ECRONICON
OPEN ACCESS

EC PULMONOLOGY AND RESPIRATORY MEDICINE

Research Article

A Review on Genomic Analysis of Sars-Cov-2 (Covid-19) by Using Bioinformatics Tools

Siddavarapu Harika[1], Hema Sekhar Reddy Rajula[2]*, PB Ramesh Babu[1] and Vassilios Fanos[2]

[1]*Depertent of Genetic Engineering, Bharath Institute of Higher Education and Research, India*

[2]*Neonatal Intensive Care Unit, Department of Surgical Sciences, AOU and University of Cagliari, Cagliari, Italy*

***Corresponding Author:** Hema Sekhar Reddy Rajula, PhD Scholar, Marie Curie Fellowship, Department of Surgical Sciences, University of Cagliari, Monserrato, Cagliari, Italy.

Received: February 16, 2021; **Published:** April 06, 2021

Abstract

The genome of the virus is smaller when compared to prokaryotes and eukaryotes. Exactly a year back from now in December 2019, a tiny virus similar to the influenza virus arose in the people of Wuhan in China. Later the virus is identified as Coronavirus disease (COVID-19) and based on the characteristics, it was renamed Severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2). It is an enveloped virus whose genome is positive-sense RNA and belongs to the family coronaviridae of the order Nidivirales and has four genera (alpha, beta, gamma, and delta). Now, our present novel SARS-CoV-2 has beta genera and based on the genomic similarity exhibited in the sequences by the virus the researchers state that the virus origin is probably from the bats. The spike (s) present on the surface of the virus leads protein to viral fusion to the cell membrane and leads to the infection. The common clinical test for the detection of the virus is the Reverse transcriptase-polymerase chain reaction test (RT-PCR) test and to detect the IgG, IgM antibodies by (serology) test. In this article, for the sake of Insilco analysis, we describe the different bioinformatics tools that are available to analyze the genome of the SARS-CoV-2. CoVDB is the database that solely deals with the genomic sequences of Coronavirus. Different types of online bioinformatics tools exclusively contain the genomic sequences of Coronavirus. These databases are used to know the phylogenetic relationships among the genomic sequences that are extracted from different COVID patients.

Keywords: *Severe Acute Respiratory Syndrome 2; Pandemic; Reverse Transcriptase-Polymerase Chain Reaction Test; CoVDB Databank; Bioinformatics Tools; Phylogenetic Relationships*

Abbreviations

RNA: Ribonucleic Acid; COVID: Coronavirus Disease; IgG: Immunoglobulin G; IgM: Immunoglobulin M

Introduction

Bioinformatics is an interdisciplinary field that connects the study of biology with computer science, which uses to analyze the various types of biological data [1]. The development of many novel bioinformatics tools became a requisite for alignment, editing, comparison, and interaction between the sequences in viral genomes. To predict appropriate results in omics research the role of bioinformatics is significant because it helps the researchers to analyze the practical values from the available data and helps to predict their results in an orderly manner [2]. Bioinformatics covers several specialized and advanced areas of biology, such areas are Functional and Structural Genomics, Comparative Genomics, DNA Microarrays, Next Generation Sequencing, Transcriptomics, Proteomics, Metabolomics, and Medical Informatics, etc [3].

Genomics is an interdisciplinary field of molecular biology focusing on the DNA content of living organisms [4]. Genomic techniques are essentially focused on DNA sequencing, DNA structure analysis, genome editing, DNA protein interactions, and phylogenomics. The evolution and interrelationships of coding and non-coding DNA sequences can be done by using DNA sequencing technologies [5].

The human Coronavirus was first characterized in the 1960s and was first identified by Dr. June Almeida in 1964 at her laboratory in St Thomas's Hospital in London [6]. Recently, In December 2019, a cluster of pneumonia cases, which was caused by a newly identified beta Coronavirus occurred in Wuhan, China [7]. This Coronavirus was initially named as the 2019- novel Coronavirus (2019-nCoV) on 12 January 2020 by the World Health Organization (WHO) [8]. The Coronavirus name was given because it has crown-like spikes on its surface. The name has been changed to SARS-CoV-2sinceit is a beta Coronavirus, which is an enveloped non-segmented positive-sense RNA virus (subgenus sarbecovirus, Orthocoronavirinae subfamily) [8]. The common symptoms of the corona infected patients are cold, cough, fever, flu-like symptoms. Sometimes the patients with viruses may be asymptomatic. The virus gets transmitted from person to person by the droplets when the infected person gets sneezed, or while coughing. The incubation period of the virus is 7 - 21 days. There are seven types of coronaviruses that affect people, among them, four types of viruses commonly occur in humans 229E, NL63 (alpha Coronavirus), OC43, HKU1 (beta Coronavirus), and other types of Coronavirus MERS-CoV (the beta Coronavirus that causes Middle East Respiratory Syndrome, or MERS), SARS-CoV (the beta Coronavirus that causes the severe acute respiratory syndrome, or SARS), SARS-CoV-2 (the novel Coronavirus that causes Coronavirus disease 2019 or COVID-19) [2].

Here are the few things we need to go through to fill the research gaps, as all know that this pandemic ruined our daily routine and the whole world is looking for ways to get rid of this pandemic i.e. probably possible when we find a vaccine for it. Moreover, to find the vaccine we need to know the research gaps and areas that need to be focused on assuring better results.

As we know that the exact genomic sequence of this novel Coronavirus was still unknown and knowing the exact sequence is required for better outcomes and it is necessary to develop better bioinformatics tools to analyze the data for approaching better results. The novel COVID-19 exhibits different types of genome sequences and gives rise to mutations and gaps. However, due to several different mutations, it is difficult to identify the exact genomic structure of this Coronavirus and developing a vaccine for it.

In this paper, we describe genome analysis of SARS-CoV-2 (COVID-19) by using different bioinformatics tools; firstly, we had given a glance at bioinformatics and genomics and then an elaborated view on the genomics of SARS-CoV-2. Secondly, to offer approaches to different computational methods and availability of data in the literature and various bio tools used in COVID- 19 analysis. To this end, we present an overview of the genomics of SARS Cov-2, and finally, we offer a perspective on the future directions of this research area.

Genomics of SARS-CoV-2

The genomics of the Coronavirus was identified by sequencing the genetic material and decoded them to get genetic information [9]. Extracted genomes from different organisms have been sequenced, the viral genome has 96% identical to bats and humans [10]. Coronavirus has an oily membrane on the surface protein-packed with genetic instructions to make millions of copies itself. The genetic material of Coronavirus is single standard RNA and it appears as a round winding structure [11]. After unwinding the genetic material of Coronavirus is encoded by 30000 letters of RNA which the infected cells read and translate into several different kinds of viral proteins. A cell infected by a Coronavirus produces millions of new viruses, all carrying copies of the original genome. The genome of the cell that transformed from one to another sometimes may skip a letter and replace it with another and thus, this mistake leads to mutations [12]. If the mutation caused in the third position of the codon then the amino acid does not change are called silent mutations. If the mutation occurred in the middle position of the codon then it results in the different amino acids are called non-silent mutations [13]. By tracking the genetic mutations, we can study their association with the virus infectivity, disease severity, and possible response to therapy and vaccine [14]. As coronavirus spread from person to person, they randomly accumulate more mutations. These errors may impact the virus structure, function, and vaccine, and these act as markers to construct the chain of transmission [15]. The genome of SARS-CoV-2 mainly contains genetic material RNA in the first ORF which further translates into two polyproteins as pp1a and pp1b and encodes for 16 non-structural proteins whereas the other ORFs encodes for other accessory and structural proteins. The other parts of the virus code for proteins spike (S) glycoprotein, small envelope (E) protein, matrix (M) protein, and nucleocapsid (N) protein [8,16].

Materials and Methods

If a person gets affected and tested positive for SARS-CoV-2 can be confirmed by conducting several clinical tests. These clinical tests can be performed in several steps. The analysis of SARS-CoV-2 can be done by a sequence of steps and detailed information described in the form of a flowchart as shown in figure 1.

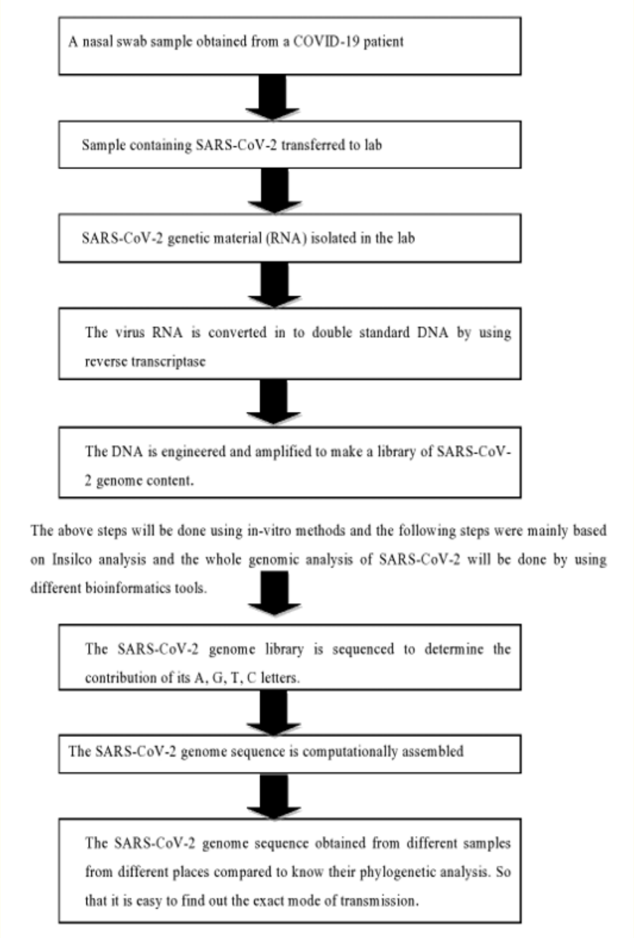
A nasal swab sample obtained from a COVID-19 patient
Sample containing SARS-CoV-2 transferred to lab
SARS-CoV-2 genetic material (RNA) isolated in the lab
The virus RNA is converted in to double standard DNA by using reverse transcriptase
The DNA is engineered and amplified to make a library of SARS-CoV-2 genome content.
The above steps will be done using in-vitro methods and the following steps were mainly based on Insilco analysis and the whole genomic analysis of SARS-CoV-2 will be done by using different bioinformatics tools.
The SARS-CoV-2 genome library is sequenced to determine the contribution of its A, G, T, C letters.
The SARS-CoV-2 genome sequence is computationally assembled
The SARS-CoV-2 genome sequence obtained from different samples from different places compared to know their phylogenetic analysis. So that it is easy to find out the exact mode of transmission.

Figure 1: *The flow chart of sequence steps in the analysis of SARS-CoV-2.*

The methodology of the systematic research

Search strategy

A systematic search was performed on PubMed, Springer, Nature, Lancet, and Cell by using the following search terms: “spike protein”, “SARS”, “Coronavirus”, “big data”, “pharmacogenomics” AND “bioinformatics tools”. This search strategy was augmented by identifying additional original reports aligned with the objectives of this paper by tracking the citations from the reference lists of included articles.

Identification of eligible studies

We included studies evaluating the primary outcome (s) of interest that was (were) categorized in any of the following domains: (i) spike protein, (ii) SARS, (iii) Coronavirus, (iv) big data, (v) bioinformatics tools were independently reviewed for eligibility by all authors. If an article was deemed potentially eligible based on the title/abstract review, a full- text review was completed. Decisions for the inclusion of the studies in the full-text review were made by consensus. Only original research articles were included. The methodology of systematic research has been performed according to Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) and illustrated in figure 2.

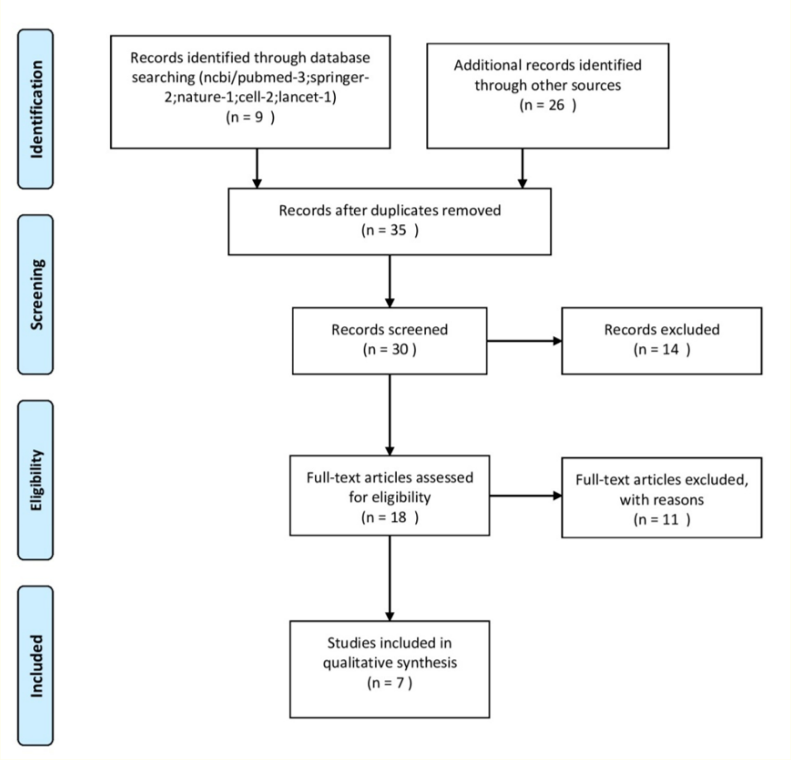
Identification
Screening
Eligibility
Included
Records identified through database searching (ncbi/pubmed-3;springer-2;nature-1;cell-2;lancet-1) (n = 9)
Additional records identified through other sources (n = 26)
Records after duplicates removed (n = 35)
Records screened (n = 30)
Records excluded (n = 14)
Full-text articles assessed for eligibility (n = 18)
Full-text articles excluded, with reasons (n = 11)
Studies included in qualitative synthesis (n = 7)

Figure 2: *PRISMA diagram of systematic research.*

Computational methods

Homology genome blast and genome information: The utmost step is that we need to retrieve the whole genome sequence of SARS-CoV-2 from the NCBI nucleotide database with the help of a reference number NC_045512.2, followed by aligning the retrieved sequences using Blastn sequence aligner.

Open reading frame reader (ORF): The main usage of ORF is to predict the starting sequence of DNA and with the help of SMARTBLAST or Blastp can not only find the new genome sequences but also can identify the ATG or initiator codon sets.

Alignment of genome analysis: As all know that genome analysis deals with nucleotide and amino acid sequences meanwhile, different tools useful for analyzing the data, in particular, WebDSV2.0 used in nucleotide analysis, and multiple sequence alignment tools like ClustalW used for amino acid sequence analysis. Moreover, comparing both nucleotide and amino acid sequences can be done by using the pairwise alignment EMBL EBI database.

Visualization of models: Visualization is the very next step after comparing the sequences and pymol software and Autodock tools are used for the prediction of models [17].

Data availability

Database: As there are many different types of genomic sequences available under the name of SARS-CoV-2 and the retrieval of the exact genomic sequence has become a challenge for the scientists and there is a lot of confusion arises with the accession numbers allotted for the genomic sequences. To overcome this problem, a new database exclusively for this Coronavirus is developed as the CoVDB database [18].

CoVDB database: A database specially designed for the retrieval of Coronavirus genomic sequences which are officially controlled by the MYSQL database. As per the source till now 3982 different sequences are collected, among them, 264 are complete genome sequences and the remaining are partial sequences [19]. However, it is an open-access database, and easy to retrieve the information from it and gather the batch genomes retrieval and polyprotein annotations.

SPHERES (SARS-CoV-2 sequencing for public health emergency response, epidemiology, and surveillance): It is a new national consortium uses to coordinate the SARS-CoV-2 sequencing across The United States. It is led by the CDC and works on eight core objectives [20]. The overview of the objectives is to bring all the institutions working on this pandemic and sharing the information and to know the new bioinformatics tools for the data analysis. Data on COVID-19 by our world in data: It is an open-access data source maintained by our world in Data. Here we can access the data of Coronavirus like daily confirmed cases, deaths, testing [21]. The Data are available in different formats like CSV, XLSX, JSON and it contains the historical data of pandemic to recent information on the pandemic.

Results and Discussion

Firstly, we found that the genomics of SARS-CoV-2 is complex and the virus has thrown a challenge for the scientists to find the exact genomic sequence. After a lot of studies, scientists concluded that different types of genomic sequences are existed because of the mutations. Mainly D614 and G614 are the strains that increase the infectivity of the virus [22]. The list of various types of bio tools that may help in the prediction of COVID-19 has been described in table 1.

S.NO	Bio tool Name	Description
1	DeepCLIP	By using deep learning we can predict the effect of mutations on protein –RNA binding. It is a neural network connected to a bidirectional LSTM layer with shallow convolution layers [31].
2	V-pipe	It is a bioinformatics pipeline uses for the analysis of next-generation sequencing data, which is derived from the intra host viral populations.
3	DisGeNET	It is a publicly available source that contains a collection of genes and variants associated with human diseases. By this collection, one can find phenotypic and genotypic relationships [32].
4	ARTIC	It is also a bioinformatics pipeline used for virus sequencing data with the nanopore.
5	SWISS-MODEL Workspace	As all know that SWISS-MODEL is an automated homology modeling server that helps to build the structure of a protein by using a template.
6	COVID ep	It is a database that is specially designed to maintain the complete set of B-cell and T-cell epitopes which can act as vaccine targets for SARS-CoV-2 [33].
7	COVID-19 Dashboard	COVID-19 Dashboard is an open-source that contains information about Coronavirus and its spread day today and the information is displayed as maps, graphs, and tables.
8	UCPH COVID19 DASHBOARD	It is a dashboard that provides visualizing tools that help to analyze the COVID-19 spread in different countries across the world.
9	The Coronavirus App	It is an app specially designed to track the spread of the Coronavirus outbreak in Wuhan and can check the affected regions.
10	CoV2ID	It is a database that contains a complete list of verified oligonucleotides for SARS-CoV-2 [34].
11	COVID-19 Knet Miner	It is a database that visualizes human biological data related to SARS-CoV-2. The data contains the gene, protein domain, phenotypes, homology, and interactions between them.
12	MALVIRUS	It is an accurate tool used for the genotyping of haploid individuals and it has two steps. It doesn't require any type of mapping to a reference genome.
13	CovidMine	It is a data source that gives us the information related to the confirmed COVID-19 cases and deaths and also the reference genomes, nucleotide sequences that are deposited in the Gene bank.
14	SIENA	It is a tool that allows to search for protein binding sites in the protein databank and can analyse the point mutations, protein flexibility by virtual screening.
15	Protoss	It is an automated hydrogen prediction tool used to predict protein-ligand complexes and it fills the missed hydrogen atoms in the protein structures. Meanwhile, it calculates the optimal degrees of freedom in the hydrogen bonding.
16.	BLAST Beta Coronavirus Database	It is a database that is exclusively designed for sequencing the nucleotide and protein sequences of Coronavirus.
17	VADR	Based on Refseq annotation, the classification and annotation of viral sequences can be done.
18	Viral Bioinformatics	The role of viral bioinformatics is to provide access to viral genomes and to discover a novel set of tools that helps in genome analyses.
19	Prokka	It is a tool that helps to annotate the bacterial, archeal, and viral genomes and results in standards outputs.
20	Viral Zone	It is a resource that contains the entire viral genus, families, and molecular structures along with epidemiological information. One can access the information from UniprotKB/Swissprot.

Table 1: *List of various types of bio tools that may help in the prediction of COVID-19 [30].*

Next-generation sequencing methods, which permit to do the phylogenetic analysis that helps to know the exact infection source and from which means the virus gets transmitted to each other [23]. According to different case studies, we found that 49 persons who are affected by Coronavirus exhibit 20 different types of Coronavirus genomes [24]. The summary of the findings of the studies included in the systematic review has been described in table 2.

Year	Reference	Major Findings	Methods and techniques	Limitations
1964	[6]	The novel coronavirus has discovered long back by Dr. June Almeida	Microscopic technique	Failed to discover with traditional methods
2016	[13]	The similarity in the sequences cannot be observed in collected samples from different patients tested positive for Coronavirus due to mutations that occurred in the codons.	Mechanisms adopted were cap dependent coV mRNAs to control their translation process.	At the post transcriptional level, we cannot understand the biologic functions of the leader sequence in other covid sequences.
2019	[15]	A cluster of cases registered with the novel viruses and the virus probably originated from a bat.	A sequence of steps involved 1. Data reporting 2. Sample collection 3. Virus isolation, cell infection, electron microscopy, and neutralization assay 4. RNA extraction and PCR 5. Serological test 6. Phylogenetic analysis	Failed to find the exact genomic sequence of coronavirus.
2020	[19]	CoVDB database is a database with a collection of different genomic sequences exhibited by a coronavirus.	1. All non-structural proteins in the sequences are encoded. 2. Encoded sequences of non-structural proteins are annotated by using Orf 1 ab protein. 3. Identical sequences are separated from the whole genome sequence.	This database contains the genomic sequences of coronaviruses. We cannot use this database for any other different viruses.
2020	[30]	Helps to analyse the coronavirus genome with the help of different tools.	Based on the expected outcome we need to choose our bio tool accordingly.	It is difficult to pick our exact bio tool from the source as it contains a large number of different databases.
2020	[22]	D614 and G614 are the strains that increase the infectivity of the virus.	*Invitro* technique	D614 and G614 strains are harder to control because they exhibit distinct genotype.
2020	[31]	RT-PCR tests give accurate results than serology tests.	1. Collection of samples by swab tests. 2. Extraction of genome 3. Analysing the data.	Serology tests are used to detect the presence of antibodies that fight against the virus. Based on that we cannot assure serology tests will give exact results when we use them as clinical tests for corona testing.

Table 2: *Summary of findings of the studies included in the systematic review.*

In the case of clinical tests, the novel virus has been detected by the RT-PCR test antibody (serology) test [25]. For accurate results and if the person has symptoms of COVID-19 and the person has been exposed to someone with the virus are advised to take the RT-PCR test and to identify any past presence of COVID-19 without our knowledge we need to go for a serology test [26].

However, these computational methods and data availability sections will play a crucial role in advanced research of this novel coronavirus because these methods and data will help the researchers as a quick reference for further research. Besides these methods, bio tools play a major role to get accurate results and they reduce wet-lab experimentation time.

Conclusion

The globe is eagerly waiting for a significant outcome. Although, several other ways have owned to control this deadly virus, ended up with passive results. Initially, scientists thought that this novel virus is similar to influenza but the anti-viral drugs became invalid for this virus [8]. Eventually, plasma therapy is helping up to some extent for the recovery of COVID patients [27] but some objections have been raised on it, so it is not accepted globally.

Now the only option to ruin this pandemic is a vaccine. But for that exact genomic sequence is required, it became challenging for scientists to discover vaccine or medicine since the genome of the virus varies in each individual due to silent or non-silent mutations caused by them and cannot predict exactly which type of proteins can act as antibodies on the virus. In recent studies, we came to know that fortunately, scientists approached better results on it [28-31].

Firstly, after the discovery of Coronavirus with the help of microscopic techniques in ancient times the world had just known the existence of a novel virus. Secondly, we need to look over some drawbacks which became a huge encumber for scientist's contributions to determine the exact sequences in the genome. Thirdly, scientists discovered the source of Coronavirus. Fourthly, here we need to notice that rapid tests and serology tests, which are owned by the doctors to confirm the Coronavirus give passive results because there are not accurate. Finally, we need to consider that the databases established by different sources can only be used for the Coronavirus but on the other side they cannot show the impact on other viruses it means they are not useful in the detection of other novel viruses [32-34].

Recently, scientists from the Rosalind Franklin institute claimed that llamas antibodies or nanobodies can be used to block Coronavirus [35]. In addition to this, another approach is that nanobodies get bind to the spike proteins of viruses and inhibit entering into human cells. Many other countries vaccines are in the clinical trial stage, among them covaxin by Bharath biotech and Astrazeneca by Oxford showing effective results.

Conflicts of Interest

The authors have no relevant financial involvement with any organization or entity with a financial interest in or financial conflict with the subject matter or materials discussed in the manuscript. This includes employment, consultancies, honoraria, stock ownership or options, expert testimony, grants or patents received or pending, or royalties.

Bibliography

1. Bayat A. "Science, medicine, and the future: Bioinformatics". *British Medical Journal* 324 (2002): 1018-1022.

2. Allen TC and Cagle PT. "Bioinformatics and Omics. In Springer, New York, NY (2008): 65-69.

3. Sindelar RD. “Genomics, other “omic” technologies, personalized medicine, and additional biotechnology-related techniques. In: Pharmaceutical Biotechnology: Fundamentals and Applications, Fourth Edition. Springer New York (2013): 179-221.
4. Merrill SA and Mazza A-M. “Innovation NRC (US) C on IPR in G and PR and. Genomics, Proteomics, and the Changing Research Environment” (2006).
5. Genomics | Medical Journals (2020).
6. RMS | First coronavirus was discovered in 1964.
7. Zhao X., *et al*. “A novel coronavirus from patients with pneumonia in China, 2019”. *The New England Journal of Medicine* 382.8 (2020).
8. Guo YR., *et al*. “The origin, transmission and clinical therapies on coronavirus disease 2019 (COVID-19) outbreak- A n update on the status”. *Military Medical Research* 7 (2020): 1-10.
9. Chen Y., *et al*. “Emerging coronaviruses: Genome structure, replication, and pathogenesis”. *Journal of Medical Virology* (2020).
10. Shereen MA., *et al*. “COVID-19 infection: Origin, transmission, and characteristics of human coronaviruses”. *Journal of Advanced Research* (2020): 24.
11. Covid-19 a Biological Perspective | by shashank Jain | Medium (2020).
12. Mutation, Repair and Recombination - Genomes - NCBI Bookshelf (2020).
13. Nakagawa K., *et al*. “Viral and Cellular mRNA Translation in Coronavirus-Infected Cells. In: Advances in Virus Research”. Academic Press Inc (2016): 165-192.
14. Young BE., *et al*. “Effects of a major deletion in the SARS-CoV-2 genome on the severity of infection and the inflammatory response: an observational cohort study”. *Lancet* 396.10251 (2020): 603-611.
15. Zhou P., *et al*. “A pneumonia outbreak associated with a new coronavirus of probable bat origin”. *Nature* 579.7798 (2020): 270-273.
16. Walls AC., *et al*. “Structure, Function, and Antigenicity of the SARS-CoV-2 Spike Glycoprotein”. *Cell* (2020).
17. Manikyam HK and Joshi SK. “Whole Genome Analysis and Targeted Drug Discovery Using Computational Methods and High Throughput Screening Tools for Emerged Novel Coronavirus (2019-nCoV)”. *International Journal of Pharmaceutical Sciences and Drug Research* 3.2 (2020): 341-361.
18. Esteban DJ., *et al*. “New bioinformatics tools for viral genome analyses at Viral Bioinformatics - Canada”. *Pharmacogenomics* 6.3 (2005): 271-280.
19. CoVDB: a comprehensive database for comparative analysis of coronavirus genes and genomes (2020).
20. SPHERES | CDC (2020).
21. covid-19-data/public/data at master · owid/covid-19-data”· GitHub (2020).

22. Korber B., *et al*. “Tracking Changes in SARS-CoV-2 Spike: Evidence that D614G Increases Infectivity of the COVID-19 Virus”. *Cell* (2020).

23. Wang JT., *et al*. “The role of phylogenetic analysis in clarifying the infection source of a COVID-19 patient”. *Journal of Infection* (2020).

24. Wertheim JO., *et al*. “A Case for the Ancient Origin of Coronaviruses”. *Journal of Virology* (2013).

25. Coronavirus (COVID-19) Testing | Lab Tests Online (2020).

26. Corman VM., *et al*. “Detection of 2019 novel coronavirus (2019-nCoV) by real-time RT-PCR”. *Eurosurveillance* 25.3 (2020).

27. Plasma therapy may look promising in treating COVID-19, but it is no magic bullet - The Hindu (2020).

28. Inspired by llamas, scientists make potent anti-coronavirus agent - STAT (2020).

29. Wang N., *et al*. “Subunit Vaccines Against Emerging Pathogenic Human Coronaviruses”. *Frontiers in Microbiology* 11 (2020).

30. bio.tools · Bioinformatics Tools and Services Discovery Portal (2020).

31. Grønning AGB., *et al*. “DeepCLIP: predicting the effect of mutations on protein–RNA binding with deep learning”. *Nucleic Acids Research* 48.13 (2020): 7099-7118.

32. Piñero J., *et al*. “DisGeNET: A comprehensive platform integrating information on human disease-associated genes and variants”. *Nucleic Acids Research* 45.1 (2017): D833-839.

33. Ahmed SF., *et al*. “COVIDep: a web-based platform for real-time reporting of vaccine target recommendations for SARS-CoV-2”. *Nature Protocols* (2020).

34. CoV2ID: Detection and Therapeutics Oligo Database for SARS-CoV-2 | BioData (2020).

35. Corman VM., *et al*. “Detection of 2019 novel coronavirus (2019-nCoV) by real-time RT-PCR”. *Eurosurveillance* 25.3 (2020).

Volume 10 Issue 5 May 2021