



# **COVID Variants, Villain and Victory: A Bioinformatics Perspective**

Nityendra Shukla<sup>1,†</sup>, Neha Srivastava<sup>2,†</sup>, Rohit Gupta<sup>1</sup>, Prachi Srivastava<sup>2</sup> and Jitendra Narayan<sup>1,\*</sup>

- <sup>1</sup> CSIR Institute of Genomics and Integrative Biology, Mall Road, Delhi 110007, India; nitinshukla218@gmail.com (N.S.); rohit.gupta.delhi1995@gmail.com (R.G.)
- <sup>2</sup> Amity Institute of Biotechnology, Amity University, Uttar Pradesh, Lucknow Campus, Lucknow 226010, India; ns011982@gmail.com (N.S.); psrivastava@amity.edu (P.S.)
- \* Correspondence: jnarayan@igib.res.in
- <sup>+</sup> These authors contributed equally to this work.

**Abstract:** The SARS-CoV-2 virus, a novel member of the Coronaviridae family, is responsible for the viral infection known as Coronavirus Disease 2019 (COVID-19). In response to the urgent and critical need for rapid detection, diagnosis, analysis, interpretation, and treatment of COVID-19, a wide variety of bioinformatics tools have been developed. Given the virulence of SARS-CoV-2, it is crucial to explore the pathophysiology of the virus. We intend to examine how bioinformatics, in conjunction with next-generation sequencing techniques, can be leveraged to improve current diagnostic tools and streamline vaccine development for emerging SARS-CoV-2 variants. We also emphasize how bioinformatics, in general, can contribute to critical areas of biomedicine, including clinical diagnostics, SARS-CoV-2 genomic surveillance and its evolution, identification of potential drug targets, and development of therapeutic strategies. Currently, state-of-the-art bioinformatics tools have helped overcome technical obstacles with respect to genomic surveillance and have assisted in rapid detection, diagnosis, and delivering precise treatment to individuals on time.

Keywords: SARS-CoV-2; genomics; immunoinformatics; drug design; bioinformatics; COVID-19



Citation: Shukla, N.; Srivastava, N.; Gupta, R.; Srivastava, P.; Narayan, J. COVID Variants, Villain and Victory: A Bioinformatics Perspective. *Microorganisms* 2023, *11*, 2039. https://doi.org/10.3390/ microorganisms11082039

Academic Editor: Qibin Geng

Received: 21 June 2023 Revised: 11 July 2023 Accepted: 11 July 2023 Published: 9 August 2023



**Copyright:** © 2023 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (https:// creativecommons.org/licenses/by/ 4.0/).

## 1. Introduction

The COVID-19 pandemic, an unprecedented catastrophe, affected nearly every country on multiple levels. Since its emergence in December 2019, the pandemic has caused severe loss of human life, damage to healthcare and educational systems, and displaced economies worldwide. The scientific research community currently faces an arduous challenge since the virus undergoes frequent mutations, leading to the emergence of new variants. The high infectivity and lethality rates of the SARS-CoV-2 virus during its initial days and now, with the given propensity of the virus to mutate, further cause it to remain a major concern. To monitor and track changes in the SARS-CoV-2 genome caused by mutations and identify novel variants, various sequence-based surveillance, laboratory, and epidemiological investigations are now regularly being conducted. Continuous and vigilant genomic surveillance and analysis of viral variants in real time are crucial in the development of effective diagnostics, therapeutics, and vaccines to combat the onslaught of the pandemic.

A variant is defined as having two or more mutations that distinguish it from other variants in terms of spreading. In collaboration with the Centers for Disease Control and Prevention (CDC) and the US Department of Health and Human Services (HHS), the two institutions established an Interagency Group (SIG) to develop a classification system for SARS-CoV-2 variants. They classified them into four categories and marked them as follows: (1) variants of interest (VOI), (2) variants of concern (VOC), (3) variants of high consequences (VOHC), and 4) variants being monitored (VBM) [1,2]. Variants of interest (VOI) are variants with specific heredity marker changes characterized by alterations in the receptor-binding domain (RBD). In addition to reduced antibody neutralization generated

post-infection and inoculation, variants can result in decreased efficacy of therapeutics, possible analytic effect, or anticipated expansion in contagiousness or chronic illness. During different peaks of the pandemic, certain variants emerged as notable players. In the spring of 2021, the Iota variant (B.1.526) took the spotlight in the United States, while in the summer of the same year, the Epsilon variant (B.1.427 & B.1.429) gained prominence in the region, exhibiting an approximate 20% increase in transmission rate. In the United Kingdom (UK), the Eta variant (B.1.525) emerged during the spring of 2021. In India, the Kappa variant (B.1.617.1 & B.1.617.3) made its presence known, eventually giving rise to the highly infectious Delta variant. As of June 2023, the VOIs circulating in the population are XBB.1.5 and XBB.1.16, both of which are suspected to possess increased immune escape potential and continue to cause surges in infection rates (Figure 1) [3].

## SARS-CoV-2 cases per million



Source: https://github.com/owid/covid-19-data



Lineages classified as VOC have been the most devastating during the course of the pandemic, leading to high hospitalization and mortality rates, in addition to the development of chronic health issues such as post-acute sequelae of COVID-19 (PASC), more commonly known as long COVID, type 2 diabetes, fatigue, and dyslipidemia, among others, resulting in lifelong disability and medical care [5]. The hallmarks of VOCs are their increased rates of transmission and immune escape potential, leading to a significant reduction in neutralization antibodies in addition to decreased effectiveness of therapeutics and vaccines, resulting in high rates of hospitalization and death. The major VOCs throughout the course of the pandemic included the Alpha (B.1.1.7) variant, identified in the UK in December 2020 [6], possessing 23 mutations as compared to the wild-type variant, with an extremely high transmissibility rate of ~70% [7]; it had become dominant in 21 countries by March 2021. The Beta (B.1.351) variant was first identified in South Africa in October 2020, defined by 12 mutations in its genome compared to the wild-type, and was the first VOC to display decreased vaccine efficacy and monoclonal antibody resistance [8]; it was responsible for pandemic waves across the African continent. The Delta (B.1.617.2), first identified in India in October 2020, was the most devastating out of all VOCs so far. It was considered to be 50% more transmissible than the Alpha variant and was the dominant lineage during India's second wave, and it contributed to third waves in the UK and South Africa and eventually became the dominant variant worldwide. The Delta variant possessed 13 mutations, 8 of which were in the spike (S) protein [9], thus increasing its transmissibility and ability to infect the lower respiratory tract, causing severe disease conditions such as acute respiratory distress syndrome (ARDS) [10]. The current

VOC, Omicron (B.1.1.529), identified in November 2021 in South Africa, is the most diverse variant of all the VOCs before it, carrying approximately 50 mutations, and has since continuously evolved into subvariants, leading to WHO changing their classification system; thus, Omicron subvariants are tracked independently [11]. Omicron and its subvariants possess significant immune escape potential, with Omicron subvariants such as BA.1, BA.2, BA.4, and BA.5 being poorly neutralized by first-generation vaccines, in addition to being resistant against most monoclonal antibodies except bebtelovimab [3]. However, Omicron seems to display a bias for infecting the upper respiratory tract over the lower respiratory tract [12], possibly due to lower fusogenicity and possibly altered tissue tropism due to its ability to be primed by endosomal proteases, such as cathepsins, a phenomenon seen in SARS-CoV [13–15]. This mechanism is hypothesized to somewhat play a role in the decreased severity of disease, though all current evidence that supports this is limited to rodent models [15,16].

Thus, these variants play the role of a villain, causing serious health complications in people who are infected and post-infection as well. Delta variants have been found to be significantly more virulent as well as contagious as compared to other variants and were mainly responsible for the global surges in 2021, particularly in India, South Africa, and the United Kingdom. The high transmission rates expose the population to repeat infections, potentially leading to a higher risk of developing long COVID [17], whose symptoms can last for years and in some cases, lifelong. Additionally, repeat COVID-19 infections heighten the risk of developing adverse comorbidities such as cardiovascular disease, type 2 diabetes [18], or myalgic encephalomyelitis/chronic fatigue syndrome (ME/CFS) [19]. Furthermore, the Delta variant displayed increased lethality as compared to the Omicron variant, with a median case-fatality rate (CFR) of 8.56 as compared to 3.04, respectively [20], probably due to increased vaccine coverage and decreasing pathogenicity as the virus continues to evolve. As the virus evolves, the effect of medical therapeutics and available interventions continues to decrease, leading to an increased risk of hospitalization, particularly in high-risk groups. New mutations drive a strong increase in transmission rates, and a constantly evolving machinery of SARS-CoV-2 means decreased drug response to new symptoms that the virus may give rise to. Therefore, mutations and variants impeded research; as a result, universal inoculation was the effective long-term solution to thwart the growing risk of an evolving SARS-CoV-2 virus and assist in decreasing mortality and hospitalization rates [21].

Bioinformatics has created a milestone and has played an essential role in COVID-19 research since the emergence of the pandemic [22], especially in detecting and tracking new variants, as well as in allowing for open data-sharing across the world. Wide varieties of bioinformatics tools and techniques have been used successfully in the interpretation of the genomic architecture of SARS-CoV-2 and its variants, as well as in the creation of mathematical models used to predict infection spread, design containment methods, and drive public health policy decisions [23,24]. They have been utilized in analyzing the data generated from genomics, transcriptomics, proteomics, and structural omics, as well as single-cell data. They have contributed to a better understanding of the underlying molecular mechanisms of viral pathogenesis, allowing for the rapid identification of vaccine and drug candidates [25], particularly for the rapid deployment of repurposed drugs during the pandemic [24,26] for clinical research and eventually, for clinical trials. From this perspective, we highlight the strong contributions and achievements of bioinformatics tools and techniques in combating the COVID-19 pandemic.

#### 2. Tracking of SARS-CoV-2

Next-generation sequencing (NGS), also defined as high-throughput sequencing technology, has become a widely adopted approach in genomics research to identify virus origins, delineate mechanistically phenotypic causative pathophysiology, and elucidate genotypic ramifications in infected individuals. These technologies have further led to the development of novel bioinformatics tools, pipelines, algorithms, and machine-learning approaches that have now become robust practices in understanding the virus genome [27].

Advancements made in metagenomics have aided in the identification of plausible co-occurring pathogens that may play a role in the clinical outcome of SARS-CoV-2 infections. Thus, studying host–pathogen interactions also become important in this context. Augmenting this with different analytical strategies that bioinformatics tools offer allows us to unravel the same genomic data via different lenses to uncover the underlying layers of host–pathogen interactions, which may modulate disease severity as well as clinical outcome [28]. Various metagenomics tools such as Kraken2 [29], MOTHUR [30], and QIIME2 [31] provide deep insights into genomic data as well as provide ways to visualize data; thus, meaningful insights can be extracted. Further, this is combined with Seurat [32], a single-cell analysis R library, which allows for spatial visualization of multiple clusters of cells that may themselves be strongly implicated in disease pathophysiology. The analysis is then extended further with downstream machine-learning tools, such as singleR [33] and clustifyr [34]. These can be leveraged to annotate cell clusters with public data libraries with high confidence and further identify causal gene–pathway relationships with respect to the cell types that may be contributing factors in disease pathophysiology.

In this context, an opportunity exists to develop a cohesive and well-developed workflow that takes into account distinct steps that occur in the data analysis, thus streamlining the process of variant discovery. Typically, an NGS data analysis pipeline consists of various integral steps, such as quality control of the data, deletion of extraneous host data, and read assembly, followed by taxonomic classification (of pathogens) and, lastly, supplemented by virus genome verification. Due to the wide adoption, use, and build of various open-source methods, various tools exist that help in further carrying out each step elucidated above.

Bioinformatics tools have accelerated efforts in unwinding the mutations and genetic variation of the SARS-CoV-2 virus. Undoubtedly, SARS-CoV-2 has the potential to adapt during the current pandemic because of its high and rapid mutation rate. How this evolution has an impact on the transmission, duration, and gravity of disease is still under study. The increasing amount of evidence of human–wildlife interactions has facilitated the zoonotic transmission from animals to humans.

In SARS-CoV-2, the spike protein of the virus has sufficient binding affinity with the receptor of angiotensin-converting enzyme 2 (ACE2), which is crucial for host cell entry and human infection. Several computational models and data have also indicated that additional mutations at the binding site strengthen the binding affinity [35,36]. According to Nextstrain [37], SARS-CoV-2 underwent roughly one genetic change per week based on the substitution rate. The mutation rate of the SARS-CoV-2 virus is estimated at  $1 \times 10^{-6}$ - $2 \times 10^{-6}$  mutations per nucleotide per replication cycle [38]. The availability of abundant data and resources on the viral genome exhibited the intense response to the pandemic by tracking and tracing of infection to drug and vaccine development. Nextstrain collects, analyzes, visualizes, and maintains metadata of SARS-CoV-2 from various public data repositories, including NCBI (www.ncbi.nlm.nih.gov, accessed 9 July 2023), GISAID (www.gisaid.org, accessed 9 July 2023), and GitHub repositories.

The genomic data allows us to understand the dynamics of the genome evolution of viruses, including mutations and natural selection, which are very important aspects for detecting clinical variants and their epidemiological significance, understanding viral evolution and responses of the immune system to mutations, and vaccine and drug development. The graphical display of sequence data is also available for a better understanding of the genomic epidemiology of SARS-CoV-2 (Figure 2). Bioinformatics strategies play a key role in the rapid detection, tracing, understanding, analysis, evaluation, and treatment of COVID-19. With the sheer amount of available data, scientists are globally investigating evolution at the genome and protein level to tackle the pandemic [39]. For early detection, tracking, tracing, sequencing, and the creation of therapeutic methods, a variety of bioinformatics workflows and tools have been developed. Interestingly, we have seen a more than four-fold increase in drug designing tools and software in the last few years, while ML tools have increased by two-fold (Figure 3). To avoid the false-positive and false-negative detection of qRT-PCR [40], a computation-based primer PriSeT has been developed for detecting the specificity and sensitivity of the qRT-PCR test [41].



Figure 2. In silico vaccine development timeline [42,43].



**Figure 3.** Stacked bar chart displaying the rising rate of published computational methods across various domains over the years.

Next-generation sequencing workflow V-Pipe [44] and Haploflow [45] are developed to evaluate and monitor genetic diversity and mutation. Amplicon-based metagenomics sequencing has been prevalent throughout the pandemic and has been extensively used to study sequence divergence [46,47]. Several open-source databases, such as GISAID, NCBI, and EMBL-EBI, provide easy access and submission for high-quality SARS-CoV-2 genome sequence metadata globally. For the detection and annotation of viral genomes, numerous bioinformatics tools, software, and pipelines have been developed, including VADR [48], VBRC tools (https://4virology.net/virology-ca-tools/, accessed 4 July 2023), and VIRULIGN [49]. Various databases, such as UniProt, Pfam, and Rfam [50], are available to allow users to investigate coding and non-coding sequence evolution and diversity to understand viral epidemiology and evolution.

Besides this, numerous platforms and models were developed to study the fundamentals of evolutionary, epidemiology, and phylogenetic divergence of viruses. The phylodynamic models include BEAST 2 [51] for phylogeographic reconstruction; epidemicmathematical models such as COPASI [52] and COVIDSIM [53]; and evolutionary tracking models like CoV-GLUE [54] and CoVe-Tracker are available [55]. The machine-learning model Covidex [56] is an open-source alignment tool based on Nextstrain and GISAID data used for the rapid classification of viral genomes in pre-defined clusters isolated from the population. Pangolin [57] rapidly assigns the most likely classification for a large number of genome sequences, which is particularly useful for local and global surveillance. All of these tools are freely available around the world for users. Thus, bioinformaticians are quickly responding to the pandemic and have provided easy and rapid access to COVID-19-specific tools and techniques for tracking and tracing the virus.

Certain limitations do currently exist in the space of NGS; bottlenecks continue to plague the efficacy of bioinformatics tools that could otherwise be leveraged to their full potential. More often than not, the virus reads (e.g., SARS-CoV-2 reads) are usually low in genomic data; this extends to single-cell data as well. In this case, to perform detailed analysis, different approaches need to be utilized; for instance, a homology-based approach or a protein structure-based method. Furthermore, robust and time-sensitive analysis becomes a challenge with the ever-increasing magnitude of availability of full-length genome sequences of SARS-CoV-2, spanning hundreds of gigabytes. One approach is to randomly subsample data and analyses rather than performing a full phylogenetic analysis on the entire genome data, which would be time-consuming and produce extremely complex results that may not lead to mechanistic insights critical to understanding virus transmission mechanisms. Thus, it is critical to continue developing bioinformatics tools with long-term strategic applications in mind so that researchers can perform efficient and rapid data analysis and clinical trial deployment.

#### 3. Vaccine Status and Development

At least eighty infectious agents that manifest on a regular basis are known to be pathogenic in humans. Out of these eighty, more than thirty have licensed individual vaccines that target 26 of these infectious diseases—most of which are either viral or bacterial in nature. Given that vaccines have been instrumental in reducing mortality rates across the world at tremendous rates, it is also important to note that some of these vaccines are being used regularly and have been deployed to primarily circumvent childhood infections. Vaccination programs have been instrumental in reducing morbidity and mortality and inferring immunity to immunocompromised individuals and to children [58].

SARS-CoV-2 rapid evolution and transmission that continued unabated due to lack of herd immunity and lack of clinical drugs that could provide effective treatment meant that a mass vaccination program needed to be dispositioned immediately. Bioinformatics databases—which contain data with respect to nucleic acid sequences, protein sequences, ontologies, host databases, pathogen databases, and functional immunological databases (to name a few)—played a pivotal role in accelerating the development of novel mRNA vaccines that could shift the momentum that SARS-CoV-2 infection was gaining at a rapid pace. Further, the combined use of bioinformatics tools and machine-learning tools to make use of these databases was further monumental in guiding the precise development of the vaccines [38]. Further, accelerated research in these areas proved instrumental in the development of candidate vaccines (Figure 3). A variety of AI modalities were utilized for effective vaccine development and for evaluating the safety of such vaccines. A few of the vaccines that gained emergency use approval during the pandemic included: BNT162b2 from Pfizer-BioNTech, mRNA-1273 from Moderna, ChAdOx1 nCoV-19 from Oxford-AstraZeneca, and JNJ-78436735 from Johnson and Johnson [59].

Multipronged in silico approaches, including bioinformatics, vaccino-genomics, immunoinformatics, structural biology, and molecular dynamic simulations, are freely available to support precise vaccine design (Table 1). The sequenced SARS-CoV-2 genome is widely used in computational tools and databases for predicting novel B-cell and T-cell epitopes in vaccine development, immunity protein analysis, and immunization modeling. It deserves to be mentioned that immunoinformatics continues to be challenged by the accurate prediction of B-cell epitopes (BCE) and T-cell epitopes (TCE), which form the basis of the development of epitope-based vaccines [60].

**Table 1.** Categories with numerous bioinformatics tools have elucidated and assisted in the design, development, and deployment of potential vaccines in the fight against SARS-CoV-2 and its variants [27,39].

Areas Supporting Candidate Vaccine Development	Description	Examples
Genome Sequencing	Rapid processing and analysis of sequenced SARS-CoV-2 genomes aid in the rapid identification of mutations and thus aid in developing potential therapeutic targets and surveillance.	Trimmomatic, BWA, SAMTools, Seurat, GATK, SPAdes, Pangolin, IGV, Deseg2/EdgeR
Molecular Modeling	Various computational tools for molecular docking and molecular dynamics simulations facilitate the design and optimize therapeutic candidates.	PyMOL, Chimera, GROMACS, CHARMM, AMBER
Epitope Prediction	Numerous bioinformatics tools are being used to predict potential epitopes (antigenic determinants) that could be used to induce an immune response in the host and accelerate research/development.	NetMHC, IEDB, BepiPred, DiscoTope, Ellipro, ABCpred
Data Mining	Large-scale data mining and synthesizing of large-scale databases and information have speeded up the process of understanding virulence factors of new mutations, predicting their behavior, and creating appropriate drugs/vaccines to mitigate severe illness and curb spread, and inform public health policy.	Weka, Orange, KNIME, Cytoscape, TANGARA

During early March 2020, when the pandemic was gaining a foothold, a structurebased immunoinformatics methodology was employed to determine epitopes in SARS-CoV-2 for potential peptide-based vaccine design initiation. Wang and colleagues predicted 9 highly antigenic B-cell epitopes on the Spike (S) protein as well as 62 T-cell epitopes [61]. Reverse vaccinology (RV) is also a broadly used technique that employs computational methods to identify open reading frames (ORFs) from viral genomes, resulting in the discovery of new antigens [62]. The VaxiJen [63] server is used to study the physicochemical properties of antigen epitopes. Various machine-learning, as well as deep-learning models, have been employed to predict novel immunogenic subunits. These models have been crucial in identifying the epitopes for cytotoxic and helper T lymphocytes as well as in exploring the genetic polymorphism in the target human population [64,65]. Structural vaccinology techniques play a prominent role in identifying structurally stable, safe, and potent peptides as novel vaccine candidates [66]. Current bioinformatics tools such as Sprint [67], ModIAMP [68], pepATTRACT [69], PEPFOLD3 [70], IEDB [71], MHCPRED [72], SVMtrip [73], and AllerTop [42] have now been integrated into the programs for vaccine development. It has to be noted, though, the drop in values across all categories in 2022 is due to a multitude of factors, such as a significant decrease in publications due to delays in peer-review processes and timelines due to a lack of manpower/staff, since an over-whelming number of COVID-19-related studies were submitted. Furthermore, financial resources were diverted from genomic sequencing to more clinical measures of control for the pandemic—such as support for healthcare infrastructure, vaccine development, and medical supplies/equipment costs to mitigate impact [74,75]. The future of immunoinformatics is important, as it will determine the pace with which we will overcome the next pandemic, in case it manifests. The rapid design, development, and deployment pipelines for vaccines will need to be refined with iterative use, reuse, and refinement of existing open-source bioinformatics databases and tools, as well as the development of new tools/pipelines that will be created in the future. Thus, bioinformatics shows a promising role in the rapid development of potent vaccine candidates against SARS-CoV-2, fighting the global mortality of COVID-19.

#### 4. Discussion and Conclusions

Finally, a major milestone in the fight against the pandemic was reached on 5 May 2023, when the Director-General of the World Health Organization (WHO) declared that COVID-19 was no longer classified as a 'global health emergency' but rather an ongoing health concern [76]. While this announcement marked a significant achievement, it is important to acknowledge that SARS-CoV-2, the virus responsible for COVID-19, continues to undergo rapid evolution, adapting to its environment and potentially enhancing its ability to evade immune responses and spread. However, there are indications that these changes in the virus may come at the expense of its severity. The development of effective vaccines in record time has played a major role in modifying the evolutionary landscape and weakening the severe effects of the virus. While predicting the evolutionary path of the virus is still hard, consistent genomic surveillance will allow for early prediction of its transition. Vaccines and therapeutics still need to be updated to maintain immunity in the population and prevent hospitalization and death, especially in high-risk groups and children, while also preparing to put systems in place so that states are better prepared for future pandemics and global health emergencies.

Bioinformatics has played a transformational role in SARS-CoV-2 diagnosis and treatment through high-quality, well-curated pipelines, software, and datasets, allowing for rapid design, development, and deployment. These are important not only in drug discovery but also in drug development and allied processes. It entails using informatics to gain new insights into health and disease, managing data during clinical trials, and reusing clinical data. We discovered that bioinformatics tools were critical in bolstering efforts to combat the COVID-19 pandemic, identifying and informing the public about emerging variants and their potential phenotypic manifestations and analyzing the virus at both the sequence and structural levels, and resulting in the rapid development of vaccines and other effective therapeutics to combat severe complications and, more importantly, develop vaccines. Thus, bioinformatics will continue to play a pivotal role in shaping the next paradigm of health, advancing personalized medicine and pushing the boundaries of elucidating the genomic architecture of infectious diseases, and ensuring we emerge victorious in safeguarding public health against unprecedented infectious diseases.

Author Contributions: Conceptualization, N.S. (Nityendra Shukla), J.N. and P.S.; resources, J.N., N.S. (Nityendra Shukla) and P.S.; data curation, J.N., N.S. (Nityendra Shukla), N.S. (Neha Srivastava) and R.G.; writing—original draft preparation, N.S. (Nityendra Shukla), N.S. (Neha Srivastava), P.S. and J.N.; writing—review and editing, N.S. (Nityendra Shukla), N.S. (Neha Srivastava), P.S. and J.N.; visualization, N.S. (Nityendra Shukla) and R.G.; supervision, P.S. and J.N.; project administration, J.N. All authors have read and agreed to the published version of the manuscript.

**Funding:** The Rockefeller Foundation provided funding for this work. The CSIR Institute of Genomics and Integrative Biology received this grant, which enabled them to carry out the research and complete the study.

**Data Availability Statement:** Please visit the following GitHub page https://github.com/ns012/dr-ns (accessed 2 July 2023) to find scripts utilized for construction of the plots as well as data used in the process.

Conflicts of Interest: The authors declare no conflict of interest.

### References

- 1. Centers for Disease Control Control and Prevention. SARS-CoV-2 Variant Classifications and Definitions. 2023. Available online: https://www.cdc.gov/coronavirus/2019-ncov/variants/variant-classifications.html (accessed on 25 March 2023).
- WHO. Updated Working Definitions and Primary Actions for SARS-CoV-2 Variants. 2023. Available online: https://www. who.int/publications/m/item/updated-working-definitions-and-primary-actions-for{-}{-}sars-cov-2-variants (accessed on 14 June 2023).
- 3. World Health Organization. Statement on the Update of WHO's Working Definitions and Tracking System for SARS-CoV-2 Variants of Concern and Variants of Interest. 2023. Available online: https://www.who.int/news/item/16-03-2023-statement-on-the-update-of-who-s-working-definitions-and-tracking-system-for-sars-cov-2-variants-of-concern-and-variants-of-interest (accessed on 14 June 2023).
- Mathieu, E.; Ritchie, H.; Rodés-Guirao, L.; Appel, C.; Gavrilov, D.; Giattino, C.; Hasell, J.; Macdonald, B.; Dattani, S.; Beltekian, D.; et al. Coronavirus Pandemic (COVID-19). 2020. Available online: https://ourworldindata.org/covid-cases/ (accessed on 16 June 2023).
- 5. Davis, H.E.; Assaf, G.S.; McCorkell, L.; Wei, H.; Low, R.J.; Re'Em, Y.; Redfield, S.; Austin, J.P.; Akrami, A. Characterizing long COVID in an international cohort: 7 months of symptoms and their impact. *Eclinicalmedicine* **2021**, *38*, 101019. [CrossRef]
- Rambaut, A.; Loman, N.; Pybus, O.; Barclay, W.; Barrett, J.; Carabelli, A.; Connor, T.; Peacock, T.; Robertson, D.L.; Volz, E. Preliminary Genomic Characterisation of an Emergent SARS-CoV-2 Lineage in the UK Defined by a Novel Set of Spike Mutations. 2020. Available online: https://virological.org/t/preliminary-genomic-characterisation-of-an-emergent-sars-cov-2-lineagein-the-uk-defined-by-a-novel-set-of-spike-mutations/563 (accessed on 3 July 2023).
- Volz, E.; Mishra, S.; Chand, M.; Barrett, J.C.; Johnson, R.; Geidelberg, L.; Ferguson, N.M. Assessing transmissibility of SARS-CoV-2 lineage B.1.1.7 in England. *Nature* 2021, 593, 266–269. [CrossRef] [PubMed]
- 8. Weisblum, Y.; Schmidt, F.; Zhang, F.; DaSilva, J.; Poston, D.; Lorenzi, J.C.; Bieniasz, P.D. Escape from neutralizing antibodies by SARS-CoV-2 spike protein variants. *eLife* **2020**, *9*, e61312. [CrossRef] [PubMed]
- 9. Mlcochova, P.; Kemp, S.A.; Dhar, M.S.; Papa, G.; Meng, B.; Ferreira, I.A.T.M.; Datir, R.; Collier, D.A.; Albecka, A.; Singh, S.; et al. SARS-CoV-2 B.1.617.2 Delta variant replication and immune evasion. *Nature* **2021**, *599*, 114–119. [CrossRef] [PubMed]
- 10. Li, X.; Ma, X. Acute respiratory failure in COVID-19: Is it "typical" ARDS? Crit. Care 2020, 24, 198. [CrossRef] [PubMed]
- Khandia, R.; Singhal, S.; Alqahtani, T.; Kamal, M.A.; El-Shall, N.A.; Nainu, F.; Desingu, P.A.; Dhama, K. Emergence of SARS-CoV-2 Omicron (B.1.1.529) variant, salient features, high global health concerns and strategies to counter it amid ongoing COVID-19 pandemic. *Environ. Res.* 2022, 209, 112816. [CrossRef]
- 12. Hyams, C.; Challen, R.; Marlow, R.; Nguyen, J.; Begier, E.; Southern, J.; Szasz-Benczur, Z. Severity of Omicron (B.1.1.529) and Delta (B.1.617.2) SARS-CoV-2 infection among hospitalised adults: A prospective cohort study in Bristol, United Kingdom. *Lancet Reg. Health Eur.* **2023**, *25*, 100556. [CrossRef]
- 13. Carabelli, A.M.; Peacock, T.P.; Thorne, L.G.; Harvey, W.T.; Hughes, J. SARS-CoV-2 variant biology: Immune escape, transmission and fitness. *Nat. Rev. Microbiol.* 2023, 21, 162–177. [CrossRef]
- 14. Willett, B.J.; Grove, J.; MacLean, O.A.; Wilkie, C.; De Lorenzo, G.; Furnon, W.; Thomson, E.C. SARS-CoV-2 Omicron is an immune escape variant with an altered cell entry pathway. *Nat. Microbiol.* **2022**, *7*, 1161–1179. [CrossRef]
- 15. Meng, B.; Abdullahi, A.; Ferreira, I.A.; Goonawardane, N.; Saito, A.; Kimura, I.; Gupta, R.K. Altered TMPRSS2 usage by SARS-CoV-2 Omicron impacts infectivity and fusogenicity. *Nature* 2022, *603*, 706–714. [CrossRef] [PubMed]
- 16. Suzuki, R.; Yamasoba, D.; Kimura, I.; Wang, L.; Kishimoto, M.; Ito, J.; Sato, K. Attenuated fusogenicity and pathogenicity of SARS-CoV-2 Omicron variant. *Nature* 2022, *603*, 700–705. [CrossRef]
- Bentley, E.G.; Kirby, A.; Sharma, P.; Kipar, A.; Mega, D.F.; Bramwell, C.; Stewart, J.P. SARS-CoV-2 Omicron-B.1.1.529 Variant leads to less severe disease than Pango B and Delta variants strains in a mouse model of severe COVID-19. *BioRxiv* 2021, *Preprint*. [CrossRef]
- Bowe, B.; Xie, Y.; Al-Aly, Z. Acute and postacute sequelae associated with SARS-CoV-2 reinfection. *Nat. Med.* 2022, 28, 2398–2405. [CrossRef] [PubMed]
- Kedor, C.; Freitag, H.; Meyer-Arndt, L.; Wittke, K.; Hanitsch, L.G.; Zoller, T.; Scheibenbogen, C. A prospective observational study of post-COVID-19 chronic fatigue syndrome following the first pandemic wave in Germany and biomarkers associated with symptom severity. *Nat. Commun.* 2022, *13*, 5104. [CrossRef] [PubMed]

- Wang, C.; Liu, B.; Zhang, S.; Huang, N.; Zhao, T.; Lu, Q.B.; Cui, F. Differences in incidence and fatality of COVID-19 by SARS-CoV-2 Omicron variant versus Delta variant in relation to vaccine coverage: A world-wide review. *J. Med. Virol.* 2023, 95, e28118. [CrossRef] [PubMed]
- Yang, W.; Shaman, J.L. Author response: COVID-19 pandemic dynamics in South Africa and epidemiological characteristics of three variants of concern (Beta, Delta, and Omicron). *eLife* 2022, 11, e78933. [CrossRef]
- Ray, M.; Sable, M.N.; Sarkar, S.; Hallur, V. Essential interpretations of bioinformatics in COVID-19 pandemic. *Meta Gene* 2020, 27, 100844. [CrossRef]
- Guzzi, P.H.; Petrizzelli, F.; Mazza, T. Disease spreading modeling and analysis: A survey. *Brief. Bioinform.* 2022, 23, bbac230. [CrossRef]
- 24. Kumar Das, J.; Tradigo, G.; Veltri, P.; Guzzi, P.H.; Roy, S. Data science in unveiling COVID-19 pathogenesis and diagnosis: Evolutionary origin to drug repurposing. *Brief. Bioinform.* **2021**, *22*, 855–872. [CrossRef]
- Ma, J.; Deng, Y.; Zhang, M.; Yu, J. The role of multi-omics in the diagnosis of COVID-19 and the prediction of new therapeutic targets. *Virulence* 2022, 13, 1101–1110. [CrossRef]
- Siminea, N.; Popescu, V.; Martin, J.A.S.; Florea, D.; Gavril, G.; Gheorghe, A.-M.; Iţcuş, C.; Kanhaiya, K.; Pacioglu, O.; Popa, L.I.; et al. Network analytics for drug repurposing in COVID-19. *Brief. Bioinform.* 2021, 23, bbab490. [CrossRef]
- Hu, T.; Li, J.; Zhou, H.; Li, C.; Holmes, E.C.; Shi, W. Bioinformatics resources for SARS-CoV-2 discovery and surveillance. *Brief. Bioinform.* 2021, 22, 631–641. [CrossRef]
- Mehta, P.; Swaminathan, A.; Yadav, A.; Chattopadhyay, P.; Shamim, U.; Pandey, R. Integrative genomics important to understand host–pathogen interactions. *Brief. Funct. Genom.* 2022, elac021. [CrossRef]
- 29. Wood, D.E.; Lu, J.; Langmead, B. Improved metagenomic analysis with Kraken 2. Genome Biol. 2019, 20, 257. [CrossRef]
- Schloss, P.D.; Westcott, S.L.; Ryabin, T.; Hall, J.R.; Hartmann, M.; Hollister, E.B.; Lesniewski, R.A.; Oakley, B.B.; Parks, D.H.; Robinson, C.J.; et al. Introducing mothur: Open-Source, Platform-Independent, Community-Supported Software for Describing and Comparing Microbial Communities. *Appl. Environ. Microbiol.* 2009, 75, 7537–7541. [CrossRef] [PubMed]
- Bolyen, E.; Rideout, J.R.; Dillon, M.R.; Bokulich, N.A.; Abnet, C.C.; Al-Ghalith, G.A.; Alexander, H.; Alm, E.J.; Arumugam, M.; Asnicar, F.; et al. Reproducible, Interactive, Scalable and Extensible Microbiome Data Science using QIIME 2. *Nat. Biotechnol.* 2019, *37*, 852–857. [CrossRef] [PubMed]
- Hao, Y.; Hao, S.; Andersen-Nissen, E.; Mauck, W.M., 3rd; Zheng, S.; Butler, A.; Lee, M.J.; Wilk, A.J.; Darby, C.; Zager, M.; et al. Integrated analysis of multimodal single-cell data. *Cell* 2021, *184*, 3573–3587.e29. [CrossRef] [PubMed]
- Aran, D.; Looney, A.P.; Liu, L.; Wu, E.; Fong, V.; Hsu, A.; Chak, S.; Naikawadi, R.P.; Wolters, P.J.; Abate, A.R.; et al. Reference-based analysis of lung single-cell sequencing reveals a transitional profibrotic macrophage. *Nat. Immunol.* 2019, 20, 163–172. [CrossRef]
- 34. Fu, R.; Gillen, A.E.; Sheridan, R.M.; Tian, C.; Daya, M.; Hao, Y.; Riemondy, K.A. clustifyr: An R package for automated single-cell RNA sequencing cluster classification. *F1000Research* **2020**, *9*, 223. [CrossRef]
- Chen, J.; Wang, R.; Wang, M.; Wei, G.W. Mutations Strengthened SARS-CoV-2 Infectivity. J. Mol. Biol. 2020, 432, 5212–5226.
  [CrossRef]
- Bate, N.; Savva, C.G.; Moody, P.C.; Brown, E.A.; Evans, S.E.; Ball, J.K.; Brindle, N.P. In vitro evolution predicts emerging SARS-CoV-2 mutations with high affinity for ACE2 and cross-species binding. *PLoS Pathog.* 2022, 18, e1010733. [CrossRef]
- Hadfield, J.; Megill, C.; Bell, S.M.; Huddleston, J.; Potter, B.; Callender, C.; Sagulenko, P.; Bedford, T.; Neher, R.A. Nextstrain: Real-time tracking of pathogen evolution. *Bioinformatics* 2018, 34, 4121–4123. [CrossRef] [PubMed]
- Markov, P.V.; Ghafari, M.; Beer, M.; Lythgoe, K.; Simmonds, P.; Stilianakis, N.I.; Katzourakis, A. The evolution of SARS-CoV-2. Nat. Rev. Microbiol. 2023, 21, 361–379. [CrossRef]
- Hufsky, F.; Lamkiewicz, K.; Almeida, A.; Aouacheria, A.; Arighi, C.; Bateman, A.; Marz, M. Computational strategies to combat COVID-19: Useful tools to accelerate SARS-CoV-2 and coronavirus research. *Brief. Bioinform.* 2021, 22, 642–663. [CrossRef] [PubMed]
- Corman, V.M.; Landt, O.; Kaiser, M.; Molenkamp, R.; Meijer, A.; Chu, D.K.W.; Bleicker, T.; Brünink, S.; Schneider, J.; Schmidt, M.L.; et al. Detection of 2019 novel coronavirus (2019-nCoV) by real-time RT-PCR. *Eurosurveillance* 2020, 25, 2000045. [CrossRef] [PubMed]
- 41. Hoffmann, M.; Monaghan, M.T.; Reinert, K. PriSeT: Efficient de novo primer discovery. In Proceedings of the 12th ACM Conference on Bioinformatics, Computational Biology, and Health Informatics, Gaineswille, FL, USA, 1–4 August 2021; pp. 1–12.
- Sumon, T.A.; Hussain, A.; Hasan, T.; Hasan, M.; Jang, W.J.; Bhuiya, E.H.; Chowdhury, A.A.M.; Sharifuzzaman, S.M.; Brown, C.L.; Kwon, H.-J.; et al. A Revisit to the Research Updates of Drugs, Vaccines, and Bioinformatics Approaches in Combating COVID-19 Pandemic. *Front. Mol. Biosci.* 2021, 7, 585899. [CrossRef]
- 43. Kumar, V.M.; Pandi-Perumal, S.R.; Trakht, I.; Thyagarajan, S.P. Strategy for COVID-19 vaccination in India: The country with the second highest population and number of cases. *NPJ Vaccines* **2021**, *6*, 60. [CrossRef] [PubMed]
- Posada-Céspedes, S.; Seifert, D.; Topolsky, I.; Jablonski, K.P.; Metzner, K.J.; Beerenwinkel, N. V-pipe: A computational pipeline for assessing viral genetic diversity from high-throughput data. *Bioinformatics* 2021, 37, 1673–1680. [CrossRef]
- 45. Fritz, A.; Bremges, A.; Deng, Z.L.; Lesker, T.R.; Götting, J.; Ganzenmueller, T.; McHardy, A.C. Haploflow: Strain-resolved de novo assembly of viral genomes. *Genome Biol.* 2021, 22, 212. [CrossRef]
- 46. Gohl, D.M.; Garbe, J.; Grady, P.; Daniel, J.; Watson, R.H.; Auch, B.; Beckman, K.B. A rapid, cost-effective tailed amplicon method for sequencing SARS-CoV-2. *BMC Genom.* **2020**, *21*, 863. [CrossRef]

- Rosenthal, S.H.; Gerasimova, A.; Ruiz-Vega, R.; Livingston, K.; Kagan, R.M.; Liu, Y.; Lacbawan, F. Development and validation of a high throughput SARS-CoV-2 whole genome sequencing workflow in a clinical laboratory. *Sci. Rep.* 2022, *12*, 2054. [CrossRef]
- Schäffer, A.A.; Hatcher, E.L.; Yankie, L.; Shonkwiler, L.; Brister, J.R.; Mizrachi, I.K.; Nawrocki, E.P. VADR: Validation and annotation of virus sequence submissions to GenBank. *BMC Bioinform.* 2020, *21*, 211. [CrossRef]
  Lib. DV(C.D. for the VC. All and VC. VALUE V
- Libin, P.J.K.; Deforche, K.; Abecasis, A.B.; Theys, K. VIRULIGN: Fast codon-correct alignment and annotation of viral genomes. Bioinformatics 2018, 35, 1763–1765. [CrossRef]
- Kalvari, I.; Argasinska, J.; Quinones-Olvera, N.; Nawrocki, E.P.; Rivas, E.; Eddy, S.R.; Bateman, A.; Finn, R.D.; I Petrov, A. Rfam 13.0: Shifting to a genome-centric resource for non-coding RNA families. *Nucleic Acids Res.* 2018, 46, D335–D342. [CrossRef] [PubMed]
- 51. Bouckaert, R.; Heled, J.; Kühnert, D.; Vaughan, T.; Wu, C.-H.; Xie, D.; Suchard, M.A.; Rambaut, A.; Drummond, A.J. BEAST 2: A Software Platform for Bayesian Evolutionary Analysis. *PLoS Comput. Biol.* **2014**, *10*, e1003537. [CrossRef]
- Hoops, S.; Sahle, S.; Gauges, R.; Lee, C.; Pahle, J.; Simus, N.; Kummer, U. COPASI—A COmplex PAthway SImulator. *Bioinformatics* 2006, 22, 3067–3074. [CrossRef] [PubMed]
- Adam, D. Special report: The simulations driving the world's response to COVID-19. *Nature* 2020, 580, 316–318. [CrossRef]
  [PubMed]
- Singer, J.; Gifford, R.; Cotten, M.; Robertson, D. CoV-GLUE: A Web Application for Tracking SARS-CoV-2 Genomic Variation. Preprints.org 2020, 2020060225. [CrossRef]
- Sathyaseelan, C.; Magateshvaren Saras, M.A.; Prasad Patro, L.P.; Uttamrao, P.P.; Rathinavelan, T. CoVe-Tracker: An Interactive SARS-CoV-2 Pan Proteome Evolution Tracker. J. Proteome Res. 2023, 22, 1984–1996. [CrossRef]
- Cacciabue, M.; Aguilera, P.; Gismondi, M.I.; Taboga, O. Covidex: An ultrafast and accurate tool for SARS-CoV-2 subtyping. *Infect Genet. Evol.* 2022, 99, 105261. [CrossRef]
- O'Toole, Á.; Scher, E.; Underwood, A.; Jackson, B.; Hill, V.; McCrone, J.T.; Colquhoun, R.; Ruis, C.; Abu-Dahab, K.; Taylor, B.; et al. Assignment of Epidemiological Lineages in an Emerging Pandemic Using the Pangolin Tool. *Virus Evol.* 2021, 7, veab064. [CrossRef]
- 58. Flower, D.R. Bioinformatics for Vaccinology; John Wiley & Sons: Hoboken, NJ, USA, 2008; ISBN 9780470027110.
- WHO. COVID-19 Vaccines with WHO Emergency Use Listing. Available online: https://extranet.who.int/pqweb/vaccines/ vaccinescovid-19-vaccine-eul-issued (accessed on 15 June 2023).
- Wang, D.; Mai, J.; Zhou, W.; Yu, W.; Zhan, Y.; Wang, N.; Yang, Y. Immunoinformatic Analysis of T- and B-Cell Epitopes for SARS-CoV-2 Vaccine Design. *Vaccines* 2020, *8*, 355. [CrossRef] [PubMed]
- 61. Ullah, A.; Sarkar, B.; Islam, S.S. Exploiting the reverse vaccinology approach to design novel subunit vaccines against Ebola virus. *Immunobiology* **2020**, 225, 151949. [CrossRef] [PubMed]
- 62. Doytchinova, I.A.; Flower, D.R. VaxiJen: A server for prediction of protective antigens, tumour antigens and subunit vaccines. BMC Bioinform. 2007, 8, 4. [CrossRef]
- Bukhari SN, H.; Jain, A.; Haq, E.; Mehbodniya, A.; Webber, J. Machine Learning Techniques for the Prediction of B-Cell and T-Cell Epitopes as Potential Vaccine Targets with a Specific Focus on SARS-CoV-2 Pathogen: A Review. *Pathogens* 2022, 11, 146. [CrossRef] [PubMed]
- 64. Yang, X.; Zhao, L.; Wei, F.; Li, J. DeepNetBim: Deep learning model for predicting HLA-epitope interactions based on network analysis by harnessing binding and immunogenicity information. *BMC Bioinform.* **2021**, 22, 231. [CrossRef]
- Kames, J.; Holcomb, D.D.; Kimchi, O.; DiCuccio, M.; Hamasaki-Katagiri, N.; Wang, T.; Kimchi-Sarfaty, C. Sequence analysis of SARS-CoV-2 genome reveals features important for vaccine design. *Sci. Rep.* 2020, 10, 15643. [CrossRef]
- 66. Taherzadeh, G.; Yang, Y.; Zhang, T.; Liew, A.W.; Zhou, Y. Sequence-based prediction of protein–peptide binding sites using support vector machine. *J. Comput. Chem.* **2016**, *37*, 1223–1229. [CrossRef]
- 67. Müller, A.T.; Gabernet, G.; Hiss, J.A.; Schneider, G. modlAMP: Python for antimicrobial peptides. *Bioinformatics* 2017, 33, 2753–2755. [CrossRef]
- De Vries, S.J.; Rey, J.; Schindler, C.E.M.; Zacharias, M.; Tuffery, P. The pepATTRACT web server for blind, large-scale peptide– protein docking. *Nucleic Acids Res.* 2017, 45, W361–W364. [CrossRef]
- 69. Lamiable, A.; Thévenet, P.; Rey, J.; Vavrusa, M.; Derreumaux, P.; Tufféry, P. PEP-FOLD3: Faster de novo structure prediction for linear peptides in solution and in complex. *Nucleic Acids Res.* **2016**, *44*, W449–W454. [CrossRef]
- 70. Vita, R.; Mahajan, S.; Overton, J.A.; Dhanda, S.K.; Martini, S.; Cantrell, J.R.; Wheeler, D.K.; Sette, A.; Peters, B. The Immune Epitope Database (IEDB): 2018 update. *Nucleic Acids Res.* **2019**, *47*, D339–D343. [CrossRef] [PubMed]
- 71. Guan, P.; Hattotuwagama, C.K.; Doytchinova, I.A.; Flower, D.R. MHCPred 2.0: An updated quantitative T-cell epitope prediction server. *Appl. Bioinform.* 2006, *5*, 55–61. [CrossRef] [PubMed]
- 72. Yao, B.; Zhang, L.; Liang, S.; Zhang, C. SVMTriP: A Method to Predict Antigenic Epitopes Using Support Vector Machine to Integrate Tri-Peptide Similarity and Propensity. *PLoS ONE* **2012**, *7*, e45152. [CrossRef]
- Dimitrov, I.; Bangov, I.; Flower, D.R.; Doytchinova, I. AllerTOP v.2—A server for in silico prediction of allergens. *J. Mol. Model.* 2014, 20, 2278. [CrossRef]
- 74. Harper, L.; Kalfa, N.; Beckers, G.; Kaefer, M.; Nieuwhof-Leppink, A.; Fossum, M.; Herbst, K.; Bagli, D. The impact of COVID-19 on research. *J. Pediatr. Urol.* 2020, *16*, 715–716. [CrossRef]

- 75. Venkatesh, V. Impacts of COVID-19: A research agenda to support people in their fight. *Int. J. Inf. Manag.* 2020, 55, 102197. [CrossRef] [PubMed]
- 76. WHO. Statement on the Fifteenth Meeting of the IHR (2005) Emergency Committee on the COVID-19 Pandemic. Available online: https://www.who.int/news/item/05-05-2023-statement-on-the-fifteenth-meeting-of-the-international-health-regulations-(2005)-emergency-committee-regarding-the-coronavirus-disease-(covid-19)-pandemic (accessed on 6 July 2023).

**Disclaimer/Publisher's Note:** The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.