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.
Role of genomics in combating COVID-19 pandemic

K.A. Saravanan, Manjit Panigrahi, Harshit Kumar, Divya Rajawat, Sonali Sonejita Nayak, Bharat Bhushan, Triveni Dutt

Division of Animal Genetics, Indian Veterinary Research Institute, Izatnagar, Bareilly 243122, UP, India
Livestock Production and Management Section, Indian Veterinary Research Institute, Izatnagar, Bareilly 243122, UP, India

ARTICLE INFO

Edited by John Doe

Keywords:
COVID-19
Genomics
Genomic surveillance
Reverse vaccinology
SARS-CoV-2

ABSTRACT

The coronavirus disease 2019 (COVID-19) quickly swept over the world, becoming one of the most devastating outbreaks in human history. Being the first pandemic in the post-genomic era, advancements in genomics contributed significantly to scientific understanding and public health response to COVID-19. Genomic technologies have been employed by researchers all over the world to better understand the biology of SARS-CoV-2 and its origin, genomic diversity, and evolution. Worldwide genomic resources have greatly aided in the investigation of the COVID-19 pandemic. The pandemic has ushered in a new era of genomic surveillance, wherein scientists are tracking the changes of the SARS-CoV-2 genome in real-time at the international and national levels. Availability of genomic and proteomic information enables the rapid development of molecular diagnostics and therapeutics. The advent of high-throughput sequencing and genome editing technologies led to the development of modern vaccines. We briefly discuss the impact of genomics in the ongoing COVID-19 pandemic in this review.

1. Introduction

Coronavirus disease 2019 (COVID-19) has wreaked havoc on the world, costing millions of lives, severely affecting public health systems, and inflicting social and economic crises. It has rapidly spread globally, becoming one of the most devastating outbreaks in the history of mankind. As of December 3, 2021, there have been more than 263 million confirmed cases of COVID-19 and over 5.2 million deaths worldwide (https://covid19.who.int/). Continuous attempts are being made to effectively tackle this deadly disease. Being the first pandemic in the post-genomic era, advancements in genomics contributed a lot to scientific understanding and the public health response to the COVID-19, to a greater degree which was not feasible during the past outbreaks like 2002–2003 severe acute respiratory syndrome (SARS) epidemic. Genomic technologies have been employed by researchers all over the world to better understand the viral origin, outbreak dynamics, transmission, and evolution. Integration of genomics and other omics technologies played a crucial role in the development of new diagnostics, therapeutics, and vaccines.

Genomics is a branch of biology that focuses on the study of structure, function, mapping, and editing of the entire genome of an organism (McKusick and Ruddle, 1987). Genomics has many sub-disciplines such as structural genomics, functional genomics, comparative genomics, epigenomics, metagenomics, pharmacogenomics, and others, which use bioinformatics and computational tools to explore the characteristics of genomes. The advent of next-generation sequencing platforms has transformed genomics from a discipline into a technology that is commonly used in labs around the world to solve scientific problems. Genomics is now widely employed in medicine, research, biotechnology, and agriculture.

Abbreviations: ACE2, Angiotensin-Converting Enzyme – 2; COVID-19, Coronavirus Disease; CRISPR, Clustered Regularly Interspaced Short Palindromic Repeats; CSIR, Council Of Scientific And Industrial Research; GISAID, Global Initiative on Sharing All Influenza Data; ICMR, Indian Council of Medical Research; ICTV, International Committee on Taxonomy of Viruses; MERS-CoV, Middle East respiratory syndrome coronavirus; NGS, Next-Generation Sequencing; NSP, Non-Structural Protein; ORF, Open Reading Frame; PANGOLIN, Phylogenetic Assignment of Named Global Outbreak Lineages; PRF, Programmed -1 Ribosomal Frameshifting; RBD, Receptor-binding domain; RdRp, RNA-dependent RNA polymerase; RTC, Replication-Transcription Complex; SARS-CoV-2, Severe Acute Respiratory Syndrome Coronavirus 2; TMEM106B, Transmembrane protein 106B; TMPRSS2, Transmembrane Protease Serine 2; UTR, Untranslated region; VOC, Variants of Concern; VOI, Variants of Interest; WHO, World Health Organization.

* Corresponding author.

E-mail address: manjit.panigrahi@icar.gov.in (M. Panigrahi).

https://doi.org/10.1016/j.gene.2022.146387

Received 3 January 2022; Received in revised form 17 February 2022; Accepted 28 February 2022

Available online 4 March 2022
0378-1119/© 2022 Elsevier B.V. All rights reserved.
In this review, we provide a brief history of the identification of SARS-CoV-2 (severe acute respiratory syndrome coronavirus 2) and its origin, outlining how genomics helps in understanding the biology of SARS-CoV-2, and discuss the importance of genomic surveillance in tracking SARS-CoV-2 variants and their spread. Finally, we highlight how genomic data is exploited in the development of molecular diagnostics, therapeutics, and vaccines to combat COVID-19 (Fig. 1).

2. Identification of causative agent and its origin

In December 2019, a cluster of cases with atypical pneumonia of unknown etiology was reported in some of the local hospitals in Wuhan city of China (Wu et al., 2020). Initial investigations identified that the pathogen was a novel coronavirus (CoV) and named 2019-nCoV by the World Health Organization (WHO). The disease has spread around the world in a very short period and crossed one hundred thousand COVID-19 cases worldwide within two months. Then, COVID-19 was declared a pandemic by WHO on March 11, 2020 (as shown in Fig. 2). During the initial stages of the outbreak, sequencing of samples from patients led to the identification of the causative organism (Wu et al., 2020; Lu et al., 2020).

Identification of the origin and source of infection is very important to take necessary public health measures to reduce disease spread. The analysis of the viral genomic sequences, from Wuhan and surrounding areas, provided insights into the early transmission dynamics and enabled the determination of the times of origin and diversification (Li et al., 2020; Boni et al., 2020). Lu et al. (2020) reported that the sequences obtained from nine patients were highly similar, with more than 99.98% sequence identity. Analysis of sequence data revealed that the virus belongs to the genus Betacoronavirus and subgenus Sarbecovirus (Wu et al., 2020). Lu et al. (2020) showed that 2019-nCoV had a sequence identity of 79% with SARS-CoV and 50% with MERS-CoV. The virus was subsequently renamed as SARS-CoV-2 (Severe Acute Respiratory Syndrome Coronavirus 2) by the International Committee on Taxonomy of Viruses (ICTV) and the disease was named as COVID-19 (Coronavirus disease 2019) by WHO (Zhou et al., 2020b).

Phylogenetic analysis of genome sequences from SARS-CoV-2 and related viruses from other animals was carried out to determine the zoonotic origin of SARS-CoV-2. These investigations revealed that SARS-CoV2 was more closely related to two bat-derived coronaviruses, bat-SL-CoVZC45 and bat-SL-CoVZXC21 with more than 88% sequence identity (Lu et al., 2020). Zhou et al. (2020b) also showed that the genomic sequences were 96% identical to a coronavirus, BatCoV RaTG13 in horseshoe bat (Rhinolophus affinis). So far, the closest known sequence to SARS-CoV-2 was BatCoV RaTG13. Several studies identified SARS-CoV-2-related coronaviruses in Chinese pangolins (Manis pentadactyla) and Malayan pangolins (Manis javanica), but pangolin coronaviruses were less closely related to SARS-CoV-2 (with 85.5–92.4% sequence similarity) than bat coronaviruses (Lam et al., 2020; Liu et al., 2020). Most of the findings suggested that the bats could be the most probable natural reservoir for SARS-CoV-2 lineage (Andersen et al., 2020; Lu et al., 2020; Zhou et al., 2020b; Wacharapluesadee et al., 2021). Pangolins are suspected to be an intermediate host of SARS-CoV-2 due to sequence similarity between pangolin coronaviruses and SARS-CoV-2 (Lam et al., 2020; Liu et al., 2020; Xiao et al., 2020). Therefore, the comparative analysis of genomic and metagenomic data from various animal sources will be crucial in unraveling the origin and evolution of SARS-CoV-2.

3. Understanding the characteristics of SARS-CoV-2

In the early stages of the pandemic, genomic and proteomic analyses have proven helpful in understanding the mechanisms of viral entry and molecular interactions with hosts which are vital to the spread of the disease. SARS-CoV-2 is an enveloped, single-stranded, positive-sense, RNA virus with a genome size of ~29.9 kb classified under the genus Betacoronavirus in the family Coronaviridae (V’kovski et al., 2020). The genome comprises 5′ UTR, replicase (ORF1a/ORF1b), four structural genes, 3′ UTR, and a poly (A) tail (Hu et al., 2020). SARS-CoV-2 genome has 14 different open reading frames (ORF) which encode 27 proteins including 4 major structural proteins (Spike (S), Envelope (E), Membrane (M), Nucleocapsid (N) proteins) (Lokman et al., 2020). Apart from these, several ORFs encoding non-structural proteins (such as papain-
1 ribosomal frameshifting (PRF) is a critical step in which the trans target cell (Shang et al., 2020). Conformational changes for the fusion and entry of the virus into the host proteases mainly transmembrane protease serine 2 (TMPRSS2), cathepsin L1 (CTSL), and furin, which make necessary by the host proteases. The spike protein is cleaved and activated to enter into the human cells, but with higher affinity than the SARS-CoV to the same receptor angiotensin-converting enzyme 2 (ACE2) as SARS-CoV to enhance or inhibiting the production of pro-inflammatory cytokines, significant role in modulating host responses to infection, such as enhancing or inhibiting the production of pro-inflammatory cytokines, and are the determinants of the pathogenicity of the virus (Shang et al., 2020). The virus binds to the host cell using surface spike glycoproteins that comprise 2 functional subunits, S1 and S2. The receptor-binding domain (RBD) of the S1 subunit recognizes and attaches to the host cell receptor, while the S2 subunit is needed for fusion with the host cell membrane (Wrapp et al., 2020). Thus the Spike proteins were mostly used as therapeutic targets to prevent the entry of the virus into the host cells (Lekto et al., 2020). SARS-CoV-2 and other related coronaviruses share genetic similarities in the spike protein RBD motif, which facilitated in identifying the cell entry receptor to which SARS-CoV-2 attaches, and hence the type of cells that it may infect. Experimental reverse genetics methods have shown that the SARS-CoV-2 uses the same receptor angiotensin-converting enzyme 2 (ACE2) as SARS-CoV to enter into the human cells, but with higher affinity than the SARS-CoV virus (Lekto et al., 2020). The spike protein is cleaved and activated by the host proteases mainly transmembrane protease serine 2 (TMPRSS2), cathepsin L1 (CTSL), and furin, which make necessary for fusion and entry of the virus into the target cell (Shang et al., 2020).

During the translation of the SARS-CoV-2 RNA genome, programmed −1 ribosomal frameshifting (PRF) is a critical step in which the translational reading frame is switched at the junction of ORF 1a and 1b (Bhatt et al., 2021). PRF is necessary for the synthesis of RNA-dependent RNA polymerase (RdRp) and downstream proteins which are crucial for virus propagation. The replication-transcription complex (RTC) and RdRp activity facilitates a more complicated replication and transcription process in coronavirus genomes than in other kinds of RNA viruses. RNA polymerase synthesizes complementary negative-strand RNAs from the positive sense template genomic RNA (gRNA). The continuous replication leads to full-length gRNAs, whereas discontinuous jumping of RdRp is called template switching which yields subgenomic RNAs (sgRNAs) with shared 5’ and 3’ ends. Next-generation sequencing (NGS) and nanopore sequencing technologies enabled the researchers to identify hundreds of template switches and to construct the subgenomic landscapes of SARS-CoV-2 (Wang et al., 2021). As a result, the molecular basis for deciphering the emergence of new strains and to track pathogen transmission and evolution (Lo and Jamrozy, 2020). Both genomic and epidemiological information should be brought together promptly to guide public health and social measures (PHSMS), diagnosis, treatment, and vaccination. Genomic epidemiology has been widely applied in various countries to track the origin and routes of transmission of COVID-19 (Deng et al., 2020; Fauver et al., 2020; Miller et al., 2020; Rockett et al., 2020; Seemann et al., 2020).

4. Genomic surveillance

The pandemic has opened a new era of genomic surveillance, wherein scientists are monitoring changes of the viral genome in real-time to understand the evolution of SARS-CoV-2 and to predict the emergence of new variants at the global and national levels (Cyranoski, 2021; Joonlasak et al., 2021). Genomic surveillance involves the use of epidemiological, genomics, and phenomics data to monitor the emergence of new strains and to track pathogen transmission and evolution. Advances in next-generation sequencing have enabled the rapid and efficient production of entire viral genomes at a low cost. Genomic sequencing plays a major role in the continuous monitoring of the evolution of SARS-CoV-2 genome. The WHO recommended the nations speed up genome sequencing and share the genomic data and findings in a coordinated way through a publicly accessible database. To coordinate sequencing operations, several initiatives and consortia have been formed in various countries (Table 1). For example, in April 2020, the COVID-19 Genomics UK Consortium (COG-UK) was formed in the United Kingdom to collect, sequence, and analyze SARS-CoV-2 genomes to understand viral transmission and evolution (https://www.cogconsortium.uk/). Other initiatives such as CDC’s “SPHERES” (SARS-CoV-2 Sequencing for Public Health Emergency Response, Epidemiology, and other fields).
Global Initiative on Sharing All Influenza Data (GISAID). GISAID was formed by the Department of Biotechnology, Ministry of Science and Technology, Government of India along with CSIR and ICMR, including National Institutes of Health and the World Health Organization (WHO). The first database was originally developed for rapid international exchange of all influenza virus genomic and proteomic data, and has now been expanded to include SARS-CoV-2 genomic data. The first SARS-CoV-2 whole-genome sequences were made publicly available on January 10, 2020, allowing for worldwide responses to the pandemic (Lu et al., 2020). Within six months more than 57,000 SARS-CoV-2 genomes from around 100 countries were deposited. GISAID combines sequence data with epidemiological information and provides real-time genomic surveillance to monitor the emergence of SARS-CoV-2 variants in different parts of the world. GISAID is the most commonly used database for the SARS-CoV-2. As of December 2021, more than 5.7 million SARS-CoV-2 genome sequences from nearly 200 countries around the world were shared on the GISAID database (https://www.gisaid.org/). The largest proportion of sequences shared from Europe (58.2%), then North America (31.8%), Asia (5.8%), South America (1.9%), Africa (1.1%), with the fewest from Oceania (0.0%) (https://www.gisaid.org/). In addition to GISAID, other existing genomic and proteomic databases have been updated and used to provide SARS-CoV-2 resources (listed in Table 2).

### Table 1

| Name of the genomics consortium/sequencing initiative | Country/ region | Source link |
|--------------------------------------------------------|-----------------|-------------|
| Africa CDC Institute for Pathogen Genomics | Africa | [https://africacdc.org/institut es/](https://africacdc.org/institutes/)|
| Canadian COVID Genomics Network (CanCOGeN) | Canada | [https://www.genomecanada.ca/en/cancogen](https://www.genomecanada.ca/en/cancogen)|
| Coronavirus Sequencing in Quebec (CoVSeq) | Canada | [https://coveq.ca/](https://coveq.ca/)|
| COVID-19 Genomics UK Consortium (COG-UK) | United Kingdom | [https://www.cogconsortium.org.uk/](https://www.cogconsortium.org.uk/)|
| COVID-19 Network Investigations (CONI) alliance | Thailand | [https://coni.team/](https://coni.team/)|
| Danish Covid-19 Genome Consortium (DCCG) | Denmark | [https://www.covi-19genomics.dk/home](https://www.covi-19genomics.dk/home) |
| COVID-19 OMICS Initiative (DeCOI) | Germany | [https://decoi.eu/](https://decoi.eu/)|
| Indian SARS-CoV-2 Genomics Consortium (INSACOG) | India | [https://dibindia.gov.in/i nsacog](https://dibindia.gov.in/insacog)|
| Irish Coronavirus sequencing consortium | Ireland | [https://www.teagasc.ie/](https://www.teagasc.ie/)|
| Mutational Dynamics of SARS-CoV-2 in Austria | Austria | [https://www.sarscov-2.at](https://www.sarscov-2.at)|
| National Institute of Infectious Diseases | Japan | [https://www.nih.go.jp](https://www.nih.go.jp)|
| RIVM – National Institute for Public Health and the Environment | Netherlands | [https://coronavirus-2019.rivm.nl/en/c](https://coronavirus-2019.rivm.nl/en/c)|
| SeqCOVID – genomic epidemiology of SARS-CoV-2 | Spain | [http://seqcovid.cscie.es/](http://seqcovid.cscie.es/)|
| SPHERES consortium (SARS-CoV-2 Sequencing for Public Health Emergency Response, Epidemiology, and Surveillance) | United States | [https://www.cdc.gov/coronavirus-2019-ncov/variants/spheres.html](https://www.cdc.gov/coronavirus-2019-ncov/variants/spheres.html)|
| Switzerland’s Swiss SARS-CoV-2 Sequencing Consortium (S3C) | Switzerland | [https://bse.ethz.ch/covid19genomics/](https://bse.ethz.ch/covid19genomics/)|
| ARTIC network’s Real-Time Molecular Epidemiology For Outbreak Response | Global | [https://artic.network/](https://artic.network/)|
| COVID-19 High Performance Computing (HPC) Consortium | Global | [https://covid19-hpc-consor tiu.org](https://covid19-hpc-consortiu.org) |
| Public Health Alliance for Genomic Epidemiology (PHAGE) | Global | [https://pha4ge.org](https://pha4ge.org)|
| The COVID-19 host genetics initiative | Global | [https://www.covid19hg.org/](https://www.covid19hg.org/)|

### Table 2

| Database | Data type (No. of entries) | References | Source link |
|----------|---------------------------|------------|-------------|
| GISAID SARS-CoV-2 database | SARS-CoV-2 genome sequences (3,445,483) | Khare et al., 2021 | [https://www.gisaid.org/](https://www.gisaid.org/)|
| DNA Databank of Japan (DDBJ) | Sequence data of SARS-CoV-2 (3,496 entries) | Okido et al., 2021 | [https://www.ddbj.nig.ac.jp/](https://www.ddbj.nig.ac.jp/)|
| EMBL-EBI COVID-19 Data Portal (CDP) | Sequenced samples (376,298), Studies (392), Genes (22), Browser (1), Variants (12,691) | De Silva et al., 2021; Harrison et al., 2021 | [https://www.ebi.ac.uk/ena/pathogens/covid/](https://www.ebi.ac.uk/ena/pathogens/covid/)|
| NCBI SARS-CoV-2 Resources | SRA runs (1,107,163), Nucleotide records (1,368,700), Clinical studies related to COVID-19 (6,533), Pubmed (175,769), PMC (202,051) | Sayers et al., 2022 | [https://www.ncbi.nlm.nih.gov/sars-cov-2/](https://www.ncbi.nlm.nih.gov/sars-cov-2/)|
| NCBI SARS-CoV-2 protein families | Macromolecules (867), compounds (725), Protein families (171) | Varadi et al., 2021 | [https://www.ebi.ac.uk/pdbe/covid-19](https://www.ebi.ac.uk/pdbe/covid-19)|
| RCSB-PDB Protein Data Bank | SARS-CoV-2 protein structures (6,476), protease (374), NCP (458) | Burley et al., 2021 | [https://rcsb.org/covid19](https://rcsb.org/covid19)|

1. As of 03.12.2021.
2021). A database update can be highly recommended in order to increase the quality of the genomic data. So many user-friendly web-based tools were created to overcome the problem of data processing and interpretation, for example, Phylogenetic Assignment of Named Global Outbreak Lineages (PANGOLIN) (https://pangolin.cog-uk.io/) for lineage assignment, Nextstrain (https://nextstrain.org/), CoVizU (http://fllogeneti.ca/covizu/), and Microreact (https://microreact.org/) for data visualization. All these databases aided the scientists in deciphering SARS-CoV-2 mutations, developing appropriate diagnostic kits, and tracking the outbreaks all around the planet.

4.2. Tracking SARS-CoV-2 variants and their spread

SARS-CoV-2, like other viruses, changes over time to adapt to changing environments. The majority of mutations are neutral that have little effect on the functional properties of the virus. There are certain mutations that may be significant, for example, when they encode essential components like the SARS-CoV-2 spike glycoprotein, which serves as a key for the virus to enter host cells and initiate infection (Zhang et al., 2020a). Genomic analyses indicate that some changes may confer a selective advantage to the virus and lead to increased fitness such as antiviral drug resistance and immune escape (Harvey et al., 2021). Even a single amino acid change may alter the severity of illness it causes, infectivity, transmissibility, host immunity responses, the effectiveness of vaccines, therapeutics, and other public health measures (Van Dorp et al., 2020). Since the beginning of the SARS-CoV-2 pandemic, the World Health Organization (WHO) and its international networks have been tracking the evolution of the SARS-CoV-2 genome and updating the variants of interest (VOI) and variants of concern (VOC) (Konings et al., 2021). As of December 2021, there are five variants of concern (VOC) designated by WHO such as Alpha (B.1.1.7), Beta (B.1.351), Gamma (P.1), Delta (B.1.617.2) and Omicron (B.1.1.529) (https://www.who.int/en/activities/tracking-SARS-CoV-2-variants/). Alpha (B.1.1.7) variant, the first VOC exhibiting greater transmissibility, was emerged in the United Kingdom (UK) in September 2020 (Davies et al., 2021). The Delta (B.1.617.2) variant, first identified in December 2020 in India, is now replacing the pre-existing lineages such as Kappa (B.1.617.1) and Alpha (B.1.1.7) and causing re-emergence even in countries with high vaccination coverage (Kannan et al., 2021). Genomic sequencing plays a major role in identifying the viral variants and hotspots of transmission. Korber et al. (2020) designed a bioinformatics pipeline to monitor SARS-CoV-2 variants using the data from the GISAID SARS-CoV-2 sequence database. The pipeline tracks the changes in spike glycoprotein overtime to find the variants that are increasing in different geographic regions at the same time. Their findings showed that during a month, a SARS-CoV-2 variant bearing a specific spike mutation (D614G) became globally dominant.

4.3. Inferring epidemiological parameters and transmission dynamics

Various studies used whole-genome sequences (WGS) as a surveillance tool to investigate outbreak dynamics (Bandoy and Weimer, 2021; Oude Munnink et al., 2020; Bukin et al., 2021) and to infer epidemiological parameters like reproductive number (Geidelberg et al., 2021). For instance, Banu et al. (2020) analyzed the phylogenetic clusters of SARS-CoV-2 genomes to rule out the emergence of COVID-19 in India and suggested that the common ancestor might have emerged at the end of January 2020 and resulted in an outbreak followed by the nationwide spread. Bousali et al. (2021) performed phylogenetic and phylodynamics analysis using SARS-CoV-2 genome sequences derived from ten European regions to investigate the Molecular Transmission Clusters (MTCs). Pan et al. (2021) conducted a phylogenetic analysis using a large number of SARS-CoV-2 genomic sequences from GenBank and GISAID databases to identify the epidemiological traits of COVID-19 and observed the diverse sources of transmission and transmission routes of SARS-CoV-2 in different countries. Geidelberg et al. (2021) estimated the growth rate and reproduction number of SARS-CoV-2 by phylogenetic analysis of genetic sequences obtained from confirmed COVID-19 cases in China. Similarly, Romero et al. (2021) estimated effective reproductive number (Rt) using genomic data of SARS-CoV-2 in Peru. Based on these studies, researchers were able to combine genomic data with epidemiological data to understand the transmission dynamics of SARS-CoV-2 and to take timely public health measures, including regional lockdowns and travel restriction.

5. Role of genomics in diagnosis and therapy of COVID-19

Genomic medicine is an advanced discipline that focuses on how genomic information is used in clinical diagnosis, therapy, and predicting outcomes (Oishi et al., 2015).

5.1. Development of molecular diagnostics

Management of COVID-19 requires prompt diagnosis, effective therapy, and future prevention. The availability of the first whole-genome sequences of SARS-CoV-2 facilitated the rapid development of molecular diagnostic techniques, particularly nucleic acid-based diagnostic assays such as real-time reverse transcription-polymerase chain reaction (RT-PCR), Transcription-Mediated Amplification (TMA), loop-mediated isothermal amplification (LAMP), and CRISPR/Cas-based assays (Broughton et al., 2020; Carter et al., 2020; Caruana et al., 2020; Corman et al., 2020; Shen et al., 2020; Wang et al., 2020). These approaches were further improved and refined to make them more specific to the viral variants in different geographical regions. Because of its sensitivity and specificity, RT-PCR is considered the ‘gold standard’ among nucleic acid tests for detection and screening of COVID-19 (Corman et al., 2020). Viral genomic sequences are needed for designing the primers and probes that would efficiently bind to SARS-CoV-2 nucleic acid. Several SARS-CoV-2 genomic areas, including the RdRp gene in the ORF1ab sequence, the S gene, N gene, and E gene, are used in RT-PCR assays to diagnose COVID-19 (Wang et al., 2020). Mutations in the primer and probe-target areas of the SARS-CoV-2 genome can lead to false-negative results (Khan and Cheung, 2020). Therefore to improve the accuracy of detection and to reduce the risk of false negatives, the virus is detected with several targets, such as multiplex real-time RT-PCR methods targeting two or more sections of the viral genome (Ishige et al., 2020; Tahamtan and Ardebili, 2020). The risk of diminished diagnostic efficiency is also avoided by developing diagnostics based on relatively stable conserved regions of the genome (Ascoli, 2021). As the virus continues to evolve, genome sequencing is necessary for monitoring the mutations that would hinder the ability of diagnostic assays to detect SARS-CoV-2 (Jain et al., 2021). Advances in genomics and proteomics enabled the cloning and expression of SARS-CoV-2 viral proteins, which aided the development of inexpensive rapid diagnostic tests for detection of SARS-CoV-2 at the point of care such as antigen and serological tests (Topian et al., 2021; Mercer and Salit, 2021).

5.2. Development of therapeutics

Having access to the genome of the SARS-CoV-2 virus allows researchers to identify therapeutic targets and to build models of epitopes and immune responses, allowing the development of new therapeutics and vaccines (Chellapandi and Saranya, 2020; Li et al., 2020; Zhou et al., 2020a; Peng et al., 2021). Both genomics and proteomics enabled the rapid understanding of viral protein function and pathogenesis, as well as the identification of virus-specific factors and potential targets for drug design. COVID-19 might be treated using drugs that target any of the key proteins involved in viral replication (Table 3). For example, the drug Remdesivir inhibits RdRp (RNA-dependent RNA polymerase) and has been approved for the treatment of COVID-19 in various countries after showing improvement in clinical studies (Beigel et al., 2020;
n novel antigens, which need to be tested by using experimental biology (Bambini and Rappuoli, 2009). Pan-genomic reverse vaccinology, which involves the comparison of genomic data from different strains of SARS-CoV-2, enhances the opportunity of developing novel vaccines (Enayathkani et al., 2020). Novel epitopes in proteins encoded in the genomes can be predicted in-silico using bioinformatics/immunoinformatics tools based on sequence similarities to previously reported immunogenic motifs or structural approaches such as molecular docking simulations (Ishack and Lipner, 2021). As the genomic and proteomic information of SARS-CoV-2 is rapidly becoming accessible, numerous studies applied reverse vaccinology and machine learning approaches to develop multi-epitope subunit vaccines against SARS-CoV-2 (Enayathkani et al., 2020; Ong et al., 2020; Sanami et al., 2020; Tahir ul Qamar et al., 2020; Almofti et al., 2021; Saha et al., 2021). Recombinant protein vaccines against SARS-CoV-2 include spike-protein-based, RBD-based, and virus-like particle (VLP)-based vaccines (Krammer, 2020) and 53 of them are in the clinical phase. For example, Novavax’s NVX-CoV2373 vaccine is made up of full-length recombinant SARS-CoV-2 spike glycoproteins nanoparticles that have been adjuvanted with Matrix-M1 (Keech et al., 2020). Despite the widespread occurrence of the B.1.1.7 (or alpha) variant, preliminary findings of phase 3 clinical trial in the UK showed an efficacy rate of 86.3% against the alpha variant and 96.4% efficacy against non-alpha variants (Heath et al., 2021). The majority of the recombinant protein subunit vaccines against SARS-CoV-2 have entered the phase 3 clinical trials (Table 4).

Nucleic acid vaccines either RNA or DNA deliver the genetic information of antigen (such as spike glycoprotein) rather than the antigen itself. 15 DNA-based and 21 mRNA-based candidate vaccines against SARS-CoV-2 are in clinical trials. Due to the ease of handling, simple manufacture, and stability of plasmid DNA, DNA-based vaccination methods have become a reality. Of the eleven DNA-based vaccines, only two vaccines, ZyCoV-D (developed by Zydus Cadila) and INO-4800 (by Inovio Pharmaceuticals) have undergone phase 3 clinical trials. ZyCoV-D uses plasmid DNA that contains the genetic information to make the 'spike protein' (Momin et al., 2021). It is the world’s first plasmid DNA vaccine for COVID-19 to be approved for emergency use (Mallapaty, 2021). Next to protein vaccines, the majority of the vaccine candidates are mRNA-based which accounts for 16% of all vaccines developed across platforms (Supplementary Table S1). Two mRNA vaccines, Pfizer-BioNTech’s BNT162b2 (Comirnaty) and Moderna’s mRNA-1273 (Spikevax) were the first to be authorized for use in many countries (Baden et al., 2021; Haas et al., 2021). These vaccines use nucleoside-modified mRNA (modRNA) encoding SARS-CoV-2 spike protein that is encapsulated in lipid nanoparticles (LNP). Other mRNA vaccines, such as CVnCoV (developed by CureVac and CEPI) and ARCoV (developed by Valneva Biotechnology) are in phase 3 clinical trials (Table 4).

The mRNA vaccines provide a variety of advantages over other vaccine platforms, including efficient delivery, flexibility, short development time, use of the host’s protein translational machinery, and no risk of genome integration (Momin et al., 2021). The era of synthetic genomics led to the development of viral vector-based vaccines that deliver antigen-coding nucleic acid fragments to host cells through viral vectors. Viruses are altered to lower their virulence and their reproduction capability but retaining their ability to infect human cells (Alter et al., 2021). At present, four non-replicating adenovirus-vector vaccines such as Oxford/AstraZeneca’s AZD1222, Janssen’s Ad26.COV2.S, CanSino’s ADS-5102 (Convidecia), Gamaleya Research Institute’s Gam-COVID-Vac (Sputnik V) are now in widespread use (Table 4). All these contain DNA that encodes a SARS-CoV-2 spike protein.

7. Conclusion

The COVID-19 pandemic startled the globe, pushing science to develop new strategies to combat the virus. The availability of genomic data enables a very rapid, thorough, and precise global follow-up of the progression of the COVID-19. Early detection of SARS-CoV-2 variants,
| Vaccine platform | Vaccine name | Type of candidate vaccine | Developer/manufacturer | Clinical stage | Reference |
|------------------|--------------|---------------------------|------------------------|---------------|-----------|
| DNA based vaccine | ZyCoV-D | Plasmid DNA Covid-19 vaccine | Zydus Cadila | Phase 3 | Momin et al., 2021 |
| | INO-4800 COVID-19 Vaccine | Plasmid DNA Covid-19 vaccine | Inovio Pharmaceuticals + International Vaccine Institute + Advaccine (Suzhou) | Phase 3 | Andrade et al., 2021 |
| Inactivated virus | CoronaVac* | Inactivated SARS-CoV-2 vaccine, produced in Vero cells | Sinovac Biotech | Phase 4 | Tanriover et al., 2021 |
| | WIBP-CorV | Inactivated SARS-CoV-2 vaccine, produced in Vero cells | Sinopharm + China National Biotech Group Co + Wuhan Institute of Biological Products (WIBP) | Phase 3 | Al Kaabi et al., 2021 |
| | BBIBP-CorV* | Inactivated SARS-CoV-2 vaccine, produced in Vero cells | Sinopharm + China National Biotech Group Co + Beijing Institute of Biological Products (BBBP) | Phase 4 | Xia et al., 2021 |
| | Covifidal or IMBCAMS COVID-19 vaccine | Inactivated SARS-CoV-2 vaccine, produced in Vero cells | Institute of Medical Biology (IMB) + Chinese Academy of Medical Sciences (CAMS) | Phase 3 | Huang et al., 2021 |
| | QaaVac or QaaCovid-in | Inactivated SARS-CoV-2 vaccine, produced in Vero cells | Research Institute for Biological Safety Problems, Kazakhstan | Phase 3 | Ganneru et al., 2021 |
| | Covaxin (BBV152)* | Whole-virion Inactivated SARS-CoV-2 Vaccine (Vero Cell) | Bharat Biotech International Limited + Indian Council of Medical Research (ICMR) | Phase 3 | https://www.biocine.com/ |
| Protein subunit | KCONVAC or Minhai COVID-19 vaccine | Inactivated SARS-CoV-2 vaccine, produced in Vero cells | Shenzhen Kangtai Biological Products Co., Ltd. + Beijing Minhai Biotechnology | Phase 3 | https://en.biokangtai.com/ |
| | VLA2001 or Valneva COVID-19 vaccine | Inactivated SARS-CoV-2 vaccine, produced in Vero cells | Valneva, National Institute for Health Research, United Kingdom | Phase 3 | https://valneva.com/research-development/covid-19-vla2001/ |
| | ERUCOV-VAC or TURKOVAC | Inactivated SARS-CoV-2 vaccine (Vero cell) | Health Institutes of Turkey + Erciyes University | Phase 3 | http://www.erciyes.edu.tr/ |
| | NVX-CoV2373 | SARS-CoV-2 rS/MATRIX M1-Adjuvant (Full-length recombinant SARS-CoV-2 glycoprotein nanoparticle vaccine adjuvanted with Matrix M) | Novavax + Coalition for Epidemic Preparedness Innovations (CEPI) | Phase 3 | Heath et al., 2021 |
| | ZIFIVAX or ZF2001 | Recombinant SARS-CoV-2 (CHO Cell) – RBD-based protein subunit vaccine | Anhui Zhifei Longcom Biopharmaceutical + Institute of Microbiology, Chinese Academy of Sciences | Phase 3 | Yang et al., 2021 |
| | VAT00002 | SARS-CoV-2 S protein with adjuvant | Sanofi Pasteur + GSK | Phase 3 | https://www.sanofi.com/en/our-covid-19-vaccine-candidates |
| | SCB-2019 | Trimeric subunit Spike Protein vaccine + Cpg 1018 adjuvant plus Alum adjuvant | Clover Biopharmaceuticals Inc./GSK/Dynavax | Phase 3 | Richmond et al., 2021 |
| | COVAX-19 (or SpikeGen) | Recombinant spike protein + adjuvant | Vaxine + CimaGen Co. | Phase 3 | https://vaxine.net/ |
| | MVC-COV1901 | Spike-ZP protein + adjuvant Cpg 1018 | Medigen Vaccine Biologics Corporation + Dynavax Technologies + National Institute of Health Instituto Finlay de Vacunas Cuba | Phase 4 | Hsieh et al., 2021 |
| | FINLAV-FR-2 or Soberana 2 | RBD chemically conjugated to tetanus toxoid plus adjuvant | Based on peptide antigens | Phase 3 | Chang-Monteguido et al., 2021 |
| | EpiVacCorona | RBD chemically conjugated to tetanus toxoid plus adjuvant | Based on peptide antigens | Phase 3 | Ryzhikov et al., 2021 |
| | Recombinant SARS-CoV-2 vaccine (Sf9 Cell) | RBD (baculovirus production expressed in Sf9 cells) | West China Hospital + Sichuan University | Phase 3 | Meng et al., 2021 |
| | CIGB-66 (or Abdala) | RBD + aluminium hydroxide | Center for Genetic Engineering and Biotechnology (CIGB) | Phase 3 | http://www.cigb.edu.cn/ |
| | BECOV2A (Corbevax) | RBD + aluminium hydroxide + Cpg 1018 | Biological E. Limited | Phase 3 | https://www.biological.com/Vaccines_BioBiotics/products.html |
| | Nanocovax | Recombinant Sars-CoV-2 Spike protein, Aluminum adjuvanted | Nanogen Pharmaceutical Biotechnology JSC | Phase 3 | https://nanogenpharma.com/products.html |
| | GBP510 | Recombinant surface protein vaccine with adjuvant AS03 (Aluminium hydroxide) | SK Bioscience Co., Ltd. and Coalition for Epidemic Preparedness Innovations (CEPI) | Phase 3 | https://www.skbioscience.co.kr/ |
| | Razi Gov Pars | Recombinant spike protein | Iranian Razi Vaccine and Serum Research Institute | Phase 3 | http://www.rsvi.ac.ir/ |
| RNA based vaccine | mRNA-1273 (Spikevax)† | Nucleoside-modified mRNA (modRNA) encoding a spike protein, encapsulated in lipid nanoparticles | Moderna + National Institute of Allergy and Infectious Diseases (NIAID) | Phase 4 | Baden et al., 2021 |
| | BNT162b2/ Comirnaty | Nucleoside-modified mRNA encapsulated in a lipid nanoparticle (LNP) | Pfizer/BioNTech + Fosun Pharma | Phase 4 | Haas et al., 2021 |
along with a better knowledge of the mutational processes behind shifting patterns of virulence, transmissibility, and antigenicity, have greatly aided in making timely public health decisions. It is critical to emphasize that genomic information must be utilized with caution while taking public health decisions. The use of improper bioinformatics tools, sampling bias, sequencing errors, and misinterpretation of findings may all lead to wrong conclusions. The genomic sequence of SARS-CoV-2 also enabled the cloning and synthesis of specific viral proteins, which aided in the development of rapid diagnostic tests for SARS-CoV-2 screening. In order to understand SARS-CoV-2 variant spread in different countries, it is essential to integrate genomic and epidemiological data. The increased adoption of genomic technologies in various facets of the worldwide response to the COVID-19 pandemic is major evidence of the role of genomics in modern medicine. Furthermore, the tremendous advances in genomics and lessons learnt from the battle against SARS-CoV-2 offer a great potential to reduce the future threats to mankind and bolster preparedness for future outbreaks.

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

All the authors of the manuscript wish to thank Director and Joint Director (Research), ICAR-Indian Veterinary Research Institute (IVRI), Izatnagar, an institute under the Indian Council of Agricultural Research (ICAR), Department of Agricultural Research and Education (DARE), Government of India for their support and providing infrastructural facility to review this study.

Appendix A. Supplementary material

Supplementary data to this article can be found online at https://doi.org/10.1016/j.gene.2022.146387.

References

Abbott, T.R., Dhamdhere, G., Liu, Y., Lin, X., Goudy, L., Zeng, L., et al., 2020. Development of CRISPR as an Antiviral Strategy to Combat SARS-CoV-2 and Influenza. Cell 181, 865-876.e12. https://doi.org/10.1016/j.cell.2020.04.020.
Al Kaabi, N., Zhang, Y., Xia, S., Yang, Y., Al Qahtani, M.M., et al., 2021. Effect of 2 Inactivated SARS-CoV-2 Vaccines on Symptomatic COVID-19 Infection in Adults. JAMA 326, 35. https://doi.org/10.1001/jama.2021.8565.
Almofti, Y.A., Abd-elrahman, K.A., Eltilib, E.E.M., 2021. Vaccinomic approach for novel multi epitopes vaccine against severe acute respiratory syndrome coronavirus-2 (SARS-CoV-2). BMC Immunol. 22 https://doi.org/10.1186/s12895-021-00412-0.
Alter, G., Yu, J., Liu, J., Chandrashekar, A., Borducchi, E.N., Trotanoki, L.H., et al., 2021. Immunogenicity of Ad26.COV2.S vaccine against SARS-CoV-2 variants in humans. Nature 596, 268-272. https://doi.org/10.1038/s41586-021-03681-2.
Andersen, K.G., Rambaut, A., Lipkin, W.I., Holmes, E.C., Garry, R.F., 2020. Development of CRISPR as an Antiviral Strategy to Combat SARS-CoV-2 and Influenza. Cell 181, 865-876.e12. https://doi.org/10.1016/j.cell.2020.04.020.
Andrade, V.M., Christensen-Quick, A., Agnes, J., Tur, J., Reed, C., Kalia, R., et al., 2021. The proximal origin of SARS-CoV-2. Nat. Med. 26, 450–452. https://doi.org/10.1038/s41591-020-08260-9.
Ascoli, C.A., 2021. Could mutations of SARS-CoV-2 suppress diagnostic detection? Nat. Biotechnol. 39, 274–275. https://doi.org/10.1038/s41587-021-00845-3.
on molecular diagnostic assays. Int. J. Infect. Dis. 102, 460–462. https://doi.org/10.1016/j.ijid.2020.02.086.

Joonhwa, K., Barry, H., Fischer, W.M., Kerr, T., Pantham, B., Kumporrin, K., Janzarain, P., et al., 2021. Genomic surveillance of SARS-CoV-2 in Thailand reveals widespread imported populations, a local lineage expansion and a virus with truncated ORF7a. Virus Res. 292, 108233 https://doi.org/10.1016/j.virusres.2021.108233.

Kannan, S.R., Sprat, J.A., Nayar, S.S., Chand, H.S., Quin, T.P., Lornon, C.L., Byrareddy, S.N., Singh, K., 2021. Evolutionary analysis of the Delta and Delta Plus variants of the SARS-CoV-2 viruses. J. Autoimmun. 124, 102715 https://doi.org/10.1016/j.jaut.2021.102715.

Keech, C., Albert, G., Cho, L., Robertson, A., Reed, P., Neal, S., et al., 2020. Phase 1–2 Trial of a SARS-CoV-2 Recombinant Spike Protein Nanoparticle Vaccine. N. Engl. J. Med. 383, 2230–2232. https://doi.org/10.1056/NEJMoa2002920.

Khan, K.A., Cheng, D., et al. Presence of mismatched bases in diagnostic PCR assays and coronavirus SARS-CoV-2 genome. R. Soc. Open Sci. 7, 200636 https://doi.org/10.1098/rsos.200636.

Khare, S., Gurry, C., Freitas, L., Schultz, M.B., Bach, G., Diallo, A., et al., 2021. GISAID's Role in Pandemic Response. Cdc. China CDC Wkly. 3 (49), 1049–1051. https://doi.org/10.4064/cdcw2021.255.

Khambiton, B.B., Niewoehner, D.F., Stein, M., O'Toole, A., Haverkate, M., Mollers, M., et al., 2020. Rapid SARS-CoV-2 whole-genome sequencing and analysis for informed public health decision-making in the Netherlands. Nat. Med. 26, 1405–1410. https://doi.org/10.1038/s41586-020-2999-y.

Khan, B.B., Ji, Z., Saklikar, S., Guo, W., Liu, J., Patterson, T.A., Hong, L., 2021. Identification of Epidemiological Traits by Analysis of SARS–CoV–2 Sequences. Viruses 13, 764. https://doi.org/10.3390/v13060764.

Khan, K.A., Cheng, D., et al. Presence of mismatched bases in diagnostic PCR assays and coronavirus SARS-CoV-2 genome. R. Soc. Open Sci. 7, 200636 https://doi.org/10.1098/rsos.200636.

Khare, S., Gurry, C., Freitas, L., Schultz, M.B., Bach, G., Diallo, A., et al., 2021. GISAID’s Role in Pandemic Response. Cdc. China CDC Wkly. 3 (49), 1049–1051. https://doi.org/10.4064/cdcw2021.255.
