Since January 2020 Elsevier has created a COVID-19 resource centre with free information in English and Mandarin on the novel coronavirus COVID-19. The COVID-19 resource centre is hosted on Elsevier Connect, the company's public news and information website.

Elsevier hereby grants permission to make all its COVID-19-related research that is available on the COVID-19 resource centre - including this research content - immediately available in PubMed Central and other publicly funded repositories, such as the WHO COVID database with rights for unrestricted research re-use and analyses in any form or by any means with acknowledgement of the original source. These permissions are granted for free by Elsevier for as long as the COVID-19 resource centre remains active.
1. Introduction

1.1 Background

Viruses of the family Coronaviridae possess a single-stranded, positive-sense RNA genome ranging from 26 to 32 kbps in sequence length [1]. Coronaviruses have been identified in several avian hosts, as well as various mammals, including camels, bats, porcupine, mice, dogs, and cats [2].

The coronaviruses are organisms with only RNA, often referred to as the genetic RNA. The genetic RNA of the coronaviruses is self-replicating and is able to produce its own replica using the RNA-dependent RNA polymerase (RdRP) synthesis [3]. The coronavirus functions directly as a messenger RNA, which in conjunction with the ribosomal apparatus of its host is able to direct the synthesis of both the RNA polymerase enzyme required for RNA replication and the viral coat [4]. Thus with the mediation of the RdRP and based on the standard base-pairing principle, the viral RNA serves as a template in the synthesis of a complementary RNA chain and hence the synthesis of a double-stranded structure [5], with which the novel mammalian coronaviruses are now regularly and easily identified.

Molecular genetics has made it possible for the easy identification of the viruses and their attributes; for example, a severe acute respiratory syndrome (SARS)-related coronavirus of bat origin was found to be responsible for a fatal acute diarrhea syndrome in pigs [6]. Among the several coronaviruses that are pathogenic to humans, many are associated with mild clinical symptoms, with two notable exceptions, namely, the SARS coronavirus (SARS-CoV), a novel betacoronavirus that emerged in Guangdong, Southern China, in November 2002 [7] and resulted in more than 8000 human infections and 774 deaths in 37 countries during 2002–03, and the Middle East respiratory syndrome coronavirus (MERS-CoV), which was first detected in Saudi Arabia in 2012 and was responsible...
for 2,494 laboratory-confirmed cases of infection and 858 fatalities in September, 2012, including 38 deaths following a single introduction into South Korea [8].

In late December, 2019, patients diagnosed with viral pneumonia due to an unidentified microbial agent were reported in Wuhan, China [9]. A novel coronavirus was subsequently identified as the causative pathogen, provisionally named 2019 novel coronavirus (2019-nCoV). As of May 16, 2020, the number of confirmed cases had reached an alarming 4,642,506 with a global death rate of 308,866 cases of 2019-nCoV infections globally, most of which involved people living in or visiting endemic countries and thus increasing the human-to-human transmission [9].

In addition, 2019-nCoV has now been reported in almost 212 countries of the world including Italy, Malaysia, Belgium, Australia, Nigeria, Portugal, Germany, the United States, France, Brazil, Egypt, South Africa, Canada, the United Kingdom, Russia, Israel, Turkey, Iran, Vietnam, China, and Spain [10]. Infections of medical workers and family clusters have also been reported and human-to-human transmission has been confirmed in most of these endemic countries. Most of the infected patients show symptoms of high fever, dry cough, tiredness, chills, chest pain, shortness of breath and loss of taste and smell, with chest radiographs revealing invasive lesion in both lungs [11].

There has been reports of epidemiologic data of positively infected inpatients from many hospitals in different endemic countries, diagnosed with viral pneumonia of unidentified cause and from which bronchoalveolar lavage fluid samples were obtained, virus isolated, cultured and sequenced using next generation sequencing protocols and results revealed the pathogenic organism as 2019-nCoV [12].

The description of the genomic characterization of this novel virus will provide important information on the origins and cell binding processes of the virus, thus helping to chart a therapeutic measure for the viral disease [13–15].

It is therefore necessary for humanity to know the essential features of the common humanity enemy called coronavirus disease 2019 (COVID-19) coronavirus. A good knowledge and understanding of the attributes/characteristics of the coronaviruses is desired to help researchers and trademedics to understand the common enemy and develop approaches and designs aimed at a specific target of the coronaviruses, with a view to bring a lasting solution to the pandemic.

1.2 Rationale

This chapter is on the essentials of coronaviruses; however, it was also borne out of the desire to find answers to questions raised globally concerning the various palliative measures that have been spelled out as peddled rumors toward the effective prevention and control of the pandemic virus. Some of the peddled rumors include (1) the coronavirus is thermophobic and will not thrive on tropical climates; (2) the coronavirus is acidic in nature and thus the use of alkaline water will be effective; (3) the virus cannot survive for long periods outside it host cells; (4) the virus is of large size (in micrometers), so it is not airborne and hence the use of gas masks; (5) the virus is genetically unstable
and easily denatures at temperatures above 26–27°C; and (6) the virus is hydrophilic and thrives under low relative humidity.

Hence, the desire to use nucleotide sequences isolated from coronaviruses from selected endemic countries including Nigeria to ascertain the essential biological, physicochemical, and phylogenetic characteristics of the virus using in silico and bioinformatics approaches to provide answers to the questions and also to determine whether the conclusion that coronaviruses are thermally and genetically unstable in the tropics is factual or unfounded.

2. Materials and methods

To ascertain the essentials of the COVID-19 coronavirus, nucleotides sequences of the SARS-CoV-2 (which causes COVID-19) S gene were retrieved from the database of the human genome hoisted by the National Center for Biotechnology Information (NCBI): the COVID-19 gene sequences were retrieved using the FASTA format with the basic local alignment search tool engine. The accession numbers for the SARS-CoV-2 S gene sequences from different selected endemic regions were recorded after retrieval, tabulated, and used for the study. The partial coding sequences (cds) of SARS-CoV-2 were retrieved and used for the study because some of the endemic countries as at the date of retrieval of sequences from the NCBI database were yet to deposit the sequences of the whole-genome sequence of the polyprotein coronavirus, thereby limiting the usage of the complete sequence for the study. It was also prompted by the fact that Nigeria as at the time also deposited the partial sequences hence our choice of the partial cds RdRP region for the study. More so, the whole-genome sequences from the isolates were not used for the preliminary in silico studies because of the anticipated large volumes of introns (noncoding regions of the gene). Thus this has made it possible to identify the essential features and attributes of the COVID-19 virus, which are documented in this chapter.

2.1 Retrieval of nucleotides and amino acid sequences

DNA sequences of the SARS-CoV-2 Spike protein gene were retrieved from the database of the human genome hoisted by the NCBI using the FASTA format of the basic alignment search tool engine. The accession numbers for the SARS-CoV-2 S gene sequences from different endemic locations were recorded after retrieval, tabulated, and used for the study. The accession numbers for the retrieved sequences include MT0088022.1, MT042777.1, MT050414.1, MT066157.1, MT127116.1, MT159778.1, MT187977.1, MN938385.1, LC522350.1, and MN970003.1.

2.2 Determination of physiochemical properties of the novel coronavirus gene sequences

The proteomic ProtParam options of the Expasy (Expert Protein Analysis System) online interactive software were used to determine the physiochemical properties of the novel
SARS-CoV-2 S gene sequences using the retrieved nucleotide and translated amino acid sequence of the different gene sequences. The proteomic parameter (protparam) analytical tool of expasy.org was used to determine the molecular weight, atomic composition, total negatively charged and positively charged amino acids residues, theoretical isoelectric points (pI), extinction coefficients, instability index and grand hydropathicity of the novel SARS-CoV-2 gene sequences [16].

2.3 Determination of guanine-cytosine content of SARS-CoV-2 sequences from 12 endemic countries

The online interactive program GENSCAN was used to determine the percentage of guanine and cytosine, the start codon, the end codon, coding regions, and the number of translated amino acids in the sequences of the SARS-CoV-2 among selected endemic countries.

2.4 Determination of evolutionary distance, mutation pathway, time of gene divergence, and phylogenetic analysis of the novel SARS-CoV-2 gene sequences

The molecular evolution and genetic analysis (MEGA X.0) software was used to determine the evolutionary relationship existing between the SARS-CoV-2 gene sequences obtained from the different endemic countries, deposited at the NCBI database. The retrieved sequences were aligned using the CLUSTALW option of multiple sequence alignments in MEGA X.0.

The phylogenetic tree was constructed using unweighted pair group mean average (UPGMA) option at 1000 bootstrap of the aligned amino acid sequence of the novel SARS-CoV-2 gene sequences. The time of divergence of the novel SARS-CoV-2 gene sequences was also inferred. The MEGA X software was used to align the sequences and subject them to phylogenetic analysis for substitution model selection, evolutionary distance estimation, phylogeny inference, mutation rate and pattern estimation, testing natural selection, and ancestral sequence inference for the selected sequences.

2.5 Prediction of secondary and tertiary protein folding RNA structure of coronavirus sequences

The secondary protein folding structures of the FASTA formatted coronavirus sequences were predicted using the online interactive program NSOPMA in lieu of GORIV, for the determination of dominance of the folded secondary structures by alpha helices, beta turns, extended strands, and random coils. Another online interactive program, the Phyre and Phyre (protein homology Y recognition engine), which is based on the canonical amino acid sequence obtained from the NCBI database [17,18], was used in lieu of I-TASSER for the determination of the tertiary protein folding structures of the coronavirus sequences.
3. Revealed essential features of COVID-19 coronavirus

3.1 Essential physical attributes of COVID-19 coronavirus

3.1.1 Molecular weight/size of COVID-19 coronavirus

Coronaviruses are single-stranded RNA viruses with genomic molecular sizes ranging from 26,000 to 32,000 bps in length. The genome of the viruses has approximately 35 functional proteins/genes within the genome. Each gene occupying approximately between 200 and 3000 bps. Most of the sequences in the genome are introns implying the noncoding region of the genome [18–20].

The preliminary in silico evaluation for the determination of physicochemical characteristics of the retrieved partial cds nucleotide and translated protein sequences isolated from the RdRP region of the novel coronavirus strains from 12 selected endemic regions of the world showed that the number of translated amino acids from the retrieved nucleotide sequence using the MEGA X.0 translation option varied from 51 KDa (Kilodalton) in LC522350.1 to 227 KDA in MT172668.1 (Table 1.1). Three nucleotides make up an amino acid codon. The molecular weight of the SARS-CoV-2 RdRP region gene sequences ranged between 5597.31 g/mol in LC522350.1 isolates and 25,734.22 g/mol in MT172668.1 isolates [21]. The molecular weights varied among the isolates (Table 1.1), and thus they exhibit differences in characteristics.

The coronavirus isolates also showed variability in the number of amino acids contained in their molecules. The highest number of amino acids was 227 (MT172668.1) and the least was 51 (LC522350.1).

| S/N | Accession no. | No. of amino acids | Molecular weight | Total no. of atoms | Total no. of negatively charged residues | Total no. of positively charged residues |
|-----|---------------|--------------------|------------------|--------------------|------------------------------------------|----------------------------------------|
| 1   | MT0088022.1   | 99                 | 10998.77         | 156.1              | 8                                        | 12                                     |
| 2   | MT042777.1    | 87                 | 9643.92          | 1328               | 7                                        | 9                                      |
| 3   | MT050414.1    | 167                | 18918.91         | 2636               | 14                                       | 17                                     |
| 4   | MT066157.1    | 82                 | 9654.11          | 1323               | 6                                        | 9                                      |
| 5   | MT172668.1    | 227                | 25734.22         | 3573               | 5                                        | 23                                     |
| 6   | MT152900.1    | 92                 | 10858.55         | 1530               | 2                                        | 21                                     |
| 7   | MT127116.1    | 140                | 15806.59         | 2240               | 9                                        | 13                                     |
| 8   | MT159778.1    | 145                | 16289.82         | 2266               | 11                                       | 14                                     |
| 9   | MT187977.1    | 131                | 14999.90         | 2083               | 10                                       | 18                                     |
| 10  | MN938385.1    | 81                 | 8905.49          | 1259               | 1                                        | 11                                     |
| 11  | LC522350.1    | 51                 | 5597.51          | 782                | 0                                        | 8                                      |
| 12  | MN970003.1    | 82                 | 9655.11          | 1323               | 6                                        | 9                                      |
3.1.2 Total number of atoms of the COVID-19 coronavirus

The total number of atoms possessed by each of the SARS-CoV-2 isolates determines their strength and stability. The more atoms found in an isolate, the more stable the isolate, as more heat and other compounds are needed and required to destroy such isolates. The total number of atoms varies directly with the molecular weight and the number of amino acids in the isolates. Isolates with higher molecular weight are those with higher number of atoms, as shown in Table 1.1.

It was gathered from the study that the total number of atoms from the isolates showed that the least was from LC522350.1 isolates with 782 atoms and the highest was from MT172668.1 isolates with 3573 atoms (Table 1.1) [21,22].

3.1.3 Amino acid side chain constituents of the COVID-19 coronavirus

The amino acid side chain constituent of the protein coat of the viruses plays a very significant role in the reactivity of the antigen with the antibody of the host immune system. The amino acid side chain is the fulcrum upon which the immunochemistry of drugs, haptens, and antibodies revolves. The heavy and light amino acid side chains take part in drug delivery reactions to target sites. The design of drugs using the ligand-based approach is highly dependent on the heavy and light amino acid side chains of the reacting species. Table 1.2 presents the results of dominant side chain amino acids for coronaviruses and their percentage bioavailability for reaction.

3.1.4 Aliphatic (side chain) index of the COVID-19 coronavirus

Aliphatic index of the viral protein sequence is the relative volume occupied by the aliphatic side chains including alanine, valine, isoleucine, and leucine amino acids. The measure of the aliphatic index of a protein is a positive factor toward increasing the

| S/N | Accession no. | Alanine (%) | Valine (%) | Isoleucine (%) | Leucine (%) | Aliphatic index |
|-----|---------------|-------------|------------|----------------|-------------|----------------|
| 1   | MT0088022.1   | 6.1         | 5.1        | 5.1            | 17.2        | 102.37         |
| 2   | MT042777.1    | 8.0         | 6.9        | 1.1            | 11.5        | 77.36          |
| 3   | MT050414.1    | 5.4         | 7.2        | 3.0            | 14.4        | 93.93          |
| 4   | MT066157.1    | 4.9         | 6.1        | 2.4            | 9.8         | 70.12          |
| 5   | MT172668.1    | 3.5         | 6.2        | 5.3            | 11.5        | 86.70          |
| 6   | MT152900.1    | 7.6         | 5.4        | 2.2            | 9.8         | 70.00          |
| 7   | MT127116.1    | 6.4         | 8.6        | 5.0            | 15.7        | 112.07         |
| 8   | MT159778.1    | 6.2         | 7.6        | 2.8            | 14.5        | 95.45          |
| 9   | MT187977.1    | 7.6         | 2.3        | 4.6            | 5.3         | 52.98          |
| 10  | MN938385.1    | 6.2         | 4.9        | 6.2            | 9.9         | 83.09          |
| 11  | LC522350.1    | 7.8         | 3.9        | 5.9            | 3.9         | 57.54          |
| 12  | MN970003.1    | 4.9         | 6.1        | 2.4            | 9.8         | 70.12          |
thermostability of the globular proteins of the viruses. Globular protein stability varied among the coronavirus sequences and ranged between 52.98 in MT187977.1 and 112.07 in MT127116.1. Globular protein stability decreases with increasing index volume occupied by the aliphatic side chain. As the volume occupied by the aliphatic side chain consisting primarily of the amino acids alanine, valine, isoleucine, and leucine increases, the thermostability of the virus decreases, as the amino acids are synthesized with time. Hence, the results in Table 1.2 show that the RdRP region sequence from MT187977.1 was more genetically stable while that from MT127116.1 was the least unstable [21,23].

### 3.1.5 Instability index of the COVID-19 coronavirus

The instability index is a characteristic that determines the genetic and thermal stability of the sequences. An RNA sequence is said to be thermally and genetically stable if its instability index is below 40 and is scientifically considered thermally and genetically unstable if its instability index is 40 and above. Hence, from the results, it implies that only the sequences from MT187977.1 were thermally and genetically stable, while all other 11 sequences were unstable [21,24]. The results as presented in Table 1.3 revealed that instability index among the sequences ranged from 39.58 in MT187977.1 to 73.65 in MT152900.1. With widespread unfavorable instability index among the evaluated coronaviruses, it implies that they are easily denatured by heat (high temperatures) and other elements of weather. This finding thus confirms the assertion that coronaviruses do not thrive in tropical climates with high temperatures and high humidity.

### 3.1.6 Guanine—cytosine content of the COVID-19 coronavirus

The guanine—cytosine (G—C) content of a gene sequence is a measure of the genetic and thermal stability of the gene sequences. The G—C linkage in the double-stranded

| S/N | Accession no. | Instability index | G—C content (%) | Estimated half-life |
|-----|---------------|-------------------|-----------------|--------------------|
| 1   | MT0088022.1   | 62.26             | 45.39           | 1.9                |
| 2   | MT042777.1    | 46.24             | 38.78           | 4.4                |
| 3   | MT050414.1    | 51.63             | 38.08           | 2.0                |
| 4   | MT066157.1    | 40.51             | 38.62           | 1.4                |
| 5   | MT172668.1    | 54.39             | 37.53           | 1.9                |
| 6   | MT152900.1    | 73.65             | 48.48           | 1.9                |
| 7   | MT127116.1    | 46.37             | 38.34           | 1.3                |
| 8   | MT159778.1    | 49.67             | 38.51           | 1.2                |
| 9   | MT187977.1    | 39.58             | 49.26           | 7.2                |
| 10  | MN938385.1    | 49.18             | 37.28           | 1.0                |
| 11  | LC522350.1    | 40.60             | 38.46           | 1.0                |
| 12  | MN970003.1    | 40.51             | 38.62           | 1.4                |
molecules shows that guanine and cytosine (G=C) are amino acids that are held by a triple hydrogen bond in the double-stranded RNA molecule, while the adenine and uracil (thymine) amino acids are held together by a double hydrogen bond. The triple hydrogen bonds, which are covalently bonded, require more thermal (heat) energy input for dissociation compared with the thermal energy required for dissociation of the double bond (A=T). Hence a nucleotide sequence with more guanine and cytosine will have more triple bonds and thus will be genetically and thermally stable [25,26]. However, based on the G=C content of a gene sequence, a gene is said to be thermally and genetically stable if its G=C content is 49% and above and below which 0% – 47% it is said to be unstable. In Table 1.3, the G–C content of the RNA sequence ranges between 37.28% in MN938385.1 and 49.26% in MT187977.1, indicating that only the sequence from the MT187977.1 isolate was genetically and thermally stable, whereas others were unstable and thus the isolates will be easily prone to gene mutations [21,27].

3.1.7 Half-life of the COVID-19 coronavirus in human reticulocytes

The half-life is a prediction of the time it takes for half of the amount of protein in a cell to disappear after its synthesis in the cell. It relies on the N-terminal amino acid (side chain) or N-end rule, which relates the half-life of a protein to the identity of its N-terminal amino acid residue. Thus the N-terminal amino acids originated from the observation that the identity of the N-terminal residue (aliphatic side chain) of a protein plays an important role in determining the protein stability in vivo [28]. The half-life of the isolates evaluated in Table 1.3 ranged from 1.0 h in MN938385.1 and LC522350.1 isolates to 7.2 h in the MT187977.1 isolate. Thermal and genetic stability of isolates is associated with the N-terminal amino acid, which is a measure of the isolate’s stability. Stability and mutation rate of the isolate sequences increase with increase in half-life. The MT187977.1 isolate showed the highest half-life of 7.2 h and hence its stability [21,29].

3.2 Essential chemical features of the COVID-19 coronavirus

3.2.1 Grand hydropathicity of the COVID-19 coronavirus

The hydropathicity character of an isolate sequence shows the free energy of transfer $\Delta G$ (Kcal/mol) from solvent to solute interface. It involves the transfer of unfolding protein chains from water to the bilipid layer interface and the transfer of unfolded chains to organic solvent [30]. This characteristic explains the use of alcohol-based sanitizers for the prevention of coronaviruses.

The grand hydropathicity, which is an expression of the hydrophobic and hydrophilic character of the viral protein coat, was determined for the various SARS-CoV 2 virus isolates, and the results presented in Table 1.4 revealed that 8 of the 12 isolates expressed hydrophobic character, while only 4 expressed hydrophilic character. Hydrophobic characteristics of the isolates varied from $-0.053$ in MT008022.1 isolate to $-0.777$ in MT187977.1 and MT152900.1 isolates. Table 1.5 also revealed that hydrophilic characteristics of the isolates ranged from 0.05 in MT050414.1 isolate to 0.344 measured in MT152900.1 isolates [21,31].
The grand hydropathicity value compares the most abundant protein in the extracellular matrix. Hydropathicity illustrates the therapeutic use of water as determined by the hydrophobicity scale [32]. This also shows the extent to which handwashing with water as a sanitizing hygiene is effective for the prevention of virus spread. Hydrophobicity scales are also used to predict the preservation of the genetic code in the viruses [33]. The result is in tandem with that of the report by Ziebuhr et al. [34],

### Table 1.4  Nucleotide base pairs, extinction coefficient, theoretic isoelectric point, aliphatic index, and GRAND hydropathicity of the RdRP region of coronavirus sequences from some endemic countries.

| S/N | Accession no.  | Nucleotide base pairs | Extinction coefficient | Theoretic isoelectric point (pl) | Grand hydropathicity |
|-----|----------------|-----------------------|-----------------------|---------------------------------|---------------------|
| 1   | MT0088022.1    | 322                   | 12740                 | 9.04                            | −0.053              |
| 2   | MT042777.1     | 294                   | 9190                  | 8.37                            | −0.255              |
| 3   | MT050414.1     | 562                   | 18170                 | 8.28                            | 0.050               |
| 4   | MT066157.1     | 250                   | 17795                 | 8.57                            | −0.106              |
| 5   | MT172668.1     | 810                   | 37815                 | 9.26                            | 0.181               |
| 6   | MT152900.1     | 322                   | 37595                 | 11.95                           | −0.777              |
| 7   | MT127116.1     | 459                   | 20315                 | 8.74                            | 0.344               |
| 8   | MT159778.1     | 483                   | 15065                 | 8.33                            | 0.106               |
| 9   | MT187977.1     | 406                   | 54430                 | 9.83                            | −0.777              |
| 10  | MN938385.1     | 287                   | 12740                 | 10.63                           | −0.042              |
| 11  | LC522350.1     | 182                   | 12615                 | 10.57                           | −0.292              |
| 12  | MN970003.1     | 290                   | 17795                 | 8.57                            | −0.106              |

**RdRP**, RNA-dependent RNA polymerase.

### Table 1.5  Nucleotide base pairs, start codon, end codon, and coding exon region of the RdRP region of coronavirus sequences from some endemic countries.

| S/N | Accession no.  | RdRP length (bps) | Start codon | End codon | Coding region |
|-----|----------------|-------------------|-------------|-----------|---------------|
| 1   | MT0088022.1    | 322               | NA          | NA        | NA            |
| 2   | MT042777.1     | 294               | 9           | 227       | 9–227         |
| 3   | MT050414.1     | 562               | 11          | 424       | 11–424        |
| 4   | MT066157.1     | 250               | NA          | NA        | NA            |
| 5   | MT172668.1     | 810               | 224         | 712       | 9–324         |
| 6   | MT152900.1     | 322               | 17          | 287       | 40–309        |
| 7   | MT127116.1     | 459               | 69          | 441       | 16–234        |
| 8   | MT159778.1     | 483               | 22          | 414       | 9–275         |
| 9   | MT187977.1     | 406               | 19          | 282       | 50–173        |
| 10  | MN938385.1     | 287               | NA          | NA        | NA            |
| 11  | LC522350.1     | 182               | NA          | NA        | NA            |
| 12  | MN970003.1     | 290               | NA          | NA        | NA            |

**RdRP**, RNA-dependent RNA polymerase.

The grand hydropathicity value compares the most abundant protein in the extracellular matrix. Hydropathicity illustrates the therapeutic use of water as determined by the hydrophobicity scale [32]. This also shows the extent to which handwashing with water as a sanitizing hygiene is effective for the prevention of virus spread. Hydrophobicity scales are also used to predict the preservation of the genetic code in the viruses [33]. The result is in tandem with that of the report by Ziebuhr et al. [34],
who measured the free energy of an unfolding protein chain from isolates and found that increase in genetic and thermal stability of the protein structure is directly proportional to an increase in hydrophobicity of the protein up to a certain size limit [35]. Hence the conclusion that the protein structure stability is measured by its hydropathicity character.

3.2.2 Theoretic isoelectric point (pl) of the COVID-19 coronavirus

The theoretic isoelectric point (pl) represents the pH of a solution at which the net charge of the isolates protein coat becomes zero [36]. In solutions in which the actual pH is above the theoretic isoelectric point, the surface of the isolate protein would predominantly be negatively charged, and therefore like-charged molecules would exhibit or show repulsive forces because they carry no net electric charge and are electrically neutral [37–39]. The theoretic isoelectric point is based on the primary protein structure, which is the primary sequence that is unlikely to match the actual pH due to some charged side chains forming a salt bridge [40]. The theoretic isoelectric point value affects the solubility of the protein molecule at this given pH by conferring minimum solubility in water and salt due to precipitation of the salt in solution. This accounts for the use of alcohol-based sanitizers in dissolving the protein coat of the novel coronavirus. Thus a specific theoretic pl for a target protein sequence can be used to model the process from which compounds used for the purification of the protein can be developed [25]. The theoretic isoelectric point (pl) varies from 8.33 in isolates from MT159778.1 to 11.95 in isolates obtained from MT152900.1 (Table 1.4) [21,41].

3.2.3 Extinction/attenuation coefficients of the COVID-19 coronavirus

The extinction coefficient is a characteristic that determines how strongly a species absorbs or reflects radiation or light at a particular wavelength. It is an intrinsic property of the isolates that is dependent on the atomic, chemical, and protein structural composition of the isolate sequences [42]. Hence, when high ultraviolet light is incident on an object containing the coronavirus, the extinction coefficient characteristics make it possible for the viral protein coat to be broken down completely or disintegrate to extinction [43]. This value was low for all isolates except for the MT187977.1 isolate, as shown in the results in Table 1.4.

The light or radiation absorption and reflection potentials (extinction coefficient) toward disintegration of the protein coat of the isolates vary from 9190 m/mol in MT042777.1 isolates to 54,430 m/mol in MT187977.1 isolates [21,44]. The attenuation coefficient is the ability of the coronavirus to be weakened in terms of their virulence after light absorption and genetic and thermal disintegration. This coefficient varied among the mutant strains of the coronavirus (Table 1.4). This singular characteristic of the COVID-19 coronavirus usually occasioned by the fast rate of mutation of the coronavirus has made the production and development of vaccines (attenuated form of the coronavirus) difficult, as the mutant strains possess variable attenuation coefficients with which resurgence is possible after attenuation.
3.2.4 Total number of negatively charged amino acid residues
The total number of negatively charged amino acid residues is the sum total of aspartic and glutamine amino acid in the coronavirus protein coat. This characteristic of the viral protein coat determines its reactivity with solvents. Highly negatively charged residues will thrive and react on acidic media.

The total number of negatively charged amino acid residues from the SARS-CoV-2 isolates was also determined and presented in Table 1.5. It shows that the LC522350.1 isolates did not have any negatively charged amino acid residues within the coding region, whereas isolates from MN938385 had only one negatively charged amino acid residue in the coding region. Isolates from MT050414.1 had the highest number of negatively charged amino acid residues of 14 in the coding region [21,35]. All the isolate sequences evaluated in this report contain lower total number of negatively charged aspartic and glutamine amino acid residues. The use of alkaline water here does not affect the viruses as like-charges will repel.

3.2.5 Total number of positively charged amino acid residues
The total number of positively charged amino acid residues is the sum of arginine and lysine amino acids in the viral protein coat. This characteristic of the viral protein coat determines its reactivity with solvents. Highly positively charged residues will thrive on alkaline media.

Table 1.5 further showed that the total number of positively charged amino acid residues in the coding region of the retrieved and translated sequences varied between 8 in LC522350.1 isolates to as high as 23 in MT172668.1 isolates [21,45]. All the isolate sequences evaluated contain a higher number of total positively charged arginine and lysine amino acid residues. The use of alkaline water greatly affects these classes of coronaviruses because of the reaction of the positively charged amino acid residues and alkaline water and hence the prescription on the use of alkaline water as a means of preventing the virus spread in some quarters.

3.2.6 Coding regions for COVID-19 viruses
The exons or coding regions reveal the portion of the virus genome that contains the spike protein gene that codes for the protein containing the inoculum or infective entity. Table 1.6 shows the results of preliminary in silico evaluation for the determination of the coding region of SARS-CoV-2 of the retrieved partial cds nucleotide and translated protein sequences isolated from the RdRP region of the novel coronavirus strains from 12 selected endemic countries of the world. Ubi et al. [21] had reported that the exon or coding regions were not available for MT0088022.1, MT066157.1, MN938385.1, LC522350.1, and MN970003.1. However, the coding regions range from 9 to 227 bps in MT042777.1 isolates, 11 to 424 bps in MT050414.1 isolates, 9 to 324 bps in MT172668.1 isolates, 40 to 309 bps in MT152900.1 isolates, 16 to 234 bps in MT127116.1 isolates, 9 to 275 bps in MT159778.1 isolates, and 50 to 173 bps in MT187977.1 isolates (Table 1.5).
3.3 Biological characteristics of the COVID-19 virus

3.3.1 Protein coat structure of the COVID-19 coronavirus

The protein coat structure of viruses can enable for the protein engineering of the fat that occupies the outer layer of the viral protein coat and helps determine the substrate specificity, double bond positioning (regioselectivity), and reaction outcome of the coronaviruses [46]. Information on the protein coat characteristics can also provide the potentials to improve effectiveness in existing reaction mechanisms and even to engineer enzymes that can undertake or undergo reactions at novel positions in the carbon side chains. Knowledge of the protein folding structures can open up the possibility of synthesizing novel fatty acids with multiple functional groups, thereby paving the way for designer lipids that currently lie beyond the reach of most of the available metabolic or protein engineering strategies and thus initiating a pathway for therapeutic enhancement and discovery [47].

The development and use of fatty acid desaturases are a prime target in protein engineering because of their central and well-defined role in the modification of aliphatic side chains (alanine, valine, isoleucine, and leucine) within the cell. This may further provide an insight into the structural determinants that govern the regioselectivity within the acyl chain of the fatty acid substrate that makes up the viral protein coat [47]. Knowledge of the structure of the protein of coronaviruses can be very useful in structural and ligand-based drug design and discovery.

3.3.2 Primary protein structures of the COVID-19 coronavirus

The primary protein structure is the particular sequence of amino acids found in the coronavirus protein coat and is determined by the covalent peptide bonding between the amino acids. The primary protein structures for the coronavirus are the basic nucleotide and amino acid sequences.

Table 1.6 Secondary protein folding characteristics of coronavirus sequences from some endemic countries of the world.

| S/N | Accession no. | Alpha helix (%) | Extended strand (%) | Random coil (%) | Beta turns (%) |
|-----|---------------|-----------------|---------------------|-----------------|--------------|
| 1   | MT0088022.1   | 25.25           | 17.17               | 51.52           | 6.06         |
| 2   | MT042777.1    | 56.32           | 13.76               | 20.69           | 9.20         |
| 3   | MT050414.1    | 50.90           | 13.72               | 24.55           | 10.78        |
| 4   | MT066157.1    | 31.71           | 37.80               | 23.17           | 7.32         |
| 5   | MT172668.1    | 39.21           | 25.58               | 25.55           | 9.69         |
| 6   | MT152900.1    | 38.02           | 9.78                | 41.30           | 10.87        |
| 7   | MT127116.1    | 63.57           | 10.00               | 20.00           | 6.43         |
| 8   | MT159778.1    | 49.66           | 19.31               | 22.07           | 8.79         |
| 9   | MT187977.1    | 1.53            | 22.14               | 67.94           | 8.40         |
| 10  | MT187977.1    | 51.85           | 12.38               | 30.86           | 4.94         |
| 11  | LC522350.1    | 25.49           | 29.41               | 35.29           | 9.80         |
| 12  | MN970003.1    | 31.71           | 36.59               | 34.39           | 7.32         |
3.3.3 Secondary protein folding structures of the COVID-19 coronavirus

The secondary folding structure of a protein is any regular repeating organization of the polypeptide chain. This characteristic of the coronaviruses determines to a large extent the reactivity of the viral protein coat and the thermal as well as the genetic stability of the spike protein gene of the coronaviruses. The dominant sheet of secondary structure determines the genetic strength and stability of the virus. Secondary structures dominated by alpha helices (building blocks) are more stable, closely followed by those dominated by random coils and then the extended strands. Some also show the presence of beta turns regulatory elements in their molecules [47].

Table 1.6 shows that the secondary protein folding structure of the RdRP region RNA sequence was dominated by alpha helices for the sequences in isolates MT042777.1 (56.32%), MT050414.1 (50.90%), MT172668.1 (39.21%), MT127116.1 (63.57%), MT159778.1 (49.66%), and MN938385.1 (51.88%) [21,47].

The secondary protein folding structure of the RdRP region RNA sequences was dominated by extended strand for sequences from MT066157.1 (37.80%) and MN970003.1 (36.59%), while random coil was the dominant secondary folding protein structure sheet for sequences obtained from LC522350.1 (35.29%), MT187977.1 (67.94%), MT152900.1 (41.30%), and MT008022.1 (51.52%) (Table 1.6).

3.3.4 Tertiary protein folding structures of the COVID-19 coronavirus

Tertiary protein folding structures are simply the more compact structure in which the helical and nonhelical regions of a polypeptide chain are folded back on themselves, which occurs in a specific pattern, thereby conferring certain characteristic (3D) properties on the protein. The 3D tertiary protein folding structure can be modeled and optimized to discover drugs for small molecules (ligands) [48]. With the development of the structural 3D model of the coding regions of the coronaviruses, as shown in Fig. 1.1A–H, it is very possible and important in the development of prophylactics and therapeutics for the dreaded disease to use the structure-based drug design and ligand-based drug design by adopting the computer-aided drug design (CADD, SBDD, and LBDD) approaches.

Fig. 1.1A–H shows the preliminary in silico determination of tertiary folding protein coat structure for the NCBI retrieved partial cds translated protein sequences obtained from the RdRP region of the novel coronavirus strains from selected endemic regions of the world [21,48].

3.3.5 Domain architectural composition of the COVID-19 coronavirus

Biosynthetic processes in the Spike protein gene of the novel SARS-CoV-2 in terms of RNA replication, transcription and protein synthesis take place in sites like N-glycosylation site, C-phosphorylation site, N-myristylation site and ADP ribosylation factor and amidation domain sites which have the same motifs for the sites among the mutant strains of the coronaviruses evaluated. The domain architectural characteristics of SARS-CoV-2 indicated that the viruses have four main protein types common to all the isolates,
which include the spike (S) protein, nucleocapsid (N) protein, envelope (E) protein, and the membrane (M) protein. Knowledge of the domain architecture of the coding region of the coronavirus will reveal the organization of the antibody molecule and its domain as well as the likely reaction site thereof [49]. These protein sites are the protein coat surface antigens through which the viruses infect their host with their toxins. It is important to note that although the viruses have in common the aforementioned protein bodies, the expression of the protein as the surface antigen site for the “peeling off” infection of host differs among the mutant strains of the coronavirus [50]. This

FIGURE 1.1 (A) MT008022.1, (B) MT042777.1, (C) MT050414.1, (D) MT127116.1, (E) MT159778.1, (F) MT187977.1, (G) MT187977.1, (H) LC522350.1, and (I) MN970003.1.
characteristic feature affords the coronavirus the opportunity to escape the host immune system and phagocytosis of the host antibodies and also attenuate vaccines used for any related respiratory tract ailment like SARS-CoV-1. These attributes of the coronavirus have also made it possible for the lowered efficacy of drugs, such as chloroquine, that were previously acknowledged and used for the prophylactic and therapeutic treatment of SARS-CoV-1 in 2003/04. Thus this calls for a thorough research into the dominating surface antigen protein types exhibited by the current mutant SARS-CoV-2 strains before the development and design of target drugs and vaccines.

3.3.6 Antigen epitope characteristics of the COVID-19 coronavirus

An epitope is that part of the virus pathogen that binds to the paratope of the host antibody. The paratope is the part of the host antibody that recognizes an antigen-binding site and is usually a small region (15–22 amino acids) of the host antibody Fv region, which contains parts of the antibody heavy and light chains. The antigen receptor part of the antibody for coronaviruses is the angiotensin-converting enzyme usually designated as ACE-2. At this site, the host immune system produces cytokine storms including interleukins and tumor necrotic factors that react with the viral antigen forming the epitope. Specific antibodies produced against the viral antigen offer a very unique opportunity for drug design and discovery, diagnosis, and treatment.

3.4 Phylogenetic characterization of the COVID-19 coronavirus

3.4.1 Evolutionary distance of the COVID-19 coronavirus

Preliminary in silico evaluation for the determination of developmental relationship among the coronaviruses was carried out by inferring evolutionary pathway using phylogenetic tree and evolutionary time of divergence for the retrieved partial cds nucleotide sequences isolated from the RdRP region of the novel coronavirus strains from 12 selected endemic regions (Fig. 1.2). The phylogenetic tree is a rooted tree indicating that the evaluated coronaviruses evolved from a common ancestor. The molecular evolutionary and genetic analysis (MEGA) software was used to infer evolutionary distance, mutation pathway, and phylogenetic (developmental) relationship among the retrieved RdRP region COVID-19 sequences. The reference sequence used for inferring the mutation pathway and evolutionary distance was the MN908947.3 betavirus SARS-CoV-2 obtained as reference sequences deposited in the NCBI database. The reference sequence showed a percentage homology of 99.8% with most of the predicted sequences in terms of similarity and identity. The evolutionary distance and consensus of the reference sequence was 1.6, indicating the most time of genetic divergence of the SARS-CoV-2 gene more than other evolving ones. The closest relative to the betavirus reference sequence was the partial cds RdRP region of SARS-CoV-2 isolate MT152900.1, and also the analysis obtained from MT152900.1 showed an evolutionary distance and mutation pathway of 0.9 and group consensus of 0.2 relative to the last common ancestor and reference sequence.
3.4.2 Time of genetic divergence and group consensus of the COVID-19 coronavirus

The phylogenetic tree (Fig. 1.2) revealed that the SARS-CoV-2 sequences retrieved from the NCBI database from partial cds RdRP region of SARS-CoV-2 isolates from MT127116.1, MT159778.1, MT050414.1, and MT042777.1 had the same genetic and mutation pathway with an evolutionary distance of 0.4 and divergence time of 0.8 relative to the last common ancestor and reference sequence.

3.4.3 Mutation and evolutionary pathway of the COVID-19 coronavirus

Fig. 1.2 revealed that sequences retrieved from the NCBI GenBank for partial cds RdRP region of SARS-CoV-2 isolates from MT172668.1, MN938385.1, and LC522350.1 have the same genetic and mutation pathway with an evolutionary distance of 0.6 and divergence time of 0.8 relative to the last common ancestor and reference sequence.

SARS-CoV-2 sequences obtained from MT008022.1 isolates showed an evolutionary distance of 1.0 and mutation pathway of 0.1 genetic divergence time relative to the last common ancestor. SARS-CoV-2 sequences from MT066157.1 and MN970003.1 isolates have a similar genetic and mutation pathway of 0.2 divergence time relative to the last common ancestor and with an evolutionary distance of 0.7 consensus. The evolutionary distance of SARS-CoV-2 sequences obtained from MT187977.1 isolate was 0.3 with a genetic and mutation pathway consensus of 0.1 relative to the last common ancestor (Fig. 1.2).

The study further revealed that the SARS-CoV-2 S gene sequences of the isolates have a high and fast mutation rate from the reference sequences of SARS-CoV-2 betaviruses. This might explain why the viruses change in size, shape, and structure, making it...
difficult to produce a suitable vaccine from the reference sequence [50]. This may also probably explain why the development of vaccines against most of the viral diseases has not been possible, which is due to the continuous metamorphosis and mutation of viruses in size, shape, and protein structural coat [25,50]. This singular characteristic of the COVID-19 coronavirus, that is, fast rate of mutation, has made the production and development of vaccines (attenuated form of the coronavirus) difficult, as the mutant strains possess variable attenuation coefficients with which resurgence is possible after attenuation.

3.4.4 Genetic polymorphism/single nucleotide polymorphisms

Great genetic variability was observed among the mutant strains of the coronaviruses evaluated. This is called polymorphism. The difference in single nucleotides among different isolates vertically in the same loci after multiple sequence alignment explains the variations in their characteristics. The single nucleotide polymorphisms (SNPs) are good molecular markers that can be employed to detect even the smallest variability among the coronavirus strains. The genetic polymorphism and SNPs associated with coronaviruses as seen in this report may affect drug response by influencing the action of the drug on the target molecule or the metabolism of the drugs on the coronavirus in vivo. Many SNPs have been found to correlate positively and strongly with the lack of drug efficacy. Multiple nucleotide sequence alignments revealed that the SARS-CoV-2 S gene sequences of the isolates have only three conserved regions at locus position 815 (guanine conserved region), locus position 808 (adenine conserved region), and locus position 679 (guanine conserved region). The multiple aligned sequences showed more than 100 SNPs without gaps obtained from the evaluated coronaviruses, thus indicating a high degree of genetic polymorphism among the novel pandemic coronaviruses. This characteristic of the coronavirus explains its ability to overcome the antibody reactions of the host and other haptenlike drugs. Multiple sequence alignment of the DNA sequences of the isolates also revealed a greater proportions of nucleotide base (A-T, G-C) deletions which proportionately translates to a higher rate of the S gene mutation (Fig. 1.2) observed among the isolates. The total number of mutation sites was 1218, with an average nucleotide diversity of 0.4221 for the coronavirus strains [21,25,50].

3.5 General essential features of the COVID-19 coronavirus

The whole structure of the coronavirus (Fig. 1.3A) portrays a crownlike shape normally worn by monarchs and kings especially during coronation. This probably may have influenced the name corona taken from word coronation (on a lighter note anyway). The reddish (black in print version) structures are the spike protein, while the gray part represents the surface protein coat.
3.5.1 Common symptoms of COVID-19
Usually after infection or contact with the coronavirus, it may take between 6 and 14 days for symptoms to appear. Some of the symptoms may include

i. fever,
ii. dry cough,
iii. tiredness,
iv. muscle aches and pains,
v. sore throat,
vi. chills,
vii. diarrhea,
viii. headache,
ix. loss of taste and smell,
x. skin rashes,
xi. difficulty in breathing or shortness of breath,
xii. chest pain or pressure in chest,
xiii. loss of speech.

3.5.2 Mode of spread of SARS-CoV-2 infection
The SARS-CoV-2 is a zoonotic virus usually transmitted from animal sources to humans and also by human-to-human transmission through direct contact with infected persons and by traveling of infected persons to endemic countries or areas where the viruses prevail [50].

3.5.3 Diagnosis of SARS-CoV-2 infection
Antibody and antigen detection is crucial for the differential diagnosis of many different pathologic conditions. The SARS-CoV-2 infection can be diagnosed and detected using any of the following methods, which include but not limited to molecular and immune response approaches:

i. nucleic acid amplification test (NAAT),
ii. rapid antigen direct test (RADT),
iii. direct fluorescent antibody test (DFAT).
These methods involve the reaction between the antigen of the virus and the antibodies produced by the host immune system to produce a color reaction for positive and negative indications.

The samples or specimens used for detection, diagnosis, and confirmatory test in persons infected with coronavirus may include but not limited to

i. upper respiratory tract swab,
ii. lower respiratory tract swab,
iii. saliva fluid,
iv. nasopharyngeal swab.

3.5.4 Prevailing prophylactic measures
The use of personal protective equipment (PPE) has been the most widely adopted prophylactic measure for COVID-19 because of the absence of authenticated vaccines for the infection. Although most pharmaceutical and biotechnological companies have developed vaccines in vitro (that is, testing the vaccines in cell cultures in the laboratory using experimental models), some have initiated clinical trials where the developed vaccines are directly administered to patients. Some of the PPE used include

1. the nose mask that prevents the viral propagules from spreading to other people because they are highly contagious and communicable.

Other habits/practices that seek to afford some perceived measure of prevention against the coronavirus are

2. social distancing 1.5 m apart for all and sundry in endemic and nonendemic areas and avoiding social gatherings,
3. regular washing of hands with soap,
4. using hand sanitizers,
5. coughing and sneezing into curved elbows or handkerchiefs,
6. eating fresh vegetables and immune-boosting foods.

3.5.5 Therapeutics/vaccine trials available at the moment
So far, no effective therapy has been developed for COVID-19. However, some of the drugs that have been tried and prescribed for the treatment of COVID-19 include

i. chloroquine.
ii. hydroxychloroquine,
iii. ivermectrin,
iv. remdesivir.
3.5.6 Challenges in combating COVID-19 pandemic

There are several challenges globally in combating the COVID-19 pandemic. Some of them include the following:

1. The fast rate of mutation of the COVID-19 virus strains acquiring new virulent forms and status among mutant strains, which enables the mutant strain to overcome the host immune system.

2. The variations in the mutant strains result in the production of different surface antigens with characteristic epitopes. Some strains possess the nucleocapsid protein antigens, some possess the spike protein antigens, some show the envelope protein antigens, and some others show the membrane protein surface antigen, thus making it difficult for a specific drug to preterminate the transcription processes of the coronavirus.

3. Chloroquine and hydroxychloroquine drugs were able to stop the biosynthesis and processing of glycoprotein through the process of glycosylation by the spike protein virus antigen. But the mutant strains possessing other surface antigen characteristics like nucleocapsid, envelope, or membrane protein will attenuate the inhibitory effect of chloroquine and hydroxychloroquine on the virulence of the strains.

4. Natural plants (herbs) have been reported and claimed in some quarters to show significant potential (concoction) in combating the dreaded pandemic coronavirus infection. Notably among some of these herbs, which are yet to be scientifically screened by regulatory government agencies such as the Food and Drug Administration (FDA), the National Agency for Food and Drug Administration (NAFDAC), etc., include (1) *Artemisia annua* (wormwood), (2) *Euphorbia hirta* (asthma weed), and (3) *Ageratum conyzoides* (goat weed). However, plants (herbs) that have shown to have promising or potential cure for COVID-19 are supposed to undergo scientific screening procedure as follows:

   i. Phytochemical screening of the bioactive components of the plants using GC-MS/HPLC (gas chromatography-mass spectrometry/high-performance liquid chromatography) to determine the volatile and nonvolatile components of the plants. This is to enable the laboratory synthesis of the most effective bioactive component in others to meet the growing demand in areas where the plant is scarce and is not found at all.

   ii. Validation of efficacy of the bioactive components on in vitro cell cultures of the coronavirus using different concentrations to enable for dosage determination.

   iii. Carrying out an in vivo clinical trial of the most efficacious herb on infected patients or experimental models before releasing it to the public.
3.5.7 Any hope for a lasting solution and the future?
There is absolutely high hopes for a lasting solution to the coronavirus pandemic. Several pharmaceutical and biotechnological companies, universities, research institutes, corporate bodies, and individuals are all seriously engaged currently with the development of vaccines and drugs for the treatment of COVID-19. Many had developed vaccines and drugs that are currently undergoing clinical trials. Thus it is hoped that after the series of clinical trials ongoing globally, a routine vaccine and drug will emerge that will eventually provide for the prevention and cure of the coronavirus disease. More so, with the knowledge that the coronaviruses mutate faster and acquire new mutant virulent status exhibiting different forms of surface protein antigens, it is possible to develop specific diagnostic tests for infected patients to ascertain the type of surface antigen of mutant coronavirus strains to enable for proper prescription of specific therapy or vaccines specific to the antibody/antigen epitope detected.

3.6 Conclusion
This chapter presents the essential characteristics of the COVID-19 coronavirus in terms of its physical, chemical, and biological attributes, as well as the general features, that make the coronavirus mutant strains causing the current global pandemic to be deadly and a threat to humanity. It also analyzes some of the efforts and strategies adopted by humanity to combat the pandemic.

Preliminary in silico determination of genetic and thermal stability potentials of the isolates has been revealed using the instability index, aliphatic index, G–C content, hydropathicity, and half-life of the isolates in human reticulocytes in vitro. This also aligns with the WHO initiatives and stands on the adverse effect of tropical climates on the pathogenicity and virulence of the novel coronavirus strains. Thus it is true that soap (foam) can dissolve the fatty layer. Heat (high temperature) above 26°C can destroy the virus, as well as hot water. Alkaline water or a basic solution with high pH can destroy the virus. Alcohols can also dissolve the fat, especially the outer lipid layer. High ultraviolet light (extinction coefficient) exposure on any object harboring the virus will break down the viral protein coat.

The scary characteristics of the coronavirus lie in its ability to mutate quickly producing many mutant copies of the coronavirus that are not exact, thus conferring on the mutant strains the ability to escape the host immune system. This is responsible for the resurgence of the viruses with varied characteristics and antigens, which differ from the previous strains, thus giving room for the risk of a pandemic.

Moreover, efforts geared toward the development of suitable vaccines for the SARS-CoV-2 and other related viral diseases have not been successful because of the variable attenuation coefficients of the coronavirus mutant strains, the continuous and
fast rate of the S gene mutations, and the isolates acquiring new size, shape, and structure over time. This thus calls for a concerted effort in studying the characteristics and mutation rates of the virus in order to be able to predict the future mutation rate and attributes of the present strains, with a view to find a suitable therapy for the pandemic and for biosecurity of humans against the virus in the future.

Abbreviations

**SARS-CoV** Severe acute respiratory syndrome coronavirus  
**Expasy** Expert protein analysis system  
**kDa** Kilodalton unit of protein  
**PHYRE** Protein homology Y recognition engine  
**FDA** Food and Drug Administration  
**NAFDAC** National Agency for Food and Drug Administration  
**GC-MS** Gas chromatography-mass spectrophotometry  
**HPLC** High-power liquid chromatography  
**kbp** Kilo base pair  
**bp** Base pair  
**SARS** Sever acute respiratory syndrome  
**RdRP** RNA-dependent RNA polymerase  
**CoV** Coronavirus  
**NCBI** National Center for Biotechnology Information  
**WHO** World Health Organization  
**UN** United Nations  
**MEGA** Molecular evolutionary genetic analysis  
**G—C** Guanine-cytosine  
**ΔG** Gibbs free energy change  
**EC** Extinction coefficient

References

[1] C.A. de Haan, P.S. Masters, X. Shen, S. Weiss, P.J. Rottier, The group-specific murine coronavirus genes are not essential, but their deletion, by reverse genetics, is attenuating in the natural host, Virology 296 (1) (2002) 177–189.

[2] K. Anand, J. Ziebuhr, P. Wadhwani, J.R. Mesters, R. Hilgenfeld, Coronavirus main proteinase (3CLpro) structure: basis for design of anti-SARS drugs, Science 300 (5626) (2003) 1763–1767.

[3] R.S. Baric, K. Fu, W. Chen, B. Yount, High recombination and mutation rates in mouse hepatitis virus suggest that coronaviruses may be potentially important emerging viruses, Adv. Exp. Med. Biol. 380 (1995) 571–576.

[4] J.C. de Jong, E.C. Claas, A.D. Osterhaus, R.G. Webster, W.L. Lim, A pandemic warning? Nature 389 (1997) 544.

[5] A. Ikai, Thermostability and Aliphatic index globular proteins, J. Biochem. 88 (6) (1980) 1895–1898.

[6] A.G. Bost, E. Prentice, M.R. Denison, Mouse hepatitis virus replicase protein complexes are translocated to sites of M protein accumulation in the ERGIC at late times of infection, Virology 285 (1) (2001) 21–29.
A.G. Bost, R.H. Carnahan, X.T. Lu, M.R. Denison, Four proteins processed from the replicase gene polyprotein of mouse hepatitis virus colocalize in the cell periphery and adjacent to sites of virion assembly, J. Virol. 74 (7) (2000) 3379–3387.

K.A. Callow, H.F. Parry, M. Sergeant, D.A. Tyrrell, The time course of the immune response to experimental coronavirus infection of man, Epidemiol. Infect. 105 (1990) 435–446.

WHO, Cumulative Number of Reported Probable Cases of Severe Acute Respiratory Syndrome (SARS), 2019. Available from: www.who.int/csr/sars/country/2019_05_20/en/.

R. Casais, V. Thiel, S.G. Siddell, D. Cavanagh, P. Britton, Reverse genetics system for the avian coronavirus infectious bronchitis virus, J. Virol. 75 (24) (2001) 12359–12369.

CDC, Update: outbreak of severe acute respiratory syndrome-worldwide, 2003, Morb. Mortal. Wkly. Rep. 52 (2003) 241–248.

W. Chen, R.S. Baric, Molecular anatomy of mouse hepatitis virus persistence: coevolution of increased host cell resistance and virus virulence, J. Virol. 70 (6) (1996) 3947–3960.

W. Chen, R.S. Baric, Evolution and persistence mechanisms of mouse hepatitis virus, Adv. Exp. Med. Biol. 380 (1995) 63–71.

S.S. Chim, S.K. Tsui, K.C. Chan, T.C. Au, E.C. Hung, Y.K. Tong, R.W. Chiu, E.K. Ng, P.K. Chan, C.M. Chu, J.J. Sung, J.S. Tam, K.P. Fung, M.M. Waye, C.Y. Lee, K.Y. Yuen, Y.M. Lo, Genomic characterization of the severe acute respiratory syndrome coronavirus of Amoy Gardens outbreak in Hong Kong, Lancet 362 (9398) (2003) 1807–1808.

X.M. Zhang, W. Herbst, K.G. Kousoulas, J. Storz, Biological and genetic characterization of a hemagglutinating coronavirus isolated from a diarrhoeic child, J. Med. Virol. 44 (2) (1995) 152–161.

Anonymous, Severe acute respiratory syndrome (SARS), Wkly. Epidemiol. Rec. 78 (2003) 81–83.

WHO (World Health Organization), Cumulative Number of Reported Probable Cases of Severe Acute Respiratory Syndrome (SARS), 2003. Available from: http://www.who.int/csr/sarscountry/2003_07_11/en.

M.L. Ballesteros, C.M. Sanchez, L. Enjuanes, Two amino acid changes at the N-terminus of transmissible gastroenteritis coronavirus spike protein result in the loss of enteric tropism, Virology 227 (2) (1997) 378–388.

K.O. Cho, M. Hasoksuz, P.R. Nielsen, K.O. Chang, S. Lathrop, L.J. Saif, Cross-protection studies between respiratory and calf diarrhea and winter dysentery coronavirus strains in calves and Rt- Pcr and nested Pcr for their detection, Arch. Virol. 146 (12) (2001) 2401–2419.

D.A. Brian, B.G. Hogue, T.E. Kienzle, The coronavirus hemagglutinin esterase clycoprotein, in: S.G. Siddell (Ed.), The Coronavirus, Plenum Press, New York, 1995, pp. 165–179.

M. Tamura, G. Stecher, D. Peterson, A. Filipski, S. Kumar, MEGA6: molecular evolutionary genetics analysis version 6.0, Mol. Biol. Evol. 30 (2013) 2725–2729.
[26] X. Xiao, S. Chakraborti, A.S. Dimitrov, K. Gramatikoff, D.S. Dimitrov, The SARS-CoV S glycoprotein: expression and functional characterization, Biochem. Biophys. Res. Commun. 312 (4) (2003) 1159–1164.

[27] C.L. Yeager, R.A. Ashmun, R.K. Williams, C.B. Cardellicchio, L.H. Shapiro, A.T. Look, K.V. Holmes, Human aminopeptidase N is a receptor for human coronavirus 229E, Nature 357 (6377) (1992) 420–422.

[28] H. Vennema, A. Poland, K. Floyd Hawkins, N.C. Pedersen, A comparison of the genomes of FECVs and FIPVs and what they tell us about the relationships between feline coronaviruses and their evolution, Feline Pract. 23 (1995) 40–44.

[29] WHO, Case Definitions for Surveillance of Severe Acute Respiratory Syndrome (SARS), 2003 (Online) Available from: http://www.who.int/csr/sars/casedefinition/en/.

[30] B. Yount, K.M. Curtis, E.A. Fritz, L.E. Hensley, P.B. Jahrling, E. Prentice, M.R. Denison, T.W. Geisbert, R.S. Baric, Reverse genetics with a full-length infectious cDNA of severe acute respiratory syndrome coronavirus, Proc. Natl. Acad. Sci. USA 100 (22) (2003) 12995–13000.

[31] B. Yount, M.R. Denison, S.R. Weiss, R.S. Baric, Systematic assembly of a full-length infectious cDNA of mouse hepatitis virus strain A59, J. Virol. 76 (21) (2002) 11065–11078.

[32] N.S. Zhong, B.J. Zheng, Y.M. Li, Poon, Z.H. Xie, K.H. Chan, P.H. Li, S.Y. Tan, Q. Chang, J.P. Xie, X.Q. Liu, J. Xu, D.X. Li, K.Y. Yuen, Peiris, Y. Guan, Epidemiology and cause of severe acute respiratory syndrome (SARS) in Guangdong, People’s Republic of China, Lancet 362 (9393) (2003) 1353–1358.

[33] J. Ziebuhr, V. Thiel, A.E. Gorbalenya, The autocatalytic release of a putative RNA virus transcription factor from its polyprotein precursor involves two paralogous papain-like proteases that cleave the same peptide bond, J. Biol. Chem. 276 (35) (2001) 33220–33232.

[34] J. Ziebuhr, E.J. Snijder, A.E. Gorbalenya, Virus-encoded proteinases and proteolytic processing in the Nidovirales, J. Gen. Virol. 81 (Pt 4) (2000) 853–879.

[35] D.B. Tresnan, R. Levis, K.V. Holmes, Feline aminopeptidase N serves as a receptor for feline, canine, porcine, and human coronaviruses in serogroup I, J. Virol. 70 (12) (1996) 8669–8674.

[36] R.J. deGroot, M.C. Horzinek, Feline infectious peritonitis, in: S.G. Siddell (Ed.), The Coronaviridae, Plenum Press, New York, 1995, pp. 293–315.

[37] R.S. Baric, E. Sullivan, L. Hensley, B. Yount, W. Chen, Persistent infection promotes cross-species transmissibility of mouse hepatitis virus, J. Virol. 73 (1) (1999) 638–649.

[38] P.J. Bonilla, A.E. Gorbalenya, S.R. Weiss, Mouse hepatitis virus strain A59 RNA polymerase gene ORF 1a: heterogeneity among MHV strains, Virology 198 (2) (1994) 736–740.

[39] K. Tamura, M. Nei, S. Kumar, Prospects for inferring very large phylogenies by using the neighbor-joining method, Proc. Natl. Acad. Sci. USA 101 (2004) 11030–11035.

[40] N. Saitou, M. Nei, The neighbor-joining method: a new method for reconstructing phylogenetic trees, Mol. Biol. Evol. 4 (1987) 406–425.

[41] H. Tsunemitsu, L.J. Saif, Antigenic and biological comparisons of bovine coronaviruses derived from neonatal calf diarrhea and winter dysentery of adult cattle, Arch. Virol. 140 (7) (1995) 1303–1311.

[42] Y. van der Meer, E.J. Snijder, J.C. Dobbe, S. Schleich, M.R. Denison, W.J. Spaan, J.K. Locker, Localization of mouse hepatitis virus nonstructural proteins and RNA synthesis indicates a role for late endosomes in viral replication, J. Virol. 73 (9) (1999) 7641–7657.

[43] V. Campanacci, M.P. Egloff, S. Longhi, F. Ferron, C. Rancurel, A. Salomoni, C. Durousseau, F. Tocque, N. Bremond, J.C. Dobbe, E.J. Snijder, B. Canard, C. Cambillau, Structural genomics of the SARS coronavirus: cloning, expression, crystallization and preliminary crystallographic study of the Nsp9 protein, Acta Crystallogr. Sect. D Biol. Crystallogr. 59 (Pt 9) (2003) 1628–1631.
[44] R.S. Baric, B. Yount, L. Hensley, S.A. Peel, W. Chen, Episodic evolution mediates interspecies transfer of a murine coronavirus, J. Virol. 71 (3) (1997) 1946–1955.

[45] K.W. Tsang, P.L. Ho, G.C. Ooi, et al., A cluster of cases of severe acute respiratory syndrome in Hong Kong, N. Engl. J. Med. 348 (2003) 1977–1985.

[46] H. Tsunemitsu, Z.R. El-Kanawati, D.R. Smith, H.H. Reed, L.J. Saif, Isolation of coronaviruses antigenically indistinguishable from bovine coronavirus from wild ruminants with diarrhea, J. Clin. Microbiol. 33 (12) (1995) 3264–3269.

[47] K. Yokomori, M.M. Lai, Mouse hepatitis virus utilizes two carcinoembryonic antigens as alternative receptors, J. Virol. 66 (10) (1992) 6194–6199.

[48] B. Yount, K.M. Curtis, R.S. Baric, Strategy for systematic assembly of large RNA and DNA genomes: transmissible gastroenteritis virus model, J. Virol. 74 (22) (2000) 10600–10611.

[49] V.N. Chouljenko, X.Q. Lin, J. Storz, K.G. Kousoulas, A.E. Gorbalenya, Comparison of genomic and predicted amino acid sequences of respiratory and enteric bovine coronaviruses isolated from the same animal with fatal shipping pneumonia, J. Gen. Virol. 82 (12) (2001) 2927–2933.

[50] M.A. Clark, Bovine coronavirus, Br. Vet. J. 149 (1) (1993) 51–70.