Inventum Biologicum, 2022, Vol. 2, Issue 1, pp. 13-23

Inventum Biologicum

Journal homepage: www.worldbiologica.com/ib

Review paper

Computational Genomic Sequencing for Covid-19

Jyoti Sarwan [a], Heenu Sharma [a], Anubha Sharma [a], Eleena Barik [a], Jagadeesh Chandra Bose K [a, *], Junaid Ahmad Malik [b]

[a] *University Institute of Biotechnology, Chandigarh University Gharuan 140413, India*
[b] *Department of Zoology, Govt. Degree College, Bijbehara, Kashmir 192124, India*

ARTICLE INFO

Article history

Received 22 January 2022
Revised 28 February 2022
Accepted 06 March 2022
Published 15 March 2022

Keywords

SARS-CoV-2
Bioinformatics
Drug design
Virus
Corona virus

ABSTRACT

SARS-CoV-2, a new virus belonging to the *Coronaviridae* family, has made itself worldwide attention seeker in the last two years owing to its exclusive infection and millions of deaths. Coronavirus has single –stranded RNA with 30 kb nucleotides with a positive sense in its genetic material. Although many years have been invested to study coronavirus still more research is pending. Therefore several tools have been invented called bioinformatics tools, specially designed to monitor and diagnose SARS-CoV-2 for fast detection and rapid reaction to treatment and understanding in its early stages. Following previous studies coronavirus RNA has enzyme *furin* that is found in organs like the small intestine, lungs, and liver of humans and is responsible for activating spike like proteins. However, coronavirus and associated enzymes can directly attack multiple organs and lead to organ failure in a small period. Therefore In *silico* studies can help to screen early stages of Covid-19 infections. In *silico* can provide data related to evolution, lineage, and drug resistance for COVID -19. In nutshell, the genomic sequencing tool is helpful to describe advanced research that is specifically for SARS-CoV-2 for its genomics, proteomics, early detections, rapid reactions, and drug discovery.

1. Background

The first case of Covid-19 was reported in Wuhan, China in 2019 by Municipal Health Commission on 31 December 2019. Earlier there were several cases in Wuhan reported with pneumonia and later they were declared acute coronavirus cases. Initially there were no evidences to claim acute respiratory syndrome, but with respect to the time researchers declared it Severe Acute Respiratory Syndrome Coronavirus, SARS-CoV-2. This virus was named as Covid-19 as a disease as it was reported in year 2019. In July 2021 numerous cases of coronavirus were reported (about 15 million) and nearly 600000 deaths. Coronviridae are viruses with prolonged RNA genomes with range of 33000 nucleotides in it (Wu et al., 2020). SARS-CoV-2 falls under Sarbecovirus sub genus including Betacoronavirus having nucleotides around 33000.

* **Corresponding author**: Jagadeesh Chandra Bose K, **E-mail**: jcboseuibtlab@gmail.com

DOI https://doi.org/10.5281/zenodo.6371393

In coronavirus main four types of proteins are present including spikes, membrane, envelop and nucleocapsid. Other different proteins found in coronavirus are pp1a and pp1ab polyproteins for some essential functioning of virus replications and other mechanisms. The main entry which is used by human SARS-CoV-2 and receptor ACE2 has been recognized by *Sarbecoviruses*. The human receptors have ability of binding at spike proteins which has cleavage site of novel *furin* in association with its transmission potential and enhances pathogenicity (Klenk and Garten, 1994).

2. Introduction

Some studies have shown that there are very lower rate of mutations in the corona virus as comparison to other RNA viruses but other kinds of mutations are also found as a result of genomic diversity of either individual or microbes. There are different criteria to enable different microorganisms in the host like genetic heterogeneity, environment in the host's body some diseases and drugs playing vital role in it. In this era of research, Covid 19 is an emerging topic to study for every researcher either belonging to health sciences or not. Although numerous studies are going on SARS-CoV-19 but still more is left to be resolved yet. It has always been critical to recognize pathogenesis of virus and its evolution simultaneously to control the pandemic on worldwide status. Therefore researchers around the world have been investigating SARS-CoV-19 on genomic and proteomics level including genomic sequencing, protein tracking, protein estimations, finding different epitopes, various protein reorganizations and so on but still some mysteries arises every time. Hence, a lot of researchers have been trying to get some informative data with the help of drug designing and epidemiological models and phylodynamic models too (Abrams et al., 2021). Numerous laboratories had shared the data of corona virus with unmatched speed and new findings. In this modern scientific era questions related to Covid-19 can be answered with bioinformatics by which the scenario of virus replication also can be determined. Bioinformatics is the study which is able to provid adequate determination and can boost several investigations for fundamental, applied sciences and public sector can be benefited also. SARS-CoV-19 has been declared a novel pathogen which is able to make entire world as a call of pandemic. This pandemic requires, first of all, to control its spreading chain and such resources and habits to breakdown chains to spread from lower level to higher level. Therefore the outbreak of virus is a challenging task to all pre-existing bioinformatics tools to understand and as well as to make them applicable to common public health and numerous researches in this critical phase. So in this review we discuss and cover major workflows involved in bioinformatics scenario and layout of tools starting from detection of symptoms of SARS-CoV-2, initiation of infection, detection of genomic sequences, tracking of virus pandemic, to study evolution of Covid-19 and drug discovery with therapeutic controls. Mostly used tool of bioinformatics have been designed either to study pathogenesis or evolution with drug discovery in favor to defeat SARS-CoV-19 (Barzkar et al., 2021).

2.1 Detection of SARS-CoV-2 in a patient

Once all the symptoms have been found in a patient of novel Coronavirus like mild fever, cough, cold, acute pneumonia and then the sample of a particular patient can be taken for further investigation and drug designs.

2.2 Detection of SARS-CoV-19 in a laboratory

The detection and analysis of SARS-CoV-19 can be done with Real-Time Quantitative Reverse Transcription PCR (qRT-PCR) where 'q' stands for quantitative detection in it to diagnose COVID-19. In this test there are two main components; first one is 2-nucleotide sequences that is virus gene envelop E, and second is RNA-dependent RNA polymerase, RdRp. The main feature of the test is to prove specificity for false negative and false positive of exclusion for validity of test. Earlier in numerous cases of Covid-19 the criteria to diagnose was false positive and negative detection and to replace this phenomena specific primers were required.

There was indeed a demand to design few specific primers related genomes of SARS-CoV-2 to encourage the sensitivity and specificity in diagnoses and test of qRT-PCR along with a high authenticity. Despite qRT-PCR has been playing a vital role from initial days of pandemic serving in public sector and analyzing the record of outbreak of corona virus all over the world. Therefore the shared data in Gene bank has provided genomic sequence of SARS-CoV-2

and it was also determined rapidly. Hence the annotation is required as similar as sequences of other corona virus. Therefore the next generation sequences can be defined as the sequencing that can assess the diversity of genomes situated in the virus. Therefore regular sequencing is also useful because it is able to detect mutations that can resist in RT-PCR tests. The study is completely relying on the host diversity and it is a challenging task of deep sequencing and analysis of all mutations in a particular host. Sometimes it is very difficult to resolve many viral strains with critical mutations in them. ARTIC network is a network that is designed to optimize and develop SARS-CoV-2. Nanopore sequencing protocol has been experienced for deploying the sequences of this particular network. This network is also helpful in new estimated sequencing and to maintain and surveillance of outbreaks of viral genomes (Bouckaert et al., 2019).

So the genomes of SARS-CoV-2 with high accuracy can be generated with Nanopore sequencing which also is quick provider and able to reduce time factor too. Similarly as high accuracy of genomes and time reduction takes place there is another advantage of this Nanopore sequencing which is track viral evolutionary patterns with respect to time and transmission of COVID- 19.

In addition of this, Nanopore based sequencing has an ability to identify and detect the pathogenic sequences and infection based sequences usually present along with transcriptomic or metagenomic sequencing based upon amplicon approaches. Samples present in clinical trials or in environment can also be identified with this approach, for example human virus like Bronchoalveolar lavage fluid (Madsen et al., 2021; Rambaut et al., 2020).

Before the era of pandemic it would not have been possible to design any tool or technique that detect and diagnose the evolutionary history and drug designing therapy for SARS-CoV-2 virus ancestors. The examination of divergence in virus sequencing of corona virus can be done with further metagenomics and recombination with other human viruses. Both approaches play a vital role to fight Covid-19 pandemic and are high quality SARS-CoV-2 genomic data and metadata in open access libraries without any resistance or restrictions. Global libraries situated all over the world have shared SARS-CoV-2 data with highest and non-comparable speed of genomic sequences on July 23, 2020. So researchers would be encouraged to submit database of sequencing of SARS-CoV-2 that does not have any limitations and NCBI offers researchers to submit new data of SARS-CoV-2 with new streamlined process.

Consequently a number of tools have been made for detection and annotation of SARS-CoV-2 infection but in this study we will discuss few of them that have wider applications and time saving processing. There are a few tools having major role in SARS-CoV-2 detection and evolutionary studies; VIRify, VBRC, VIRULIGN, V-Pipe etc. To detect differences between other virus and coronaviruses the comparative genomics are very helpful and pathogenesis or virulence strains also can be identified. During the study of viral genomes, both virulence and non-virulence as well as coding and non-coding sequences, every pattern is playing an important role to study evolution patterns of viruses. Coronavirus is a virus that has been found with un-translated regions of RNA secondary structures and non-coding sequences. These un-translated regions found in proteins of SARS-CoV-2 are responsible for mutations and evolution in coronaviruses. Before making any conclusion of evolutionary patterns, all sequences and molecular mechanisms of viruses have to be studied. The optimizations of viral samples extracted from the environment are based upon barcoding sequencing and its analysis. The most frequent used PCR in viral strains studies is to prepare complimentary primer for viral strain and choosing a template for new estimated sequence. The reference data will be matched to the regions present in amplicon or barcode regions for further investigations (Corman et al., 2020).

3. The Primer Search Tool or PriSet

It can be defined as the software that is able to identify chemically suitable primers to the viral genomic sequencing as reference data. Fasta file is defined as a file of either of complete genomes or set having short regions of reference data. SARS-CoV-2 is a test that totally depends upon diagnosing sample of mucus of patient that is secreted from nose, throat and mouth which is further used for barcoding and metabarcoding testing. RT PCR is required for analysis of DNA amplification and straight DNA samples are more prominent to primer sequences.

Some studies show that *in silico* transcripts of primers of SARS-CoV-2 genomes are more compatible with RT PCR settings. Data stored in NCBI also shows that there is no evidence of reoccurrence of primers of SARS-CoV-2. All the data related to SARS-CoV-2 is available on researchgate including genomic sequences of SARS-CoV-2 and primers. Specific primers are required in the SARS-CoV-2 RT-PCR test. The primers designed to use in RT PCR are made according to make compatibility with SARS-CoV-2 genomes to serve as special identification of specific primers and try to avoid false negative and false positive identifications.

4. Amplicon based Genomic Construction or CoVPipe

CoVPipe is a tool that is typically reference based and fully automated with SARS-CoV-2 genomic sequences. The data reference based data and work flow used in amplicon is also based on next generation amplicon sequences and it is used in Hayword and Paragon genomic studies and database. The pipeline is design to remove all terminal primer sequences in PCR and to eliminate adapters of low quality during sequencing test in RT PCR. This sequencing is typically reference based and processed to align BWA-MEM data. Evaluation of BEM data helps to report mapping and measurements like read depth, insert size and coverage etc. The workflow of pipeline is implemented by using Snakemake which is able to balance and control the input and output mechanisms during testing in software.

5. Rapid Sample Analysis for Nanopore Sequencing- poreCov

This is a tool or software that has already been used to diagnose diseases like zika, ebola, yellow fever, swine flu and viral genomic database. So it is a time saving protocol and within approximately less than 7 hours we can get all genomic data of virulent genes enabling us to manage and compile up with other bioinfomatic tools.

6. VADR and Vpipe for SARS-CoV-2 Genome Annotation and Validation

- Vadr is a sequence based model usually build to take data used from reference sequence and there are different entries that revealed it is a powerful tool with future perspectives and able to read ORF polyproteins and provides data for further drug discovery investigations.
- V pipe is a tool that is a boon of bioinformatics and it is defined as a tool or software which is able to analyze data of virus having high sequencing of numerous elongated repetitive patterns in a genome. It is a software used not only for analysis of the elongated sequences but also helps to read mapping including intra host data quality control. The alignment of viral genes can also be done with the help of these VPipe softwares.

7. Haploflow: Denovo Assemply to Make Awareness of Multi Strain

A complete viral infection is made up of a large number of distinct types of viruses. So in viral infection not only one strain is able to cause the infection but variants of viral strains combine to cause viral infection or diseases. Therefore presence of numerous viral strains in a host body can create mutations or evolution and every host has a unique genomic pattern based upon either heredity or environment. As a result, strain-specific genomic features have been identified in a collection of viral strains (El-Gebali et al., 2019). The actual criteria or process of evolution of viral strains are yet not known but several researches are still continued in it. One major conclusion has been come out that SARS-CoV-2 has an ability to cause several infections and co-infections in host body and higher risks to reactions with other coronaviruses at a specific time span. The main struggle during dealing with SARS-CoV-2 is to resolve complete viral haplotypes and it becomes a critical task in few cases of therapies (Di Renzo et al., 2020). Haploflow is a technique in which graphical based data assembly takes place in the detection method of either viral infection of viral strain. It is a rapid graph based technique and can resolve issues like based pair detection and combined two viral strains. Reconstructions and full length of SARS-CoV-2 strains and multiple strain infections can be obtained by Haploflow.

8. VIRify: Annotation of Viruses in Meta-omic Data

VIRify is a newly created general process for detecting, annotating and classifying viral and phage

contiguous in the fields of metagenomic and metatranscriptomic for a common purpose. This execution is a part of analytical services provided by MGnify (Mitchell et al., 2020). Taxonomic categorization of VIRify depends on the identification of a profile which is taxon specific known as Hidden Markov Models (HMMs) that are constructed on a set of 22,014 different orthologous protein fields known as ViPhOGs. This collection of HMM database contains 139 various models that acts as particular markers for species in the *Coronaviridae* family. This demonstrates the usefulness of the VIRify pipeline, utilized in the process of isolation from MGnify, for investigating coronaviruses in the human respiratory microbiome. Coronaviruses in the clinical as well as in the environmental samples can be identified by VIRify. Because of the inherent distinctions between metatrascriptomes and metagenomes, extra considerations for quality control, assembly, post processing and categorization must be made.

9. Genome Analysis Tools by VBRC

The Viral Bioinformatics Research Centre (VBRC) is an established resource designed particularly for virologists to aid in the analysis by comparing the viral genome. A MySQL database built from GenBank files within VBRC offers a variety of analytic tools. Virus Orthologous Clusters (Wang et al., 2020), a sophisticated yet user-friendly database GUI, is used to access the curated database. Base-by-base (Nawrocki and Eddy, 2013) is a program that can generate, visualize and modify numerous sequence alignments. It can use alignments and graphs to compare genomes, genes and proteins. Users can remark on sequences and save alignments to their own computer. Viral Genome Organizer (Upton et al., 2000) visualizes and analyzes the gene arrangement in several entire viral genomes. The user may transfer sequences of proteins or DNA and view START/ STOP codons for 6-frames, open reading frames, and various user-defined outputs. When genomes are imported from the database, shared orthologs can be shown. The Genome Annotation Transfer Utility (Tcherepenov et al., 2006) is a programme that allows you to annotate using information related to genomes from a reference genome. It supports responding to a user's input annotation and annotating genes by itself which are extremely related to the reference virus while leaving others up to human interpretation. This VBRC was originally designed for dsDNA viruses but has now been modified for coronavirus. SARS-CoV-2 and other viruses with similar characteristics have been included to the database.

10. VIRULIGN: Codon-correct Multiple Sequence Alignments

VIRULIGN was created to allow for rapid, codon-correct numerous sequence alignments and annotation of viral genomes using a reference sequence (Libin et al., 2017). A codon-aware alignment is required for researching the evolution of coding nucleotide sequences for different vaccine and antiviral production (Cuypers et al., 2016), understanding the establishment of drug resistance (Ngcapu et al., 2017) and quantifying epidemiological potential (Rambaut et al., 2020). Theys et al. (2010) shows that studying emerging diseases requires a representative and vetted annotation of open reading frames and proteins. To that aim, a SARS-CoV-2 reference sequence and genome annotation evolved from the first known genome sequence (Wu et al., 2020) have been added to VIRULIGN, covering all reading frames and proteins. VIRULIGN is peculiarly well adapted to studying the fast increasing number of SARS-CoV-2 genomes which are made public (Rambaut et al., 2020) because of its effective alignment technique with simple computing complexity in relation to the number of sequence analyzed.

Besides that, VIRULIGN's easily adaptable output formats (e.g., CSV file with headers analogous to genome annotation) makes it easier to integrate into analysis workflow, lowering the barrier for scientists to implement innovative bioinformatics pipelines (Libin et al., 2017) and database (Libin et al., 2013), which are required to monitor the COVID-19 pandemic.

11. Rfam COVID-19 Resources: Coronavirus-specific RNA Families

Rfam (Kalvari, 2018) is an RNA family database that contains curated various sequence alignments and covariance models. Rfam created a unique version 14.2 which includes ten different families spanning the complete 5' and 3' untranslated regions (UTRs) from *alpha-, beta-, gamma-* and *delta-* coronaviruses to aid in the study of coronavirus sequences. A

specific collection of Serbecovirus models, including SARS-CoV-2 sequences is also given.

The families are based on a combination of high quality whole-genome alignments evaluated by professional virologists. An updated collection of non-UTR corona virus structured RNAs are included in Rfam, like that of the frame shift stimulating element S2m RNA, and the 3` UTR pseudoknot. The updated Rfam families may be in combination with the Infernal Programme (Nawrocki et al., 2013) for annotating and predicting the secondary structure of structured RNAs in corona virus sequences. Furthermore, Rfam which is available online has sequence search that allows users to explore genome sequences for RNA elements.

12. UniProt COVID-19 Protein Portal: Rapid Access to Protein Information

UniProt (Uniprot Consortium, 2019) has acknowledged the need of annotating and making available the most up-to-date data on proteins which can be important to the disease for both the virus as well as the human host. As a result, the COVID-19 UniProt platform offers early pre-release access to;

(i) annotated protein sequences of SARS-CoV-2,
(ii) fairly close SARS proteins from SARS 2003,
(iii) human proteins which are appropriate to viral infection biology, such as various receptors and enzymes,
(iv) ProtVista (Watkins and Wulaningsih, 2020) which provides visual representation of sequence features for every individual protein,
(v) aligns to sequence analysis tools,
(vi) accessibility to a collection of community-contributed related papers, and
(vii) connections to pertinent resources.

13. Pfam Protein Families Database

For large-scale functional annotation of proteins, the Pfam protein families database is extremely used in molecular biology (Kieliszek et al., 2021). The sole SARS-CoV-2 protein that does not have a match with Orf10, a tiny potential protein discovered at the 3' end of the SARS-CoV-2 genome which seems to be unrelated to any other sequence in UniProtKB (https://covid-19.uniprot.org/). The hidden Markov model (HMM) library which is available on the Pfam profile, in conjunction with the HMMER software (Westerhoff and Kolodkin, 2020) enables fast annotation of corona virus which may be used to produce multiple sequence alignments that permits the recognition of mutations and clusters of similar sequences, which is especially useful for the tracking of outbreak and also researching evolution of the corona virus. The Pfam library of HMM is available for download purposes and may be utilized by combining with pfam_scan to conduct Pfam analysis directly from https://pfam.xfam.org. Hmmalign (http://hmmer.org/) may create multiple sequence alignments of similar matches. Matches and alignments are pre-calculated which can be accessible from the FTPsite of Pfam. Under the license of Creative Common Zero (CC0), Pfam is provided free of cost.

14. Tracking, Epidemiology and Evolution

Because there are no common methods for categorizing the genetic diversity of a viral species, different terms like 'subtypes', 'genotypes' or 'group' is being known for the phylogenetic subfamilies. Phylogenetic assignment, on the other hand is critical for studies on viral epidemiology, evolution and pathogenesis (see Covidex, Pangolin). As a result, for designating the rising number of phylogenetic lineages that comprises of SARS-CoV-2 population diversity, a nomenclature system is required. Rambaut et al., (2020) have presented a SARS-CoV-2 lineage nomenclature based on the concepts of basic evolutionary, phylogenetic and epidemiological factors.

Models of phylodynamic can support by understanding chronology of pandemics, which provides information into epidemiological factors such as e.g. R0 (Yang et al., 2020), or helps in determining the efficiency of viral control measures. At the most general level, studies related to phylodynamic seek to deduce epidemiological mechanisms from viral phylogenies by correlating genetic similarity to geographic similarities.

Main motive of these models is to help analyzing the impact of contact reduction measures or many other interventions, predicting hospital resource utilization, and directing decision making politically. SARS-CoV-2 is spontaneously acquiring mutations as the pandemic advances. The reported alterations are predicted to have negligible or little impact on viral biology, in general.

15. Covidex: Alignment-free Subtyping using Machine Learning

Viral subgroups or clades are groups of isolates from worldwide population of a specified species. Importance of subtypification can be seen in the fields of virus epidemiology, evolution and pathogenesis. The majority of subtype classification algorithms required the input data to be aligned to a collection of pre-defined subtype reference sequences. Major drawbacks of these techniques are that they may be computationally costly, especially for the sequence that are long like SARS-CoV-2 (~30 kb per genome). Machine learning methods for viral subtyping can be useful to address this issue (Solis-Reyes et al., 2018). A free and open-source machine learning subtyping tool was created known as Covidex. The default uploaded model for SARS-CoV-2 is based on Nextstrain (Hadfield et al., 2018) and GISAID data (Elbe et al., 2017). Models uploaded by users can also be utilized. Covidex is buit on a rapid random forest implementation trained on a k-mer database (Spiteri et al., 2020). Covidex significantly decreases computational and time requirements by training classification algorithms across k-mer frequency vectors, allowing it to identify hundreds of SARS-CoV-2 genomes within seconds. Therefore, in the context of the present worldwide, when the amount of accessible SARS-CoV-2 genomes rapidly increases, the research on it can profit from this unique technology which is developed to minimize the time required in the data analysis.

16. Pangolin: Phylogenetic Assignment of Named Global Outbreak Lineages

Based on the lineage nomenclature established by Rambaut et al., (2020), Pangolin allocates a worldwide lineage to query SARS-CoV-2 genomes by determining the most likely location in aphylogenetic tree of reference sequences from all presently known global SARS-CoV-2 lineages. It can easily be scaled, allowing it to execute on thousands or few of the sequences. Technically, Pangolin uses mafft (Katoh and Standley, 2013) and iqtree (Güvenç et al., 2018), which provides a guide tree and alignment to reduce analysis overhead to a minimum burden. Wide ranges of applications are being offered by Pangolin, including front-line hospital use as well as domestic and worldwide surveillance. In hospitals, for example, sequencing SARS-CoV-2 samples may be utilized for ruling out intra-hospital transmission and efforts in guiding infection control. It may also be used for monitoring, summarizing which lineages are present in a certain location. The online application also communicates with Microreact (https://www.microreact.org/) which displays query sequence in the context of global lineages. To allocate lineages to UK sequences pangolin is utilized as part of COG- UK's (https://www.cogconsortium.uk/) data analysing lineages. People can also create and submit their own finer-scale lineages, such as within-country lineages, as well as their own guide tree and alignment.

It is simple to extract relevant data from Pangolin through real-time viral genome sequencing and can help with establishing innovative developments and tracking SARS-CoV-2 propagation.

17. BEAST 2: Phylodynamics based on Bayesian Inference

Bayesian phylodynamic inference (Grenfall et al., 2004), which permits the appropriate integration of proof from many independent sources of information, such as sequences of genome, sampling dates, and geographical locations, can be used to answer vital evolutionary and epidemiological concerns about SARS-CoV-2. BEAST 2 (Bouckaert et al., 2019) is a complex computational software system which allows for comprehensive Bayesian analysis to be performed using a variety of phylodynamic programs, such as the one used by Volz and Siveroni (2018). The history of phylogenetic tree is estimated concurrently with evolutionary and epidemiological parameters, allowing all elements of the combined model's uncertainty to be accounted for and also represented in the results. Phylodynamic analysis of SARS-CoV-2 is critical for interpreting (i) evolutionary dynamics of SARS-CoV-2, especially the evolutionary level at which mutations become fixated in the viral genome, (ii) with time origin, selection of COVID-19 cases as an estimation of the time with which a sub-epidemic surfaced, (iii) sub- epidemic's geographical origin, (iv) SARS-CoV-2 transmission dynamics, such as calculations of the effective reproducing number directly Re and its variations over time, and lastly (v) the fraction of undiscovered COVID-19 cases. Although, evolutionary and epidemiological processes appear within the same time period, viral genome diversity gives information

and host transmission dynamics allowing Bayesian phylodynamic study of SARS-CoV-2 critical supplement to traditional epidemiological approaches.

18. Phylogeographic Reconstruction using air Transportation Data

To reconstitute the presumed spread paths and epidemic roots of fast evolving pathogens, phylogeographic techniques combines genomic data along with viral isolate sampling locations and models of dispersal, such as air travel or local diffusion. Reimering et al., (2020) has made for internal nodes in a phylogenetic tree using a parsimonious reconstruction and effective distances can be estimated using passenger *f* lows among airports. A short distance includes a great connection between two airports. The rebuilding of parsimonious determines ancestral sites for intrinsic modes of the tree which minimizes the distances throughout the phylogeny by using these distances as a cost of matrix. This approach enables quick inference of distribution pathways on quite well geographical scale. Effective distance reconstruction presumes phylogeographic spread more correctly than geographic distance reconstruction or Bayesian reconstruction that does not employ any location data. Air transportation data may be utilized by phylogeographic reconstruction to analyze the worldwide outbreak of the SARS-CoV-2 pandemic, particularlyin the early stages when air travel still played a significant role in viral spreading. The technique is now being modified to take into accounts almost both air travel and local movement information inside nations during inference in order to determine shifting global movements during various stages of the pandemic.

19. COPASI: Modeling SARS-CoV-2 Dynamics with Differential Equations

A dynamic simulator known as COPASI was initially designed to simulate chemical and biological reactors network (Hadfield et al., 2018). Moreover, it is increasingly frequently used in other disciplines such as epidemiology. It enables for the simulation of models using both the classic differential equation method, which portrays populations as continuous and the stochastic kinetics strategy, which believes that populations to be made up of individuals. COPASI offers a single model representation for each of these techniques, allowing for easy transition between them. Models can also contain arbitrary discrete occurrences. This software contains numerous algorithms which offer extensive assessments of models, and it is capable for estimating parameters by utilizing a variety of optimization methods. Several aspects of virology including mechanism of action (Naveed et al., 2021), pharmaceutical interventions (Reiling et al., 2002), virus life-cycle (Cuypers et al., 2016), vaccine design (Upton et al., 2000) and epidemic dynamics (Abrams et al., 2021) has been used to model with the help of COPASI. It has also been used to predict the epidemic dynamics and the impact of treatment in Covid-19 (Westerhoff and Kolodkin, 2020). Some writers have also utilized COPASI to stimulate local epidemics and anticipate resources used by hospitals as well as to assess the potential advantages of contact network agent based models over differential equation models.

20. COVIDSIM: Epidemiological Models of Viral Spread

Conventional models of epidemiology were widely used to describe the COVID-19 pandemic. Individuals in a population are assigned to distinct subgroups in deterministic or compact mental mathematical models, which explain their dynamic changes using framework of differential equations. The SEIR Model and expanded variants of it are commonly used for SARS-CoV-2. The fundamental model systems are not novel, and similar models to represent infectious illness were reported as early 20th century (Zsidisin et al., 2000). Individuals in a population are classified as Susceptible (S), Exposed (E), Infected (I), Recovered (R), or Deceased (D) in their SEIR or SEIRD-model. Firstly, each person is considered vulnerable to infection, with the exception of a tiny number who are already infected. The model may then mimic the population infections dynamics by parameterizing the differential equations considering factors like incubation period or mean illness duration. Such SEIR models have been used to forecast COVID-19 trends for example, in Spain and Italy, as well as to assess the impact of management method (Zhao et al., 2020). Extensive variants of the SEIR model have been created to aid in political decision making (Libin et al., 2017). In Germany, for

example, this model is applied in COVIDSIM which includes hospitalized and critical care patients as well as the results of contact reduction strategies. It may be layered with data from several German federal states as well as data from other nations. This model includes an easy to use online interface, which allows the user to alter model parameters and get a feel for the model dynamics allowing it to predict infection parameters, analyse the impacts of contact reduction strategies, and drive political decision-making.

21. CoV-GLUE: Tracking Nucleotide Changes in the SARS-CoV-2 Genome

As the epidemic proceeds, SARS-CoV-2's genome naturally accumulates nucleotide mutations. In viral genome sequences, point mutations, especially non-synonymous substitutions result in amino-acid replacements, whereas other changes result in insertions or deletions (*indels*). On average, the reported alterations are predicted to have negligible or little impact on viral biology. Monitoring these changes, on the other hand will help in understanding the pandemic since mutations can have an influence on virus biology and lead to resistance to antiviral medicines and future vaccine. The database was created with the help of GLUE, a data-centric bioinformatics environment for viral sequence dates that focuses on variations, evolution and sequence analyses (Varatharaj et al., 2020). Every week sequences are obtained from GISAID EpiCoV and contributed to a restricted alignment inside the GLUE system. People can now examine the growing variety or contribute a FASTA file of a novel genome to COV-GLUE for comparison to existing data. The provided data generates an amino acid substitutions, *indels* and diagnostic primer design report. Users may access the identified variations and see their sequence relative to a reference data set using a phylogenetic placement maximum-likelihood technique. In accordance with Rambaut et al. (2020), the user's sequence is then allocated to a lineage. CoV-GLUE will aid in SARS-CoV-2 research by tracking changes in the SARS-CoV-2 genome.

22. Conclusion

Bioinformatics is a vast study and spreading multidisciplinary branches all over the world. We have already discussed numerous applications of bioinformatics in every stage of SARS-CoV-2 infection from the beginning of symptoms in host body till drug discovery tool. Research tools including SARS-CoV-2 detection and genomic studies empower to boost up the study for COVID-19 treatment. This review is neither claiming complete set of tools for detection and treatment of COVID-19 nor enclosing ending research for it.

Funding Information

Uma Maheshwari is thankful to DBT, New Delhi, for the Junior Research Fellowship.

Declaration of Conflict

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

References

1. Abrams, S., Wambua, J., Santermans, E., Willem, L., Kuylen, E., Coletti, P., Libin, P., Faes, C., Petrof, O., Herzog, S. A., Beutels, P., & Hens, N. (2021). Modelling the early phase of the Belgian COVID-19 epidemic using a stochastic compartmental model and studying its implied future trajectories. Epidemics, 35, 100449. https://doi.org/10.1016/j.epidem.2021.100449
2. Barzkar, N., Sohail, M., Tamadoni Jahromi, S., Gozari, M., Poormozaffar, S., Nahavandi, R., & Hafezieh, M. (2021). Marine bacterial esterases: Emerging biocatalysts for industrial applications. Applied Biochemistry and Biotechnology, 193(4), 1187–1214. https://doi.org/10.1007/s12010-020-03483-8
3. Bouckaert, R., Vaughan, T. G., Barido-Sottani, J., Duchêne, S., Fourment, M., Gavryushkina, A., Heled, J., Jones, G., Kühnert, D., De Maio, N., Matschiner, M., Mendes, F. K., Müller, N. F., Ogilvie, H. A., du Plessis, L., Popinga, A., Rambaut, A., Rasmussen, D., Siveroni, I., … Drummond, A. J. (2019). BEAST 2.5: An advanced software platform for Bayesian evolutionary analysis. PLOS Computational Biology, 15(4), e1006650. https://doi.org/10.1371/journal.pcbi.1006650
4. Corman, V. M., Landt, O., Kaiser, M., Molenkamp, R., Meijer, A., Chu, D. K., Bleicker, T., Brünink, S., Schneider, J., Schmidt, M. L., Mulders, D. G., Haagmans, B. L., van der Veer, B., van den Brink, S., Wijsman, L., Goderski, G., Romette, J. L., Ellis, J., Zambon, M., … and Drosten, C. (2020). Detection of 2019 novel coronavirus (2019-nCoV) by real-time RT-PCR. Eurosurveillance, 25(3). PubMed: 2000045
5. Cuypers, L., Li, G., Neumann-Haefelin, C., Piampongsant, S., Libin, P., Van Laethem, K., Vandamme, A. M., & Theys,

K. (2016). Mapping the genomic diversity of HCV subtypes 1a and 1b: Implications of structural and immunological constraints for vaccine and drug development. Virus Evolution, 2(2), vew024. https://doi.org/10.1093/ve/vew024

6. Di Renzo, L., Gualtieri, P., Pivari, F., Soldati, L., Attinà, A., Cinelli, G., Leggeri, C., Caparello, G., Barrea, L., Scerbo, F., Esposito, E., & De Lorenzo, A. (2020). Eating habits and lifestyle changes during COVID-19 lockdown: An Italian survey. Journal of Translational Medicine, 18(1), 229. https://doi.org/10.1186/s12967-020-02399-5
7. El-Gebali, S., Mistry, J., Bateman, A., Eddy, S. R., Luciani, A., Potter, S. C., Qureshi, M., Richardson, L. J., Salazar, G. A., Smart, A., Sonnhammer, E. L. L., Hirsh, L., Paladin, L., Piovesan, D., Tosatto, S. C. E., & Finn, R. D. (2019). The Pfam protein families database in 2019. Nucleic Acids Research, 47(D1), D427–D432. https://doi.org/10.1093/nar/gky995
8. Grenfell, B. T., Pybus, O. G., Gog, J. R., Wood, J. L., Daly, J. M., Mumford, J. A., & Holmes, E. C. (2004). Unifying the epidemiological and evolutionary dynamics of pathogens. Science, 303(5656), 327–332. https://doi.org/10.1126/science.1090727
9. Hadfield, J., Megill, C., Bell, S. M., Huddleston, J., Potter, B., Callender, C., Sagulenko, P., Bedford, T., & Neher, R. A. (2018). Nextstrain: Real-time tracking of pathogen evolution. Bioinformatics, 34(23), 4121–4123. https://doi.org/10.1093/bioinformatics/bty407
10. Kalvari, I. K. (2018). Computational approaches for the identification of LIR-motifs in selective autophagy receptor and adaptor proteins [Doctoral Dissertation]. Faculty of Pure and Applied Sciences, University of Cyprus.
11. Katoh, K., & Standley, D. M. (2013). MAFFT multiple sequence alignment software version 7: Improvements in performance and usability. Molecular Biology Evolution, 30(4), 772–780. https://doi.org/10.1093/molbev/mst010
12. Kieliszek, M., Pobiega, K., Piwowarek, K., & Kot, A. M. (2021). Characteristics of the proteolytic enzymes produced by lactic acid bacteria. Molecules, 26(7). https://doi.org/10.3390/molecules26071858
13. Klenk, H. D., & Garten, W. (1994). Host cell proteases controlling virus pathogenicity. Trends in Microbiology, 2(2), 39–43. https://doi.org/10.1016/0966-842x(94)90123-6
14. Libin, P., Beheydt, G., Deforche, K., Imbrechts, S., Ferreira, F., Van Laethem, K., Theys, K., Carvalho, A. P., Cavaco-Silva, J., Lapadula, G., Torti, C., Assel, M., Wesner, S., Snoeck, J., Ruelle, J., De Bel, A., Lacor, P., De Munter, P., Van Wijngaerden, E., … and Vandamme, A. M. (2013). RegaDB: Community-driven data management and analysis for infectious diseases. Bioinformatics, 29(11), 1477–1480. https://doi.org/10.1093/bioinformatics/btt162
15. Libin, P., Vanden Eynden, E., Incardona, F., Nowé, A., Bezenchek, A., Group, E. S., Sönnerborg, A., Vandamme, A. M., Theys, K., & Baele, G. (2017). PhyloGeoTool: Interactively exploring large phylogenies in an epidemiological context. Bioinformatics, 33(24), 3993–3995. https://doi.org/10.1093/bioinformatics/btx535
16. Madsen, E. E., Krustrup, P., Larsen, C. H., Elbe, A. M., Wikman, J. M., Ivarsson, A., & Lautenbach, F. (2021). Resilience as a protective factor for well-being and emotional stability in elite-level football players during the first wave of the COVID-19 pandemic. Science and Medicine in Football, 5(sup1)(Suppl. 1), 62–69. https://doi.org/10.1080/24733938.2021.1959047
17. Mitchell, A. L., Almeida, A., Beracochea, M., Boland, M., Burgin, J., Cochrane, G., Crusoe, M. R., Kale, V., Potter, S. C., Richardson, L. J., Sakharova, E., Scheremetjew, M., Korobeynikov, A., Shlemov, A., Kunyavskaya, O., Lapidus, A., & Finn, R. D. (2020). MGnify: The microbiome analysis resource in 2020. Nucleic Acids Research, 48(D1), D570–D578. https://doi.org/10.1093/nar/gkz1035
18. Naveed, M., Nadeem, F., Mehmood, T., Bilal, M., Anwar, Z., & Amjad, F. (2021). Protease—A versatile and ecofriendly biocatalyst with multi-industrial applications: An updated review. Catalysis Letters, 151(2), 307–323. https://doi.org/10.1007/s10562-020-03316-7
19. Nawrocki, E. P., & Eddy, S. R. (2013). Infernal 1.1: 100-Fold faster RNA homology searches. Bioinformatics, 29(22), 2933–2935. https://doi.org/10.1093/bioinformatics/btt509
20. Ngcapu, S., Theys, K., Libin, P., Marconi, V. C., Sunpath, H., Ndung'u, T., & Gordon, M. L. (2017). Characterization of nucleoside reverse transcriptase inhibitor-associated mutations in the RNase H region of HIV-1 subtype C infected individuals. Viruses, 9(11), 330. https://doi.org/10.3390/v9110330
21. Rambaut, A., Holmes, E. C., O'Toole, Á., Hill, V., McCrone, J. T., Ruis, C., du Plessis, L., & Pybus, O. G. (2020). A dynamic nomenclature proposal for SARS-CoV-2 lineages to assist genomic epidemiology. Nature Microbiology, 5(11), 1403–1407. https://doi.org/10.1038/s41564-020-0770-5
22. Reiling, N., Hölscher, C., Fehrenbach, A., Kröger, S., Kirschning, C. J., Goyert, S., & Ehlers, S. (2002). Cutting edge: Toll-like receptor (TLR) 2-and TLR4-mediated pathogen recognition in resistance to airborne infection with Mycobacterium tuberculosis. Journal of Immunology, 169(7), 3480–3484. https://doi.org/10.4049/jimmunol.169.7.3480
23. Reimering, S., Muñoz, S., & McHardy, A. C. (2020). Phylogeographic reconstruction using air

transportation data and its application to the 2009 H1N1 influenza A pandemic. PLOS Computational Biology, 16(2), e1007101. https://doi.org/10.1371/journal.pcbi.1007101

24. Solis-Reyes, S., Avino, M., Poon, A., & Kari, L. (2018). An open-source k-mer based machine learning tool for fast and accurate subtyping of HIV-1 genomes. PLOS ONE, 13(11), e0206409. https://doi.org/10.1371/journal.pone.0206409
25. Spiteri, G., Fielding, J., Diercke, M., Campese, C., Enouf, V., Gaymard, A., Bella, A., Sognamiglio, P., Sierra Moros, M. J., Riutort, A. N., Demina, Y. V., Mahieu, R., Broas, M., Bengnér, M., Buda, S., Schilling, J., Filleul, L., Lepoutre, A., Saura, C., … and Ciancio, B. C. (2020). First cases of coronavirus disease 2019 (COVID-19) in the WHO European Region, 24 January to 21 February 2020. Eurosurveillance, 25(9). PubMed: 2000178
26. Tcherepanov, V., Ehlers, A., & Upton, C. (2006). Genome Annotation Transfer Utility (GATU): Rapid annotation of viral genomes using a closely related reference genome. BMC Genomics, 7(1), 150. https://doi.org/10.1186/1471-2164-7-150
27. Theys, K., Deforche, K., Libin, P., Camacho, R. J., Van Laethem, K., & Vandamme, A. M. (2010). Resistance pathways of human immunodeficiency virus type 1 against the combination of zidovudine and lamivudine. Journal of General Virology, 91(8), 1898–1908. https://doi.org/10.1099/vir.0.022657-0
28. UniProt Consortium. (2019). UniProt: A worldwide hub of protein knowledge. Nucleic Acids Research, 47(D1), D506–D515. https://doi.org/10.1093/nar/gky1049
29. Upton, C., Hogg, D., Perrin, D., Boone, M., & Harris, N. L. (2000). Viral genome organizer: A system for analyzing complete viral genomes. Virus Research, 70(1–2), 55–64. https://doi.org/10.1016/s0168-1702(00)00210-0
30. Varatharaj, A., Thomas, N., Ellul, M. A., Davies, N. W. S., Pollak, T. A., Tenorio, E. L., Sultan, M., Easton, A., Breen, G., Zandi, M., Coles, J. P., Manji, H., Al-Shahi Salman, R., Menon, D. K., Nicholson, T. R., Benjamin, L. A., Carson, A., Smith, C., Turner, M. R., … and CoroNerve Study Group. (2020). Neurological and neuropsychiatric complications of COVID-19 in 153 patients: A UK-wide surveillance study. Lancet Psychiatry, 7(10), 875–882. https://doi.org/10.1016/S2215-0366(20)30287-X
31. Volz, E. M., & Siveroni, I. (2018). Bayesian phylodynamic inference with complex models. PLOS Computational Biology, 14(11), e1006546. https://doi.org/10.1371/journal.pcbi.1006546
32. Wang, Y., Cui, X., Chen, X., Yang, S., Ling, Y., Song, Q., Zhu, S., Sun, L., Li, C., Li, Y., Deng, X., Delwart, E., & Zhang, W. (2020). A recombinant infectious bronchitis virus from a chicken with a spike gene closely related to that of a turkey coronavirus. Archives of Virology, 165(3), 703–707. https://doi.org/10.1007/s00705-019-04488-3
33. Watkins, J., & Wulaningsih, W. (2020). Three further ways that the COVID-19 pandemic will affect health outcomes. International Journal of Public Health, 65(5), 519–520. https://doi.org/10.1007/s00038-020-01383-6
34. Westerhoff, H. V., & Kolodkin, A. N. (2020). Advice from a systems-biology model of the corona epidemics. npj Systems Biology and Applications, 6(1), 18. https://doi.org/10.1038/s41540-020-0138-8
35. Wu, D., Lu, J., Liu, Y., Zhang, Z., & Luo, L. (2020). Positive effects of COVID-19 control measures on influenza prevention. International Journal of Infectious Diseases, 95, 345–346. https://doi.org/10.1016/j.ijid.2020.04.009
36. Yang, G. Z., J Nelson, B., Murphy, R. R., Choset, H., Christensen, H., H Collins, S., Dario, P., Goldberg, K., Ikuta, K., Jacobstein, N., Kragic, D., Taylor, R. H., & McNutt, M. (2020). Combating COVID-19—The role of robotics in managing public health and infectious diseases. Science Robotics, 5(40), eabb5589. https://doi.org/10.1126/scirobotics.abb5589
37. Zhao, S., Lin, Q., Ran, J., Musa, S. S., Yang, G., Wang, W., Lou, Y., Gao, D., Yang, L., He, D., & Wang, M. H. (2020). Preliminary estimation of the basic reproduction number of novel coronavirus (2019-nCoV) in China, from 2019 to 2020: A data-driven analysis in the early phase of the outbreak. International Journal of Infectious Diseases, 92, 214–217. https://doi.org/10.1016/j.ijid.2020.01.050
38. Zsidisin, G. A., Panelli, A., & Upton, R. (2000). Purchasing organization involvement in risk assessments, contingency plans, and risk management: An exploratory study. Supply Chain Management, 5(4), 187–198. https://doi.org/10.1108/13598540010347307