrates are slower in amplification and recognition than the RNA pattern, confirming that RNA is proficiently transformed to cDNA through the reverse transcriptase and amplified by the DNA-dependent DNA polymerase. The result was completely comparable with the real-time detection. Also, cell lysate has been tested, and the detection sensitivity was as the synthetic RNA alone (Zhang and Holmes, 2020). Loop-mediated isothermal amplification (LAMP) assay is rapid, conducted at a single PCR, and does not require high-priced reagents or devices (Udugama et al., 2020). There are disadvantages to LAMP, such as the challenge reaction condition, designing primers, and the difficulty with designing the readout device. The RT-LAMP assay has been improved using a quenching probe (QProbe) to reveal signs, which has a good performance as an RT-PCR assay (Ozma et al., 2020). On the other hand, the microarray test has been used broadly in detecting SARS-COV-2, in which the Viral RNA will produce categorized cDNA through the reverse transcriptase and amplified by the DNA-dependent DNA polymerase. The result was completely comparable with the real-time detection. Also, cell lysate has been tested, and the detection sensitivity was as the synthetic RNA alone (Zhang and Holmes, 2020).

## 9. Conclusion

SARS-CoV-2 via utilizing molecular techniques and based on its genetic material, represents a new addition to the betacoronavirus family. It is genetically related, with more than 96% identity, to bat-SARS-like coronavirus. Although it is highly conserved, new strains and mutations appeared in different parts of the world, revealing the virus’s capability to adapt different environmental conditions to survive. As of today, no curable treatment or prophylactic vaccine available for this virus, which makes the success in resolving this pandemic depends on the people awareness of adopting proper safety measures to stop viral dissemination and the availability of new diagnostic tools able to detect the virus on its early stage in more efficient-fast way. As of today, no curable treatment or prophylactic vaccine available for this virus, which makes the success in resolving this pandemic depends on the people awareness of adopting proper safety measures to stop viral dissemination and the availability of new diagnostic tools able to detect the virus on its early stage in more efficient-fast way.

## Funding

This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.
Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

The authors thank the Dean of the Institute for Research and Medical Consultations (IRMC), Imam Abdulrahman Bin Faisal University, Dammam, Saudi Arabia for her continuous support and encouragement. The senior author JFB thanks Dr. Sayed AbduI-Azeez for his support and critical review.

References

Adejei, A.O., Lazarus, H., 2016. Biochemical characterization of middle east respiratory syndrome coronavirus helicase. mSphere. 1 https://doi.org/10.1128/mSphere.00235-16.

Ahmed, S.F., Quadeer, A.A., McKay, M.R., 2020. Preliminary identification of potential vaccine targets for the COVID-19 Coronavirus (SARS-CoV-2) Based on SARS-CoV Immunological Studies. Viruses. 12 https://doi.org/10.3390/v12030054.

Akhtar, M., Ahmadi, J.K., Taylor, D.R., Oh, J.W., 2012. Biochemical characterization of a recombinant SARS coronavirus nsp12 RNA-dependent RNA polymerase capable of copying viral RNA templates. Arch. Virol. 157, 2095–2104. https://doi.org/10.1007/s00705-012-1404-x.

Abu-Raddad, L.J., Yang, Z., Hou, H., Zhan, C., Chen, C., Lv, W., Tao, Q., Sun, Z., Xia, L., 2020. Correlation of chest CT and RT-PCR testing in coronavirus disease 2019 (COVID-19) in China: a report of 1014 cases. Radiology. https://doi.org/10.1148/radiol.2020200642.

Angelini, M.M., Atia, M.J., Neuman, B.W., Prado, L., 2013. Severe acute respiratory syndrome-coronavirus nonstructural proteins 3, 4, and 6 induce double-membrane vesicles. MBio 4. https://doi.org/10.1128/mBio.00235-13.

Arabi, Y.M., Murthy, S., Webb, S., 2020. COVID-19: a novel coronavirus and a novel challenge for critical care. Intensive Care Med. https://doi.org/10.1007/s00134-020-05955-1.

Aree, J., Murthy, S., Webb, S., 2020. COVID-19: a novel coronavirus and a novel challenge for critical care. Intensive Care Med. https://doi.org/10.1007/s00134-020-05955-1.

Aree, J., Murthy, S., Webb, S., 2020. COVID-19: a novel coronavirus and a novel challenge for critical care. Intensive Care Med. https://doi.org/10.1007/s00134-020-05955-1.

Aree, J., Murthy, S., Webb, S., 2020. COVID-19: a novel coronavirus and a novel challenge for critical care. Intensive Care Med. https://doi.org/10.1007/s00134-020-05955-1.

Aree, J., Murthy, S., Webb, S., 2020. COVID-19: a novel coronavirus and a novel challenge for critical care. Intensive Care Med. https://doi.org/10.1007/s00134-020-05955-1.

Aree, J., Murthy, S., Webb, S., 2020. COVID-19: a novel coronavirus and a novel challenge for critical care. Intensive Care Med. https://doi.org/10.1007/s00134-020-05955-1.

Aree, J., Murthy, S., Webb, S., 2020. COVID-19: a novel coronavirus and a novel challenge for critical care. Intensive Care Med. https://doi.org/10.1007/s00134-020-05955-1.

Aree, J., Murthy, S., Webb, S., 2020. COVID-19: a novel coronavirus and a novel challenge for critical care. Intensive Care Med. https://doi.org/10.1007/s00134-020-05955-1.
Phan, T., 2020. Genetic diversity and evolution of SARS-CoV-2. Infect. Genet. Evol. 81, https://doi.org/10.1016/j.meegid.2020.104260 104260.

Pillay, T.S., Pillay, T.S., 2020. Gene of the month: The 2019-nCoV/SARS-CoV-2 novel coronavirus spike glycoprotein. J. Clin. Pathol. 73, 366–369. https://doi.org/10.1136/jcp-2020-206658.

Real-time RT-PCR Primers and Probes for COVID-19 | CDC [WWW Document], n.d. URL https://www.cdc.gov/ covid19/clinical-diagnosis/primers.html (accessed 12.25.20).

Sawicki, S.G., Sawicki, D.L., Siddell, S.G., 2007. A contemporary view of coronavirus transcription. J. Virol. 81, 20–29. https://doi.org/10.1128/jvi.01358-06.

Schulze, S.R., Pehlke, R., 2009. SARS coronavirus gene expression and function. In: Molecular Biology of the SARS Coronavirus. Springer, Berlin Heidelberg, pp. 153–166. https://doi.org/10.1007/978-3-642-03683-5_10.

Seah, I., Agrawal, R., 2020. Can the Coronavirus Disease 2019 (COVID-19) Affect the Eyes? A Review of Coronaviruses and Ocular Implications in Humans and Animals. Ocul. Immunol. 28, 391–395. https://doi.org/10.1080/09258646.2020.1738501.

Serrano, F., Johnson, M.A., Chatterjee, A., Neuman, B.W., Joseph, J.S., Buchmeier, M.J., Kahn, P., Wirth, K., 2009. Nuclear magnetic resonance structure of the nucleic acid-binding domain of severe acute respiratory syndrome coronavirus nonstructural protein 3. J. Virol. 83, 12998–13008. https://doi.org/10.1128/jvi.01253-09.

Shen, Z., Wang, G., Yang, Y., Shi, J., Fang, L., Li, F., Xiao, S., Fu, Z.P., Peng, G., 2019. A conserved region of nonstructural protein 1 from alphacoronaviruses inhibits host gene expression and is critical for viral virulence. J. Biol. Chem. 294, 13606–13618. https://doi.org/10.1074/jbc.R119.007173.

Shen, Y., Zhang, M., Cao, J., Zhou, P., Zhang, J., Yang, D., Guo, L., Zhang, G., Ji, I., Hu, X., Chen, M., Gao, Z., Wang, L., Ren, L., Li, M., 2020. Genomic diversity of severe acute respiratory syndrome coronavirus 2 in patients with coronavirus disease 2019. Clin. Infect. Dis. 71, 713–720. https://doi.org/10.1093/cid/ciaa1116.

Shereen, M.A., Khan, S., Kazmi, A., Bashir, N., Siddiquee, R., 2020. COVID-19 infection: Origin, transmission, and characteristics of human coronaviruses. J. Adv. Res. 24, 91–98. https://doi.org/10.1016/j.jare.2020.03.005.

Shi, P., Su, Y., Li, R., Liu, P., Yang, Z., Dong, S., Huang, J., 2019. PDEV nsP16 negatively regulates innate immunity to promote viral proliferation. Virus Res. 266, 57–66. https://doi.org/10.1016/j.virusres.2019.03.005.

Shi, R., Ma, W., Wu, Q., Zhang, B., Song, Y., Guo, Q., Xiao, W., Wang, Y., Zheng, W., 2003. Design and application of 60mer oligonucleotide microarray in SARS coronavirus detection. Chin. Sci. Bull. 48, 1165–1169. https://doi.org/10.1007/s11434-003-0398-7.

Simon, F., 2020. Pervasive RNA secondary structure in the genomes of SARS-CoV-2 and other coronaviruses. MBio 11. https://doi.org/10.1128/mBio.01661-20.

Singh Tomar, P.P., Arkin, I.T., 2020. SARS-CoV-2 E protein is a potential ion channel that can be inhibited by Gliclazide and Memantine. Biochem. Biophys. Res. Commun. 530, 10–14. https://doi.org/10.1016/j.bbrc.2020.05.206.

Smith, E.C., Blanc, H., Vignuzzi, M., Denison, M.R., 2013. Coronaviruses lacking exoribonuclease activity are susceptible to lethal mutagenesis: evidence for that can be inhibited by Gliclazide and Memantine. Biochem. Biophys. Res. Commun. 527, 618–623. https://doi.org/10.1016/j.bbrc.2020.04.136.

Sotaboro, C.C., Sexton, N.R., Munjal, H., Lu, X., Molland, K.L., Tomar, S., Mesecar, A.D., Denison, M.R., 2013. Chimeric exchange of coronavirus nsp5 Proteases (3Clpro) identifies common and divergent regulatory determinants of protease activity. J. Virol. 87, 2936–2946. https://doi.org/10.1128/jvi.02397-12.

Taher ul Qamar, M., Alqahtani, S.M., Alamri, M.A., Chen, L.-L., 2020. Structural basis of SARS-CoV-2 3CLpro and anti-COVID-19 drug discovery from medicinal plants. J. Pharm. Anal. https://doi.org/10.1016/j.jpha.2020.03.009.

Tan, Y.-J., Lim, S.C., Hong, W., 2005. Characterization of viral proteins encoded by the coronavirus-2 (SARS CoV-2 or n-COV19), the Cause of COVID-19. JCM. 00297-20.

Yan, Y., Chang, L., Wang, L., 2020. Laboratory testing of SARS-CoV-MERS-CoV, and SARS-CoV-2 (2019-nCoV): Current status, challenges, and countermeasures. Rev. Med. Virol. 30, 2020.02.11.20021493.

Ye, Z.-W., Yuan, S., Yuen, K.-S., Fung, S.-Y., Chan, C.-P., Jin, D.-Y., 2020a. Zoonotic origins of human coronaviruses. Int. J. Biol. Sci. 16, 1686–1697. https://doi.org/10.1177/1741073319868137.

Zeng, Z., Deng, F., Shi, K., Ye, G., Wang, G., Fang, L., Xiao, S., Fu, Z., Peng, G., 2018. Determinization of coronavirus nsp9 with diverse modes enhances its nucleic acid binding affinity. J. Virol. 92, e00962–e00718. https://doi.org/10.1128/jvi.00962-19.

Zeng, Z., Dierich, S., Mau, M., Yang, Y., Guo, Q., Wang, L., Ren, L., Li, M., 2020. Biochemical characterization of SARS-CoV-2 nucleocapsid protein. Biochem. Biophys. Res. Commun. 527, 618–623. https://doi.org/10.1016/j.bbrc.2020.05.206.

Zhao, D., Yang, Y., Shi, J., Fang, L., Gao, Z., Wang, L., Ren, L., Li, M., 2020. Genomic diversity of severe acute respiratory syndrome coronavirus 2 in patients with coronavirus disease 2019. Clin. Infect. Dis. 71, 713–720. https://doi.org/10.1093/cid/ciaa1116.
Zhang, L., Li, L., Yan, L., Ming, Z., Jia, Z., Lou, Z., Rao, Z., 2018. Structural and biochemical characterization of endoribonuclease Nsp15 encoded by middle east respiratory syndrome coronavirus. J. Virol. 92, e00893–e00918. https://doi.org/10.1128/JVI.00893-18.

Zhang, X.Y., Huang, H.J., Zhuang, D.L., Nasser, M.I., Yang, M.H., Zhu, P., Zhao, M.Y., 2020. Biological, clinical and epidemiological features of COVID-19, SARS and MERS and AutoDock simulation of ACE2. Infect. Dis. Poverty. https://doi.org/10.1186/s40249-020-00691-6.

Zhang, Y.-Z., Holmes, E.C., 2020. A genomic perspective on the origin and emergence of SARS-CoV-2. Cell 181, 223–227. https://doi.org/10.1016/j.cell.2020.03.035.

Zhou, P., Yang, X.-L., Wang, X.-G., Hu, B., Zhang, L., Zhang, W., Si, H.-R., Zhu, Y., Li, B., Huang, C.-L., Chen, H.-D., Chen, J., Luo, Y., Guo, H., Jiang, R.-D., Liu, M.-Q., Chen, Y., Shen, X.-R., Wang, X., Zheng, X.-S., Zhao, K., Chen, Q.-J., Deng, F., Liu, L.-L., Yan, B., Zhan, F.-X., Wang, Y.-Y., Xiao, G.-F., Shi, Z.-L., 2020. A pneumonia outbreak associated with a new coronavirus of probable bat origin. Nature 579, 270–273. https://doi.org/10.1038/s41586-020-2012-7.