Role of Bioinformatics, Computational Biology and Computer Technologies in Combating COVID-19 Virus-a Review

Santhosh Poojary

Assistant Professor, Biotechnology Engineering Department
NMAM Institute of Technology, (affiliated to VTU Belagavi) Nitte-Karkala,
Karnataka-India
drsanthoshnellikatte@gmail.com

Abstract — *The world is distressing to eradicate COVID-19 pandemic and at this point of time, the available medicine (repurposed drugs) and existing technology are the only first hand weapons for immediate exploration. The review paper discusses the role of Bioinformatics with software packages, biological databases to understand genetic traits, mutations of the virus, also the role of Computation biology techniques for realistic mathematic modelling, simulations and the generalized overview on Computer Technologies accessible from modern computer including supercomputer was highlighted with reference to combating the COVID-19 Virus. The development of target drugs and vaccine with Computer aided drug design (CADD) in line with Artificial intelligence (AI) and Cloud computing (CC) techniques was discussed. On the other hand the role of Internet of things (IOT), Big data analytics (BDA) for data exchange, processing and analysis was overviewed. Hence it is been realized that a multifaceted understanding of Bioinformatics, Computation Biology and Computer Technologies is the real need of the hour for this emergency situation to understand, analyze the Corona virus crisis as well as to develop Therapeutic drugs and Vaccines.*

Keywords— *COVID-19, Bioinformatics, Biological databases, Computation Biology, Artificial intelligence, Cloud computing, Internet of things, Big data analytics, Corona Virus, Therapeutic drugs, Vaccines.*

I. INTRODUCTION

The COVID-19 (Corona Virus Disease-19) is infecting millions of people and made billions of people to stay under lockdown. The invisible virus is threatening the entire human race and challenging the man made technological advancements. It is posing major burden to public, government and healthcare sectors. The COVID-19 causing virus is a novel virus and it resembles to earlier know Severe Acute Respiratory Syndrome Corona Virus (SARS-CoV) and Middle East Respiratory Syndrome Corona Virus (MERS-CoV), hence COVID-19 virus is been named as SARS-CoV2. Till 1968, these viruses are considered as uncategorized cold viruses and later the name Corona-Virus is been nomenclatured which is the word synonymous with Solar corona [1]. The deadliest effect from these viruses was seen in 2003 (SARS) and in 2012 (MERS) outbreaks and ongoing pandemic COVID-19 affecting millions of people as shown in the Table 1.

The corona virus (Fig. 1) is made up of genetic material ribonucleic acid (RNA) containing more than 30,000 genetic codes (nucleic acid bases) with more than 10 hereditary characters (genes) and it is covered by a nucleocapsid (N) protein. This RNA is referred as single stranded positive sense RNA (ssRNA) and it can also act as messenger RNA (mRNA) in the host cell where it can be directly translated into proteins by host ribosome [2]. All positive-sense ssRNA virus genomes encode RNA-dependent RNA polymerase (RdRP), a viral protein that synthesizes RNA from a RNA template. Recombination in RNA viruses is common adaptation for coping with genome damage [3]. These viruses possess an oily (fatty) cell membrane which is composed of membrane (M) protein along with they have protruding protein component in their outer surface called as Spike (S) protein which is a carbohydrate attached protein (glycoprotein) [4]. This protein helps for the virus to attach onto the infecting (host) cells through a specific receptor and in case of corona virus SARS COV-2 it is hACE-2 (human Angiotensin Converting Enzyme 2) and later it aid virus to combine (fusion) with host cell membrane. One of the trickier point is that even though the function, three dimensional structure of the spike protein is well studied this protein structure is not same for all the corona virus groups; hence the entry mechanism may vary. The COVID- 19 disease virus has special region in the protein structure (cleavage site) which distinguishes it from other SARS or MERS corona viruses. They also have envelop (E) protein component which plays a central role in virus movement (morphogenesis) and assembly. The entry of corona virus to susceptible cells is a complex process which requires specified action of receptor binding and proteolytic processing of the spike protein to promote the fusion of virus to the host cells [5].

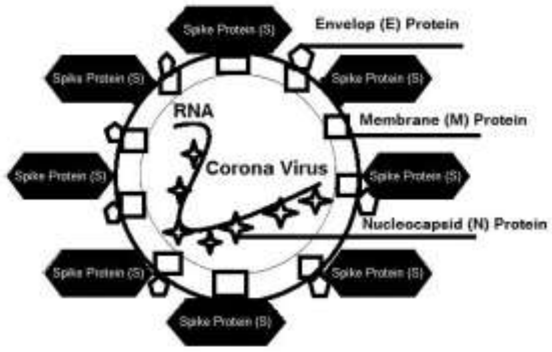

Fig. 1 A diagrammatic depiction of Corona Virus

Since the viruses are in the size of nano-meters and only grow in living cells, the enquiry of body fluid sample can be subjected virus specific identity test (antigen based), genetic material test (DNA or RNA) or antibody testing and virus genetic sequence is very much essential which is obtained by genomic sequencing techniques and Bioinformatics approach. Bioinformatics in drug development process utilizes chemical information of drug molecules, pharmacological aspects, in sillico modelling and computational biology techniques. The biological system environment is tried to virtually visualize in computer system. The simulation can start from a simple biochemical reaction to complex models for real organisms itself. The appropriate algorithms are getting developed continuously to understand the living systems.

At present moment, there is no vaccine or antiviral drug to prevent or treat human coronavirus infections and it is only supportive treatment with already known antiviral drugs (re-purposed drugs) [6]. The target drug development process is underway to discover preventative vaccine or a therapeutic prescription drug. A number of antiviral targets have been identified such as viral proteases, polymerases and entry proteins. Drugs are in development which targets these proteins and the different steps of viral replication. Globally hundreds of drug companies, biotechnology research firms, university research teams and health institutes are striving to develop more 150 vaccine candidates and over 300 potential therapies for COVID-19 disease in various stages of preclinical or clinical trial phase. A self replicating RNA vaccine is been recently developed by many modern tools entering into human clinical trials (Phase 1) which is giving lot of hopes but the effectiveness of such vaccines need to be assessed for long time because the another species of corona virus may create drastic fatality by the phenomena called as "Virus interference". Since the virus is changing its genetic characters (mutations), it is the necessity of advanced technology to develop broad spectrum drug or vaccine candidate. That is how the development of drug/ vaccine is challenging for these viruses. [7]

TABLE I

CORONA VIRUS OUTBREAK HISTORY [8]

Year	Out break Name	Virus Strain	Countries affected	People affected
2002	SARS	SARS-CoV	29	~8,000
2012	MERS	MERS -CoV	26	~1,600
2019	COVID-19	SARS-CoV2	>216 Including territories (As of June 2020)	>10 million (As of June 2020)

II. BIOINFORMATICS

Bioinformatics is an interdisciplinary biological science dealing with mathematics, statistics as well as computer science and information engineering. It helps to understand, integrate and analyze complex biological data by means of databases and software tools. It is been widely used in many field of investigation but pertaining to COVID-19, the genome analysis of the virus, diagnostic kit development, genetic variation (mutation) and predict target drugs the role of bioinformatics is immense. It can direct researchers in a systematic manner, beginning a hypothesis in the virtual environment and then proceeding to the real lab experiments.

There are two important aspects that bioinformatics can deal with is the Genomics and Proteomics of corona virus. In genomics the mapping and editing of viral genome is carried out whereas in proteomics the study was confined to viral protein analysis including amino acid sequencing, structural elucidation etc. A study on evolution history and genetic variation (phylogenomics) of corona virus by phylogenetic tree constructed by a computer is helping researchers to understand the origin, divergence in virus species etc. Already based on the phylogenetic basis the corona viruses are grouped into four catagories representing the initials names with greek alphabets as alpha, beta, gamma and delta corona viruses. The present SARS COV-2 belongs to the beta group [9].

A. Biological databases

The biotechnological techniques are already advanced for prominent separation and efficient detection of each building blocks (Nucliec acids of DNA and RNA) for hereditary (genomic) characters. The proper arrangement of these units (genetic sequencing) generated a whole lot of data since many decades. Nowadays these sequencing techniques have been developed from basic versions to next generation sequencing (NGS) with much more deep and comprehensive approach. The basic unit of protein (amino acids) sequencing study also took place subsequently. As a consequence the biological databases materialized in terms of large amount of nucleic acids (Adenine-A, Guanine-G, Thymine-T, Cytosine-C and Uracil- U) for DNA, RNA and proteinogenic twenty amino acids. The development of biological structure databases was also in pace with sequence databases. The Genbank was the first database of DNA and protein sequences and it is under the maintenance of the National Center for Biotechnology Information (NCBI) a part of National Institutes of Health (NIH). Along with Genbank, the primary databases which store original sequence data includes, DNA Data Bank of Japan (DDBJ), European Bioinformatics Institute (EMBL-EBI) European Nucleotide Archive (ENA) and Protein Data Bank (PDB) etc. Later the secondary databases such as InterPro, UniProt Knowledgebase (UniProtKB), Kyoto Encyclopedia of Genes and Genomes (KEGG), Ensembl, Prosite, TrEMBL and many more advanced hybrid databases creation took a dramatic shape and today thousands of sequence, gene expression, phenotype and structural databases are helping bioinformatics researches all around the world [10].

B. Bioinformatics software packages

The bioinformatics software packages can be considered as special tools available for retrieval of useful information (Data mining) and visualization of these computer readable data formats for our analysis. There are many specific famous tools most of the scientists use such as sequence search tools like BLAST (Basic Local Alignment Search Tool) or FASTA (FAST-All), sequence analysis tools like EMBOSS (European Molecular Biology Open Software Suite), sequence alignment tools like ClustalW, structure view tools like Rasmol, structure prediction tools like PROSPECT [11],[12].

Taking the benefit of these bioinformatics tools for COVID-19, recently researchers launched a web based software tool called Genome Detective for assembling virus genomes. It precisely assembles the existing virus genomes from advanced next generation sequence (NGS) datasets. This software compares viruses which caused huge disease outbreaks and this tool is been now effectively used for corona genome analysis within minutes. It classified all the SARS-corona viruses and made available genome data for public usage. The software tracks new viral genetic variations (mutations) which will definitely help for new diagnostics tool, target drug or vaccine development [13], [14].

III. COMPUTATIONAL BIOLOGY

The mathematical way of understanding theoretical and experimental aspects of biology is computational biology [15]. Broadly computational methods may deal with number of biological fields such as modelling biological systems, genomics, pharmacology, anatomy, evolution, neuroscience, neuropsychiatry, cancer biology etc. It uses many open source and proprietary software and tools with web, graphical as well as command line programs. As it was perceived that command line interface (CLI) bioinformatics tools usage was tiresome for many fundamental biologists, graphical platforms developed by computer scientists such as BioGUI was well appreciated [16]. Computational biology uses many features of Bioinformatics where in for COVID- 19, it may help to validate mathematical models, simulations leading to understand the biological system relationships. The drug target efficiency and molecular interaction with the virus can be predicted by computational biology approach.

IV. COMPUTER TECHNOLOGIES

A. Computer Aided Drug Design (CADD)

The new drug discovery and development process is long journey of investing money, time and efforts. This process is been now aided by computer technologies collectively called as computer aided drug design (CADD) and also known as In silico screening [17]. In case of COVID-19, the key protein molecules of the virus can be modelled to a three dimensional structure (homology modelling) and then the target drug molecule either screened from library of either re-purposed drug molecules or new one developed. Later the binding interaction of protein and drug molecule for stable complex formation (molecular docking) can be achieved. Assistance of computers is definitely needed for large number of repetitive biological experiments (High throughput experiments).

B. Artificial Intelligence (AI)

The ability of human intelligence to create artificial intelligence algorithms in computers and leading to machine learning is the greatest achievement in this century. The usage of deep learning techniques by the help of Artificial Neural Network (ANN) is also helping to build computer models that compare the X-ray crystallography structures of approved and experimental drugs to combat corona virus targeting its key proteins [18].The AI tool is helping to track the genetic variations of the corona virus as it circulates in large human populations. An interesting online platform was specifically designed for COVID-19 virus named as COVID-19 Genotyping Tool (CGT) using AI. Already more than 20,000 COVID-19 patients virus genome sequence was uploaded to Global Initiative on Sharing All Influenza Data (GISAID) COVID-19 database. In this software the hospitals can compare their corona virus sequences to the global stains sequences and analysis takes only minutes. By this quick AI based tool virus strain distribution and evolving characters can be assessed. The software developing team claims that when there is a re-occurrence for corona virus (second wave), the mutations can be immediately tracked which may be very much essential to re-battle with the virus [19].

C. Cloud Computing(CC)

Cloud computing technique deploys on demand availability of system resources and memory space over the internet (cloud storage) for computing purpose. In case of corona crisis, this cloud computing method helped to considerably accelerate the new target vaccine development as it was realized by Australian research team, where they collaborated Oracle-cloud based technology. They also utilized AI based advanced manufacturing techniques for vaccine design and it progressed towards animal testing and it is on the verge of human clinical trial [20].

D. Internet of Things (IOT)

The Internet of Things (IOT) is unified system of network connected devices through internet for collection and exchange of data, in most cases without the human intervention (information sensing devices). Widely used in security applications, traffic control, agricultural monitoring, weather forecast etc. In general it can be used to monitor or control (automation) of any electrical, electronic and mechanical systems.

The IOT is helping out COVID-19 pandemic in many ways, especially the deployment in healthcare sector. The foremost application came for quarantine tracking in many countries called as Geofencing where GPS, Wi-Fi, bluetooth, cellular network and many such wireless technologies used for tracking infected patients as well as susceptible peoples.

The “Arogya Setu” app is India’s useful IOT tool developed by Naional informatics Center for COVID-19 patient contact tracing, disease mapping and self assessment. One more interesting application ‘Bebot’, developed by Bespoke Inc., is an IOT system combined with AI chatbot which gives updated information about the pandemic in fingertips for residents and travellers. Even though there are some privacy concerns with some apps, more than 25 such applications with centralized and decentralized strategies for contact tracing is been already developed all over the world and it helped many people to get awareness about COVID-19 [21].

Recently to combat corona virus some robots are deployed for cleaning and disinfecting surfaces which are controlled by remote access applications using IOT technologies [22]. Drones with AI tools are also used for lockdown enforcement and monitoring, disinfection, medical samples transport and remote temperature monitoring in quarantine facilities [23].

To avoid in house viral infection spread, the IOT based smart devises can be used such as smart switches, door knobs etc. There was very big role which was played by our smart IOT devices during lockdown period for virtual communication with others through video conferencing platforms. The only care need to take using IOT is proper encryption techniques to avoid security and privacy breaches [24].

E. Big Data Analytics (BDA)

The Big Data Analytics (BDA) is the systematic analysis of large amount of complex data, when the traditional method of analysis shows its limitation for proper conclusion. There may be peta to exa byte of data containing billions to trillions data records. It is widely applied for data analysis in education, banking, insurance, agriculture, business intelligence, etc. This is been already used in healthcare sectors for analysis of large amount of patient data. During this COVID-19 crisis, lot of data gets generated in many aspects and it can be considered as “Big Data” and analysis of these data through high end computer technologies including AI will definitely yield valuable information. It can also predict the possible changes in our lifestyle, employment, economy and many more forecasted aspects of post- COVID-19 situations [25].

F. Supercomputer calculations

A supercomputer can be used for high end computational tasks deployed even in molecular modelling along with its wide application in Climate, Quantum science and many other technological fields. For COVID-19 pandemic, recently the IBM Corporation is participating in a “COVID-19 High Performance Computing Consortium” with its super computer platforms. As it is revealed by IBM, the scientists all around the world performing calculations in molecular modelling, bioinformatics and epidemiology (disease occurance, distribution and control) and for some calculations the regular computers may take months and if done manually it may take years, so supercomputers are definitely showing their promise [26]. There was one more effort from a supercomputer MOGON II (one of the most powerful supercomputer in the world-operated by Johannes Gutenberg University and Helmholtz Institute Mainz) to find out repurposed drugs for COVID-19 using drug screening, molecular docking and machine learning techniques [27].

V. CONCLUSIONS

Certainly the Bioinformatics, Computational Biology and Computer Technologies are helping us to understand COVID-19 pandemic in a scientific manner and also assisting to screen many potent repurposed drug molecules out of thousands of drugs available in the market. Even in the new drug development the role of Bioinformatics and Computation biology is far-fetched. The duration of developing new drug/vaccine may be longer but certainly the modern computers are making things possible which were thought to be unrealistic in earlier decades. This can be better visualized in the sense that, it is better to get some idea from computational models instead of mere guessing. The computer based methods accelerates the progression of scientific studies, enabling humans to imitate the biological systems and interact with virtual reality, instead of initially wasting time and money in wet lab experiments or trials. Along with many computer based digital technologies such as Computer aided drug design, Artificial intelligence, Cloud computing, Internet of things, Big data analytics are helping in COVID-19 combating race [28].

But even after having these many scientific advancements in hand why we are not able to quickly find a target drug for COVID- 19? The answer is simple, our technologies are at the verge of completely understanding the COVID-19 virus biology and also we are trying to recognize the tricky mechanisms this clever virus posing to us, which means we are still in the battle field exploring many weapons. It is realistic that, this is the war between viruses and humans and it continues; now it is just the beginning. If all human creatures think in the same direction for defeating these kinds of viruses in future by any means, the ahead of 21st century is for humans otherwise it is the Darwinian evolutionary theory coming into picture again, comprehending "Survival of the fittest".

REFERENCES

[1] Almeida JD, Berry DM, Cunningham CH, Hamre D, Hofstad MS, Mallucci L, McIntosh K and Tyrrell DA. "*Virology: Coronaviruses*". Nature, Vol 220, pp- 650, 1968.

[2] Ahlquist, Paul. "*RNA-dependent RNA polymerases, viruses, and RNA silencing*." Science, Vol. 296, pp-1270-1273, 2002.

[3] Barr, John N., and Rachel Fearns. "*How RNA viruses maintain their genome integrity*." Journal of General Virology, Vol 91, pp-1373-1387, 2010.

[4] Fehr, Anthony R, and Stanley Perlman. "*Coronaviruses: an overview of their replication and pathogenesis*." Methods in molecular biology (Clifton, N.J.) vol. 1282, 2015.

[5] Walls, Alexandra C., Young-Jun Park, M. Alejandra Tortorici, Abigail Wall, Andrew T. McGuire, and David Veesler. "*Structure, function, and antigenicity of the SARS-CoV-2 spike glycoprotein.*" Cell, 2020.

[6] Jean, Shio-Shin, and Po-Ren Hsueh. "*Old and re-purposed drugs for the treatment of COVID-19.*" Expert Review of Anti-infective Therapy just-accepted, 2020.

[7] (2020) New age opinion website [online] Available: https://www.newagebd.net/article/108023/rushing-a-vaccine-to-market-for-a-vanishing-virus

[8] Arora, Pooja, Mohammad Jafferany, Torello Lotti, Roxanna Sadoughifar, and Mohamad Goldust. "*Learning from history: Coronavirus outbreaks in the past*." Dermatologic Therapy: e13343, 2020.

[9] Woo, Patrick CY, Yi Huang, Susanna KP Lau, and Kwok-Yung Yuen. "*Coronavirus genomics and bioinformatics analysis.*" Viruses, Vo 2, pp-1804-1820, 2010.

[10] Letovsky, Stanley I., ed. *Bioinformatics: databases and systems*. Springer Science & Business Media, 2006.

[11] (2020), "*Bioinformatics software and tools*" Website [online] https://bioinformaticssoftwareandtools.co.in/index.php

[12] Ipsita Saha, Joy Dewanjee "*A web based nucleotide sequencing tool using BLAST algorithm*", International Journal of Biotech Trends and Technology (IJBTT), Vol. 6, pp- 1-4 Jul - Sep 2016.

[13] Cleemput, Sara, Wim Dumon, Vagner Fonseca, Wasim Abdool Karim, Marta Giovanetti, Luiz Carlos Alcantara, Koen Deforche, and Tulio De Oliveira. "*Genome Detective Coronavirus Typing Tool for rapid identification and characterization of novel coronavirus genomes*." Bioinformatics, Vol 36, pp-3552-3555, 2020.

[14] Vilsker, Michael, Yumna Moosa, Sam Nooij, Vagner Fonseca, Yoika Ghysens, Korneel Dumon, Raf Pauwels et al. "*Genome Detective: an automated system for virus identification from high-throughput sequencing data*." Bioinformatics, Vol 35, pp-871-873, 2019.

[15] Markowetz, Florian. "*All biology is computational biology*." PLoS biology Vol-15, e2002050, 2017.

[16] (2020) Biogui software website [Online] Available https://www.bio.ifi.lmu.de/software/biogui/index.html

[17] Bisht, Neema, and B. K. Singh. "*Role of computer aided drug design in drug development and drug discovery*." International Journal of Pharmaceutical Sciences and Research, Vol 9, pp-1405-1415, 2018.

[18] (2020) "*The COVID-19 High Performance Computing Consortium*" website , Available : https://covid19-hpc-consortium.org/

[19] Hassaan Maan, Hamza Mbareche, Amogelang R. Raphenya, Arinjay Banerjee, Jalees A. Nasir, Robert A. Kozak, and Bo Wang. "*COVID-19 Genotyping Tool*" (CGT) (Version v0.2.2). 2020.

[20] (2020) "*Flinders University*" Website, [online] Available https://news.flinders.edu.au/blog/2020/04/03/flinders-targets-covid-19-vaccine/

[21] (2020) Bespokes Chat Bot Website [online] Available: https://www.prnewswire.com/news-releases/bespokes-ai-chatbot-provides-assistance-during-current-coronavirus-pandemic-301000382.html

[22] G.-Z.Yang, B.J. Nelson, R.R. Murphy, H. Choset,H.Christensen, S. H. Collins, P.Dario, K.Goldberg, K.Ikuta,N.Jacobstein, D.Kragic, R.H. Taylor and M.McNutt, "*Combating COVID-19—The role of robotics in managing public health and infectious diseases*". Science Robotics, Vol. 5, eabb5589, 2020.

[23] Lu, Donna. "*Drones keep an eye on people failing to social distance*." New scientist (1971) vol. 246, pp-3282, 2020.

[24] (2020) IOT for All Website, writer Sushree Mishra, [online] Available: https://www.iotforall.com/the-growing-role-of-iot-in-covid-19-response/

[25] Wang, C. Jason, Chun Y. Ng, and Robert H. Brook. "*Response to COVID-19 in Taiwan: big data analytics, new technology, and proactive testing*." Jama Vol 323, pp-1341-1342, 2020.

[26] (2020) IBM Website, [online] Available https://newsroom.ibm.com/IBM-helps-bring-supercomputers-into-the-global-fight-against-COVID-19

[27] Kadioglu O, Saeed M, Johannes Greten H and Efferth T. "*Identification of novel compounds against three targets of SARS CoV-2 coronavirus by combined virtual screening and supervisedmachine learning*". [Preprint]. Bull World Health Organ. E-pub: 2020.

[28] Ting, Daniel Shu Wei, Lawrence Carin, Victor Dzau, and Tien Y. Wong. "*Digital technology and COVID-19" Nature Medicine"*, Vol 26, pp-459–461, 2020.