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.
Essential interpretations of bioinformatics in COVID-19 pandemic

Manisha Ray, Mukund Namdev Sable, Saurav Sarkar, Vinaykumar Hallur

ARTICLE INFO
Keywords:
SARS-CoV-2
COVID-19
Bioinformatics
Next generation sequencing
Genome wide association study
Drug design

ABSTRACT
The currently emerging pathogen SARS-CoV-2 has produced the global pandemic crisis by causing COVID-19. The unique and novel genetic makeup of SARS-CoV-2 has created hurdles in biological research, due to which the potential drug/vaccine candidates have not yet been discovered by the scientific community. Meanwhile, the advantages of bioinformatics in viral research had created a milestone since last few decades. The exploitation of bioinformatics tools and techniques has successfully interpreted this viral genomics architecture. Some major in silico studies involving next-generation sequencing, genome-wide association studies, computer-aided drug design etc. have been effectively applied in COVID-19 research methodologies and discovered novel information on SARS-CoV-2 in several ways. Nowadays the implementation of in silico studies in COVID-19 research has not only sequenced the SARS-CoV-2 genome but also properly analyzed the sequencing errors, evolutionary relationship, genetic variations, putative drug candidates against SARS-CoV-2 viral genes etc. within a very short time period. These would be very needful towards further research on COVID-19 pandemic and essential for vaccine development against SARS-CoV-2 which will save public health.

1. Introduction
Due to the small genome size, viruses have complex methods to maximize the coding potential of genomes and evaluation (Gautam et al., 2019). Meanwhile, the introduction of genomics and bioinformatics have contributed enormously to understand the infectious disease from disease pathogenesis, mechanisms and the spread of antimicrobial resistance to host immune responses (Buh et al., 2018).

SARS-CoV-2, which has created world pandemic scenario by affecting not only public health but also the socio-economic status of the entire humankind. The genome of the novel severe acute respiratory syndrome 2 (SARS-CoV-2) has been observed to be between 29.8 kb to 29.9 kb in size, and its sequence differs substantially from some of the previously identified human corona viruses including SARS and the Middle East respiratory syndrome (MERS) (Khailany et al., 2020; Chaw et al., 2020). However, the proper investigation of epidemiological, virological and pathogenic characteristics of SARS-CoV-2 is crucial to introduce novel treatment approaches and to develop effective prevention strategies (Messina et al., 2020). For the above bioinformatics tools and techniques have been implemented.

2. Next-generation sequencing
Advances in Next-Generation sequencing (NGS) innovations have brought about a remarkable multiplication of genomic sequence data (Suwinski et al., 2019). NGS has revolutionized the scale and deepness of biomedical sciences. During an outbreak condition in a health care system, the fast and effective identification of causative pathogen with epidemiological surveys are needed to permit a focused on disease control reaction. The accuracy of NGS in viral variants has productively analyzed and quantify the extremely high diversity within viral quasi-species. Many low frequency discovered drug or vaccine resistant mutations of therapeutic importance (Lu et al., 2020). High throughput sequencing technologies, including whole-genome sequencing (WGS) metagenomics technique, are providing the possibility to rapidly obtain the full sequence of pathogen genomes.

2.1. Metagenomics
The in silico virus sequencing is often based on alignments mapping of reads against a reference sequence (Maurier et al., 2019). Whereas a simple, cost-effective approach metagenomics is the only approach,
which does not require reference sequence for analysis. It represents a powerful application for pathogen identification from the environmental samples and directly accessing the genetic content of the organism during emerging pandemics situations (Peddu et al., 2020; Thomas et al., 2012). Metagenomics applications have also introduced in recent COVID-19 pandemics to reveal some critical novel information regarding SARS-CoV-2. The metagenomics has been used for rapid identification and quick characterization of the first few cases of COVID-19 (Chen et al., 2020; Manning et al., 2020), for examining the identification and quick characterization of the first few cases of WNau sea food market.

### Table 2

| Applications useful in COVID19 Next Generation Sequencing Data Analysis (Meta Genomics and Whole Genome Sequencing). |
|---|
| **Table 2** Application of metagenomics in different experimental studies on SARS-CoV-2. |
| **Author and publication year** | **Objectives of the study** | **Sequencing platform** | **Findings** |
| **Peddu et al., 2020** | Studied on SARS-CoV-2 epidemic, laboratory-confirmed positive and negative samples from Seattle, Washington | Illumina MiSeq | • Betacoronavirus of Bats are the closely related species of SARS-CoV-2 |
| **Chen et al., 2020** | Investigated two pneumonia patients who developed acute respiratory syndromes after independent contact history with Wuhan sea food market | Illumina Miseq | • Colonization with human parainfluenza virus 3 with SARS-CoV-2 |
| **Manning et al., 2020** | Quick characterization of Cambodia’s first case of COVID-19 | iSeq100 | • 2019-nCoV was closely related to strains bat-SL-CoVZXC21 and bat-SL-CoVZC45 at ORF1a, S, and N genes |
| **Van Tan et al., 2020** | Isolation of other pathogen co-infections in people with COVID-19 | Illumina MiSeq | • Identified presence of SARS-CoV-2 from pneumonia patients |
| **Tsan-Yuk-Lam et al., 2020** | Identification of any intermediate host for SARS-CoV-2 infection transmission to human | Illumina HiSeq | • No other pathogens were identified from the infected sample |
| **Wahba et al., 2020** | Examined close matches to the severe acute respiratory syndrome coronavirus 2 and its association with longitudinal faecal microbiome alterations in patients with COVID-19 | NA | • All human SARS-CoV-2 genomes are very similar, including the SARS-CoV-2 genome from the Cambodian case |
| **Zuo et al., 2020** | | Illumina NextSeq 550 | • SNP was noted at position 25,654 in ORF3a resulting in a valine-to-leucine substitution |

### Table 2

| **Sequences Read Archive (SRA) Database** | It is the largest publicly available repository of high throughput sequencing data, stores raw sequencing data and alignment information. | Leinonen et al., 2011a,b |
| **European Nucleotide Archive (ENA)** | Provides a comprehensive record on DNA and RNA raw sequencing and assembly data. | Leinonen et al., 2011a,b |
| **Metagenomics** | | |
| **FastQC** | Used to check quality control on raw sequences generated from high throughput sequencing pipelines. | Brown et al., 2017 |
| **Cutadapt** | Used to clean the sequences. It finds and removes adapter sequences, primers, poly-A tails and other types of unwanted sequence from the high-throughput sequencing reads. | Martin, 2011 |
| **Qime** | An open-source bioinformatics pipeline for performing microbiome analysis from raw DNA sequencing data. It interprets demultiplexing and quality filtering, OTU picking, taxonomic assignment, and phylogenetic reconstruction, and diversity analyses and visualizations through command lines. | Kuczynski et al., 2011 |
| **Whole genome sequencing** | | |
| **FastQC** | Used to check quality control on raw sequences generated from high throughput sequencing pipelines. | Brown et al., 2017 |
| **Cutadapt** | Used to clean the sequences. It finds and removes adapter sequences, primers, poly-A tails and other types of unwanted sequence from the high-throughput sequencing reads. | Martin, 2011 |
| **MaSuRCA** | Genome Assembler | Zimin et al., 2013 |
| **Ragout** | A reference assisted assembly tool. Records contigs to create high quality scaffolds by using a genome rearrangement approach and multiple closely related genome references as a guide. | Kolmogorov et al., 2014 |
| **Prokka** | Rapid annotation of prokaryotic genomes. | Seemann, 2014 |
| **AUGUSTUS** | A tool to predict genes in eukaryote genome sequences. | Stanke and Morgenstern, 2005 |
3. Genome-wide association study

GWAS has rehabilitated the complex disease genetics in to modest by providing various convincing links between complex characteristics of human and disease. Comprehensive and accurate detection of variants from whole-genome sequencing is a definite prerequisite for translational genomic research (Hwang et al., 2019). GWAS has involved in the screening of genetic variants across the genomes of many individuals to identify genotype-phenotype associations. Genetic variants discovered by GWAS are used to identify individuals at high risk of deadly diseases, which influences the early detection and prevention of diseases (Tam et al., 2019).

A genome wide association study (GWAS) is an extensive genetic analysis of the disease-associated observable alleles in the host.

### Table 3

| Author and Publication Year | Objectives of the Study | Platform | Findings |
|----------------------------|-------------------------|----------|----------|
| Sah et al., 2020           | Whole genome sequencing of SARS-CoV-2 specimen isolated from COVID-19 patients of Nepal | Illumina miSeq | Identical sequence between BetaCoV/Nepal/61/2020 and 2019-nCoV WHU01 |
| Yadav et al., 2020         | Characterization of SARS-CoV-2 sequences isolated from India with travel history of China | Illumina miniseq | Silent mutations at coding region of Spike, ORF1a, ORF1b and ORF3b proteins |
| Sekizuka et al., 2020      | Characterization of SARS-CoV-2 genome, isolated from Japan with travel history of Egypt | Illumina | Sequence heterogeneity with in SARS-CoV-2 globally |
| Chong et al., 2020         | Whole genome sequencing and analysis of SARS-CoV-2 isolated from Malaysia | Illumina iseq | Mutations in Spike protein |
| Caly et al., 2020          | To describe the first isolation and sequencing of SARS-CoV-2 in Australia and rapid sharing of the isolate | Oxford Nanopore Technologies and Illumina short-read | B and T cell epitope prediction on Spike protein |

### 2.2. Whole genome sequencing

Obtaining virus genome sequence directly from clinical samples is still a challenging task due to the low load of virus genetic material compared to the host DNA and the difficulty to get an accurate genome assembly (Maurizier et al., 2019). By the time genome sequencing procedure of virus has become a convenient method for better understanding of virus pathogenicity and epidemiological surveillance. Whole-genome sequencing (WGS) is a potent implement for studying virus evolution and genetic association to diseases or for tracking outbreaks. The depth of the sequencing data and the quality of the obtained sequences make this approach particularly efficient in this context (Kremer et al., 2017).

For the early understanding and diagnosis of COVID-19, the whole genome sequencing of SARS-CoV-2 was done for the samples collected from different countries throughout the world by using NGS platforms like Illumina miseq, Roche etc. (Sah et al., 2020; Yadav et al., 2020; Sekizuka et al., 2020; Chong et al., 2020; Caly et al., 2020) (Table 2).

The use of nanopore sequencing is used for genome sequencing of SARS-CoV-2 (Caly et al., 2020) (Table 3). The available whole genome sequences of SARS-CoV-2 in various online databases, and data analysis software provides insights into the further genomic data analysis to offer better medications to the patients (Table 2).

### Table 4

| Author and Publication Year | Objective | Findings |
|----------------------------|-----------|----------|
| Khailany et al., 2020      | Understand the genomic structure and variations in SARS-CoV-2 complete genome sequences | 116 mutations found |
| Ellinghaus et al., 2020    | Identification of potential genetic factors involved in the development of Covid-19 | 3 most common mutations: 8782C > T in ORF1ab, 28,144 T > C in ORF8 and 29095C > T in N gene |
| Aiewsakun et al., 2020     | Identification of Genetic variation associated with COVID-19 severity | Analyzed 8,582,968 SNPs |
| Ray et al., 2020b           | Elucidation of Nucleotide polymorphisms in whole genome sequences of SARS-CoV-2 | A3p2131 gene cluster as a genetic susceptibility locus in COVID-19 patients |
| Tabbradeh et al., 2020      | Investigate and track SARS-CoV-2 in Iranian COVID-19 patients | Potential involvement of ABO blood group |
| Satpathy, 2020             | Investigation on source of origin of this novel coronavirus | Nucleotide variation at genomic position 11,083 |
| Joshi and Paul, 2020        | Highlight the similarities and changes observed in the submitted Indian viral strains | Variation in 11083G in symptomatic patients |
| Zhou et al., 2020          | Analyse the evolution and variation of SARS-CoV-2 during the epidemic starting at the end of 2019 | 11,083 T variant in asymptomatic patient |
| Lopes et al., 2020         | Investigate bats and pangolin as hosts in SARS-CoV-2 cross-species transmission | mir-485-3p, mir-539-3p, mir-3149 differentially target the variants |

**GWAS**
segments in viral genome (Lanza et al., 2020; Lopez-Rincon et al., 2020; Ellinghaus et al., 2020; Aiewsakun et al., 2020; Ray et al., 2020a) diversity analysis with phylogenetic analysis has been frequently used in quantifying rare viral variants within the species (Khailany et al., 2020; Forsythia; fructus and Isatidis radix herbs are widely used for treating Covid-19)

structure and alterations, primer design etc. have represented novel alignment, genetic/nucleotide variations in the form of SNPs, genomic demography of SARS-CoV-2 globally (Ramírez et al., 2020; Fang et al., 2020) the SARS-CoV-2 research analyses to study the evolution and population structure of the nucleocapsid protein for inhibit viral replication in SARS-CoV-2 study. This provides additional data for proper genomic assessment of SARS-CoV-2 (Ray et al., 2020b; Tabibzadeh et al., 2020; Satpathy, 2020; Joshi and Paul, 2020; Zhou et al., 2020; Lopes et al., 2020) (Table 4).

Also to prevent the false positive results during testing of COVID-19 through real-time polymerase chain reaction (rtPCR) and decreasing the need for standardization across different PCR protocols, some primers have been designed through in silico algorithms by targeting conserved segments in viral genome (Lanza et al., 2020; Lopez-Rincon et al., 2020; Toms et al., 2020). This generated novel information on SARS-CoV-2 infectious genes are helping the researchers in the vaccine development against SARS-CoV-2, according to the identified viral genes coding regions, genetic sequence variations and molecular differentiations between the isolated species throughout the world. All the reported genomic experiments and analyses including SNP study, phylogenetic analysis, primer designing etc. have been carried out through high throughput bioinformatics tools and techniques which provide an appropriate pipeline for data analyses and annotations (Table 5).

4. Computer aided drug design

Drug design is very challenging, expensive, time consuming and an integrated rising discipline (Bisht and Singh, 2019). In the interim, the field of bioinformatics has become a crucial part of the drug design that plays a vital role for the validation of drug targets. It can help in the understanding of complex biological processes to improve drug discovery (Choudhry and Saikia, 2018). The in silico screening or computer-aided drug design (CADD) has signified as a dominant practice because of its proper algorithms including the development of digital repositories for the study of chemical interaction relationships, computer programs for designing compounds with unusual physicochemical characteristics as well as tools for systematic assessment of potential lead candidates etc. in drug discovery and development (Song et al., 2009). Also, the additional benefits like cost-saving, time to market, in-sight knowledge of drug-receptor interaction, speed up in drug discovery and development increases its popularity in scientific researches (Ramirez et al., 2020).
SARS-CoV-2 (Hall Jr and Ji, 2020; Fantini et al., 2020; BR et al., 2020; some previously used ancient synthetic drugs with antiviral activities bioinformatics in a single flow diagram (Fig. 1).

end CADD have interconnected and represented the applications of predicted novel putative natural inhibitors but also re-experimented a task in a sequential manner. From the beginning metagenomics to the completion of CADD, various bioinformatics tools and databases have been used since last decades and would be used in further research (Table 7).

Panda et al., 2020; Ray et al., 2020b) (Table 6). For the successful completion of CADD, various bioinformatics tools and databases have been used since last decades and would be used in further research (Table 7).

Huang et al., 2007

Discovery of Single Nucleotide Polymorphisms dbSNP (https://www.ncbi.nlm.nih.gov/snp/)

A crucial repository for each single base nucleotide substitutions and quick deletion and insertion polymorphisms

Sherry et al., 2001

Phylogenetic Analysis

MEGA (Molecular Evolutionary Genetics Analysis) (https://www.megasoftware.net/)

Multiple sequence alignment, phylogenetic tree generation and statistical analyses.

Kumar et al., 2008

PHYLOGENY.fr (https://www.phylogeny.fr/)

Reconstruct and analyse phylogenetic relationships between molecular sequences.

Dereeper et al., 2008

PAUP (https://www.megasoftware.net/)

Consensus classifier for prediction of disease related amino acid mutations.

Rath et al., 2020

PolyPhen2 (https://genetics.bwh.harvard.edu/pph2/)

Predicts possible impact of an amino acid substitution on the structure and function of a human protein using straightforward physical and comparative considerations.

Ray et al., 2019

PROVEAN (http://provean.jcvi.org/index.php)

Predicts impact of an amino acid substitution or indel on the biological function of a protein.

Ray et al., 2019

SNAP2 (https://rostlab.org/services/snap/)

Predicts functional effects of sequence variants.

Ray et al., 2019

UCSC genome Browser (https://genome.ucsc.edu/)

A broad collection of vertebrate and model organism assemblies and annotations, along with a large suite of tools for viewing, analyzing and downloading genomic data.

Karolchik et al., 2009

NCBI Gene database (https://www.ncbi.nlm.nih.gov/gene/)

Resource of protein sequence and functional information

UniProt Consortium, 2008

_NEAREST (http://www.ncbi.nlm.nih.gov/blast/)

Conserved domain search through multiple and pair wise sequence alignments.

Ray et al., 2020a

DAVID (Database for Annotation, Visualization and Integrated Discovery)

Functional annotation of genes (Biological process, Molecular function, Cellular component)

Huang et al., 2007

Kegg (Kyoto Encyclopaedia of Genes and Genome)

Metabolic pathway analysis

Kanehisa and Goto, 2000

5. Limitations

Wide application of robust algorithm based tools and information perceived from several public repository have enriched the knowledge spheres of modern life science research. The available bioinformatics tools and techniques are simple, accurate, cost effective, economical and freely available on internet, enabling their universal use for different research purposes. The above mentioned online repositories including PDB, PubChem, DrugBank, NCBI gene/genome databases, UCSC genome database, Uniprot, dbSNP, GEO, SRA, ENA (Table 2), (Table 5), (Table 7) etc. have updated frequently with huge novel datasets, which provides much authenticated and useful information to the users to carry out their research purposes. However, there is some limitations in use of certain tools particularly used for drug design such as Modeller (Table 7) or any other software generated 3D structure of proteins is approximate, which needs to be properly validated through crystallographic method for further study. The analyzed docking parameters based on predefined algorithms of autodock (Table 7) should be simulated further to analyse the proper stability between target and drug candidate interactions. Likewise, some softwares including Schrodinger, Discovery studio (Table 7), PAUP (Table 5) etc. are creating limitations for researchers during data analysis and accession, as they are customized or paid software. Apart from the above major drawbacks/limitations some minor flaws are associated with the using of tools and software i.e. error during software installation, software dependencies

2020; Ray et al., 2020a; Bhowmik et al., 2020; Lavecchia and Fernandez, 2020), Envelop protein (Bhowmik et al., 2020), Main protease (M pro) (Prasanth et al., 2020; Cavasotto and Di Filippo, 2020; Vardhan and Sahoo, 2020; Panda et al., 2020; Kumar et al., 2020), 3CL protease (Hall Jr and Ji, 2020; Vardhan and Sahoo, 2020; Jo et al., 2020) of SARS-CoV-2 by using the bioinformatics tools and software (Table 6). This immediate and effective action has not only perceived from several public repository have enriched the knowledge spheres of modern life science research. The available bioinformatics tools and techniques are simple, accurate, cost effective, economical and freely available on internet, enabling their universal use for different research purposes. The above mentioned online repositories including PDB, PubChem, DrugBank, NCBI gene/genome databases, UCSC genome database, Uniprot, dbSNP, GEO, SRA, ENA (Table 2), (Table 5), (Table 7) etc. have updated frequently with huge novel datasets, which provides much authenticated and useful information to the users to carry out their research purposes. However, there is some limitations in use of certain tools particularly used for drug design such as Modeller (Table 7) or any other software generated 3D structure of proteins is approximate, which needs to be properly validated through crystallographic method for further study. The analyzed docking parameters based on predefined algorithms of autodock (Table 7) should be simulated further to analyse the proper stability between target and drug candidate interactions. Likewise, some softwares including Schrodinger, Discovery studio (Table 7), PAUP (Table 5) etc. are creating limitations for researchers during data analysis and accession, as they are customized or paid software. Apart from the above major drawbacks/limitations some minor flaws are associated with the using of tools and software i.e. error during software installation, software dependencies

Table 6

| Databases/ Tools | Application | References |
|------------------|-------------|------------|
| GEO (Gene Expression Omnibus) database (http://www.ncbi.nlm.nih.gov/geo/) | It is a repository of functional genomics data generated from experiments and stores curate gene expression profiles. | Cloough and Barrett, 2016 |
| NCBI Gene database (https://www.ncbi.nlm.nih.gov/gene/) | Repository of gene related information from a wide range of species. | Brown et al., 2015 |
| UCSC genome Browser (https://genome.ucsc.edu/) | Broad collection of vertebrate and model organism assemblies and annotations, along with a large suite of tools for viewing, analyzing and downloading genomic data. | Karolchik et al., 2009 |
| UniProt (https://www.uniprot.org/) | Resource of protein sequence and functional information | UniProt Consortium, 2008 |
| CD (Conserved Domain) Search (https://www.ncbi.nlm.nih.gov/Structure/cds/cdsearch.cgi) | Conserved domain search through multiple and pair wise sequence alignments. | Ray et al., 2020a |
| DAVID (Database for Annotation, Visualization and Integrated Discovery) | Functional annotation of genes (Biological process, Molecular function, Cellular component) | Huang et al., 2007 |
| KEGG (Kyoto Encyclopaedia of Genes and Genome) | Metabolic pathway analysis | Kanehisa and Goto, 2000 |

Discovery of Single Nucleotide Polymorphisms

dbSNP (https://www.ncbi.nlm.nih.gov/snp/)

A crucial repository for each single base nucleotide substitutions and quick deletion and insertion polymorphisms

Sherry et al., 2001

SIFT (https://sift.bi.a-star.edu.sg/)

Predicts effects of an amino acid substitution on protein function based on sequence homology and the physical properties of amino acids.

Sim et al., 2012

PredictSNP1 (https://loschmidt.chemi.muni.cz/predictsnp1/)

Consensus classifier for prediction of disease related amino acid mutations.

Rath et al., 2020

PredictSNP2 (https://loschmidt.chemi.muni.cz/predictsnp2/)

Platform for prediction of effects of SNPs in genomic region.

Bendl et al., 2016

PolyPhen2 (https://genetics.bwh.harvard.edu/pph2/)

Predicts possible impact of an amino acid substitution on the structure and function of a human protein using straightforward physical and comparative considerations.

Ray et al., 2019

PROVEAN (http://provean.jcvi.org/index.php)

Predicts impact of an amino acid substitution or indel on the biological function of a protein.

Ray et al., 2019

SNAP2 (https://rostlab.org/services/snap/)

Predicts functional effects of sequence variants.

Ray et al., 2019

PHYLOGENY.fr (https://www.phylogeny.fr/)

Reconstruct and analyse phylogenetic relationships between molecular sequences.

Dereeper et al., 2008

PAUP (https://www.megasoftware.net/)

Reconstruct and analyse phylogenetic relationships between molecular sequences using parsimony method.

Wilgenbusch and Swoford, 2003

Dnasp (http://www.ebi.ac.uk/dnasp/)

Analyse DNA polymorphisms using data from a single locus, and also generate haplotype diversity between the sequences.

Roaz et al., 2017

PopArt (http://popart.otago.ac.nz/index.shtml)

Population genetic software which visualizes haplotype diversity network.

Leigh and Bryant, 2015

Primer Design

Primer3 (https://bioinfo.ut.ee/primer3-0.4.0/)

Primer design, often in high-throughput genomics applications.

Untergasser et al., 2012

NCBI Primer Blast (https://www.ncbi.nlm.nih.gov/tools/primer-blast/)

Design new target-specific primers in one step as well as to check the specificity of pre-existing primers and also placing primers based on exon/intron locations and excluding single nucleotide polymorphism (SNP) sites in primers.

Ye et al., 2012

Meta Gene 27 (2021) 100844
particularity the type of operating systems, high speed internet network connection, high core computer facility etc. The designed tools and software are meant for respective analyses, the user cannot modify the software used for next generation sequencing analyses.

6. Future aspects

The observations on SARS-CoV-2 will be explored extensively through bioinformatics and its applications variously. The researchers can also elucidate the SNPs in host body after affected with COVID-19. According to the modified nucleotides/genes novel primers can be designed for polymerase chain reaction through computational primer design algorithms. Apart from the drug design, putative inhibitory peptide can be created against SARS-CoV-2 viral genes. These further ideas would exploit many more denovo information of SARS-CoV-2, which will help the clinicians to add novel medication insights in the diagnosis procedures.

7. Conclusion

The outbreak of COVID-19 throughout the world is a big challenge for people to overcome this. Advances in bioinformatics techniques have been proved as the most advanced and effective technique in biomedical research. The high throughput screening and accuracy of data analysis have made this possible. The vast utilization of computational approaches in the current pandemic situation has effectively used from the preliminary stage of viral sample identification to the end stage of drug design by discovering novel information on SARS-CoV-2 genomic contents, variations, diversity within the species and predicted potential drug/ vaccine candidates against the viral genes within a very short period. In the present economically down condition, the successfully implementation of bioinformatics approaches against SARS-CoV-2 is a great achievement for scientific community.

Funding

No funding has been received for this work.

Declaration of Competing Interest

The authors declare that they have no conflict of interest.
Eswar, N., Webb, B., Marti-Renom, M.A., Madhusudhan, M.S., Eramian, D., Shen, M.Y., Fantini, J., Di Scala, C., Chahinian, H., Yahi, N., 2020. Structural and molecular docking models of the SARS-CoV-2 spike glycoprotein and 3CL protease. Travel Med. Infect. Dis. 35, 101664.

Hsin, J., Arkhipov, A., Yin, Y., Stone, J.E., Schultz, K., 2008. Using VMD: an introductory tutorial. Curr. Protoc. Bioinformatics. https://doi.org/10.1002/0471295953.i309.306.

Huang, D.W., Sherman, B.T., Tan, Q., Kir, J., Liu, D., Bryant, D., Guo, Y., et al., 2007. DAVID bioinformatics resources: expanded annotation database and novel algorithms to better extract biology from large gene lists. Nucleic Acids Res. 35, W557–W561.

Hwang, K.B., Lee, I.H., Li, H., Won, D.G., Hernandez-Ferrer, C., Negron, J.A., Kong, S.W., 2019. Comparative analysis of whole-genome sequencing pipelines to minimize false negative findings. Sci. Rep. 9, 8276. https://doi.org/10.1038/s41598-019-42513-4.

Jo, S., Kim, S., Kim, D.Y., Kim, M.S., Shin, D.H., 2020. Flavonoids with inhibitory activity against SARS-CoV-2 3CLpro. J. Enzyme Inhibition Medicinal Chemistry 35 (1), 1539–1544.

Joshua, S., 2020. Phylogenetic analysis of the novel coronavirus reveals important variants in Indian strains. bioRxiv [Preprint]. https://doi.org/10.1101/2020.04.14.041301.

Kanehisa, M., Goto, S., 2000. KEGG: Kyoto encyclopedia of genes and genomes. Nucleic Acids Res. 28 (1), 27–30.

Karolchik, D., Hinrichs, A.S., Kent, W.J., 2009. The UCSC genome browser. Curr. Bioinformatics. 1, 4. https://doi.org/10.1016/j.binf.2010.05.006.

Kim, S., Thiessen, P.A., Cheng, T., Yu, B., Shoemaker, B.A., Wang, J., Bolton, E.E., Wang, Y., Bryant, S.H., 2016. Literature information in PubMed: associations between PubMed records and scientific articles. J. Cheminformatics 8, 32.

Kolmogorov, M., Ramey, B., Paten, B., Pham, S., 2014. Ragout—a reference-assisted assembly tool for bacterial genomes. Bioinformatics 30 (12), i302–i309.

Kremer, F.S., McBride, A.J.A., Pinto, L.S., 2017. Approaches for in silico finishing of microbial genome sequences. Genet. Mol. Biol. 40 (3), 553–576.

Kuczynska, J.A., Stougham, J.A., Walters, W.A., Gonzalez, A., Caprioli, J.G., Knight, R., 2011. Using QIME to analyze 16S rRNA gene sequences from microbial communities. Curr. Protoc. Bioinformatics 10, 10.7.

Kumar, S., Nei, M., Dudley, J., Tamura, K., 2008. MEGA: a biologist-centric software for evolutionary analysis of DNA and protein sequences. Brief. Bioinform. 9 (4), 299–306.

Kumar, Y., Singh, H., Patel, C.N., 2020. In silico prediction of potential inhibitors for the Main protease of SARS-CoV-2 using molecular docking and dynamics simulation based drug-repurposing. J. Infect. Public Health. https://doi.org/10.1016/j.jiph.2020.06.015.

Lam, T.T.Y., Shum, M.H.H., Zhu, H.C., Tong, Y.G., Xi, N.B., Liao, Y.S., et al., 2020. Identifying SARS-CoV-2 related coronaviruses in Malayang pangolins. Nature. https://doi.org/10.1038/s41586-020-2652-3.

Lanza, D.C.F., Lima, J.P.M.S., Jeronima, S.M.B., 2020. Design and in silico validation of polymerase chain reaction primers to detect severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2). Research Square [Preprint].

Lavecchia, M., Fernandez, J., 2020. In silico study of SARS-CoV-2 Nucleocapsid protein-protein interactions and potential candidates for their stabilization. [Preprint] 2020070558.

Leigh, J.W., Bryan, D., 2015. Popart: full-feature software for haplotype network construction. Methods Ecol. Evolut. 6, 1110–1116.

Leinonen, R., Sugawara, H., Shumway, M., et al., 2011a. The sequence read archive. Nucleic Acids Res. 39, D21-D29.

Leinonen, R., Akhtar, R., Birney, E., Bower, L., Cerdeno-Talavera, F., Cheng, Y., et al., 2011b. The European nucleotide archive. Nucleic Acids Res. 39, D93–D110.

Dereeper, A., Guignon, V., Blanch, G., Audic, S., Buffet, S., Chevenet, F., Dufayard, J.F., et al., 2008. Phylogenetic: robust phylogenetic analysis for the non-specialist. Bioinformatics 24 (14), 1682–1683.

Dellinghaus, D., Degenhardt, F., Bujanda, I., Buti, M., et al., 2020. Genomewide association study of severe Covid-19 with respiratory failure. NEJM. https://doi.org/10.1056/NEJMoa2002830.

Erwar, N., Webb, B., Mari-Renom, M.A., Madhusudhan, M.S., Ermanian, D., Shen, M.Y., Pieper, U., Sali, A., 2006. Comparative protein structure modeling using Modeller. Current Protocols Bioinformatics 5 (5.6). https://doi.org/10.1002/0471295953.bi050615.

Fang, B., Liu, Y., Xu, L., Li, Y., Yu, G., Xu, J., et al., 2020. Genome-wide data inferring the evolution and population demography of the novel pneumonia coronavirus (SARS-CoV-2). bioRxiv [Preprint]. https://doi.org/10.1101/2020.03.04.976662.

Fanetti, J., Di Scala, C., Chahinian, H., Yahi, N., 2020. Structural and molecular modeling studies reveal a new mechanism of action of chloroquine and hydroxychloroquine against SARS-CoV-2 infection. Int. J. Antimicrob. Agents 55 (5), 105960.

Forli, S., Huyer, R., Pique, M.E., Sanner, M.F., Goodsell, D.S., Olson, A.J., 2016. Computational protein docking and virtual drug screening with the AutoDock suite. Nat. Protoc. 11 (5), 905–919.

Gautam, A., Tiwari, A., Malik, Y.S., 2019. Bioinformatics applications in advancing animal virus research. Recent Adv. Anim. Virol. 6, 447–471.

Grondin, A., Zoete, V., 2011. SwissPDBViewer: small molecule docking web service on EDOQ DSS. Nucleic Acids Res. 39, W270–W277.

Gupta, M.K., Vemula, S., Donde, R., Gouda, G., Behera, L., Yadav, R., 2020. In-silico approaches to detect inhibitors of the human severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) spike glycoprotein channel. BioInfo. Struct. Dyn. 15, 02254.

Hall Jr, D.C., Ji, H.F., 2020. A search for medications to treat COVID-19 via in silico molecular docking models of the SARS-CoV-2 spike glycoprotein and 3CL protease. Travel Med. Infect. Dis. 35, 101664.
SARS-CoV-2 infection and bacterial or viral superinfection and colonization. Clin. Chem. 66 (7), 966–972.

Ray, M., Sarkar, S., Rath, S.N., Sable, M.N., 2020a. Elucidation of polymorphisms in emerging SARS-CoV-2. bioRxiv [preprint]. https://doi.org/10.1101/2020.07.22.215731.

Ray, M., Sarkar, S., Rath, S.N., 2020b. Druggability for COVID-19 – in silico discovery of potential drug compounds against Nucleocapsid (N) protein of SARS-CoV-2. ChemRxiv [preprint]. https://doi.org/10.26434/chemrxiv.12387290.v1.

Richard, A., Friensier, Jay I., Banks, Robert B., Murphy, Thomas A., Halgren, Jasna J., Friesner, Jay L., Phillips, J.C., Braun, R., Wang, W., Gumbart, J., Tajkhorshid, E., Villa, E., Chipot, C., Skeel, R.D., Kale, L., Schulten, K., 2005. Scalable molecular dynamics with NAMD. J. Comput. Chem. 26 (16), 1781–1802.

Ramirez, J.D., Muñoz, M., Hernández, C., Flores, C., Gomez, S., Rico, A., Pardo, I., Barros, E.C., Panix-Mondolfi, A.E., 2020. Genetic diversity among SARS-CoV-2 strains in South America may impact performance of molecular detection. Pathogens 9 (7), 560.

Rath, S.N., Ray, M., Patri, M., 2020. Computational discovery and assessment of non-synonymous single nucleotide polymorphisms from target gene pool associated with Parkinson’s disease. Gene Reports. https://doi.org/10.1016/j.genepr.2020.100947.

Ray, M., Mishra, J., Priyadarshini, A., Sahoo, S., 2019. In silico identification of potential drug target and analysis of effective single nucleotide polymorphisms for autism spectrum disorder. Gene Reports 16. https://doi.org/10.1016/j.genepr.2019.100420.

Ray, M., Sarkar, S., Rath, S.N., Sable, M.N., 2020a. Elucidation of polymorphisms in emerging SARS-CoV-2. bioRxiv [preprint]. https://doi.org/10.1101/2020.07.22.215731.

Ray, M., Sarkar, S., Rath, S.N., 2020b. Druggability for COVID-19 – in silico discovery of potential drug compounds against Nucleocapsid (N) protein of SARS-CoV-2. ChemRxiv [preprint]. https://doi.org/10.26434/chemrxiv.12387290.v1.

Richard, A., Friensier, Jay I., Banks, Robert B., Murphy, Thomas A., Halgren, Jasna J., Friesner, Jay L., Phillips, J.C., Braun, R., Wang, W., Gumbart, J., Tajkhorshid, E., Villa, E., Chipot, C., Skeel, R.D., Kale, L., Schulten, K., 2005. Scalable molecular dynamics with NAMD. J. Comput. Chem. 26 (16), 1781–1802.

Ramirez, J.D., Muñoz, M., Hernández, C., Flores, C., Gomez, S., Rico, A., Pardo, I., Barros, E.C., Panix-Mondolfi, A.E., 2020. Genetic diversity among SARS-CoV-2 strains in South America may impact performance of molecular detection. Pathogens 9 (7), 560.

Rath, S.N., Ray, M., Patri, M., 2020. Computational discovery and assessment of non-synonymous single nucleotide polymorphisms from target gene pool associated with Parkinson’s disease. Gene Reports. https://doi.org/10.1016/j.genepr.2020.100947.

Ray, M., Mishra, J., Priyadarshini, A., Sahoo, S., 2019. In silico identification of potential drug target and analysis of effective single nucleotide polymorphisms for autism spectrum disorder. Gene Reports 16. https://doi.org/10.1016/j.genepr.2019.100420.