Special Edition on "Frontiers of Biosciences in Sustainable Development (FBSD)"
CIBTech Journal of Zoology ISSN: 2319–3883; http://www.cibtech.org.htm
2023 Vol.12/S1, pp.232-238/Hiba and Asifa
***Research Article** (Open Access)*

IN SILICO ANALYSIS OF THE GENOMES OF SARS-CoV-2 VARIANTS OF CONCERN (VOC) REPORTED IN INDIA, WITH RESPECT TO THE REFERENCE GENOME AT NUCLEOTIDE LEVEL

M. Hiba and Asifa Ahmmed*
P.G. and Research Department of Zoology, Justice Basheer Ahmed Sayeed College for Women (Autonomous), Teynampet, Chennai-600018, Tamil Nadu, India
**Author for correspondence: asifaahmmed62@gmail.com*

ABSTRACT
The SARS-CoV-2 virus responsible for the major outbreak of COVID-19 disease globally belongs to a large family of viruses called the coronaviruses and accumulation of novel mutations in their genome has led to the emergence of new variants of the virus which has caused increased morbidity and mortality.
This study aims to present an in-silico analysis of the genomes of SARS-CoV-2 variants of concern (VOC) reported in India with respect to the reference genome of SARS-COV-2 at a nucleotide level by performing pairwise sequence alignment using the NCBI BLAST tool. The VOCs taken for this study are delta variant (B.1.617.2) and sub-lineages of omicron variant – BA.1, BA.2, BA.1.1 From the pairwise sequence alignment of the VOCs with the reference genome, we are able to determine their degree of similarity (percent identity) to the reference genome and also the single nucleotide polymorphisms (SNPs) in the genomes of VOCs, which greatly influences their structural and functional properties. From this study, the omicron sub-variant BA.2 is concluded to have the closest nucleotide sequence homology with the reference genome with a percent identity of 99.88% and BA.1.1 sub-variant is quite distantly related with a percent identity of 99.81%, to the reference genome than compared to the other variants taken in this study. Such in-silico analysis of the genomes of variants can help us to identify the mutations in their genomes that causes changes in their structural and functional properties. Hence this provides us the information based on which we can deduce effective preventive strategies, diagnostic tools and therapeutic approaches to keep the widespread of the disease in check.

Keywords: *In-silico analysis, SARS-CoV-2, Variants of concern (VOC), Pairwise sequence alignment, single nucleotide polymorphisms (SNPs)*

INTRODUCTION

The 2019 novel coronavirus (2019-nCoV/SARS-CoV-2) originally arose as part of a major outbreak of respiratory disease centred on Hubei province, China. It has caused a global pandemic and is a major public health concern. Taxonomically, SARS-CoV-2 was shown to be a Beta coronavirus (lineage B) closely related to SARS-CoV and SARS-related bat coronaviruses (Temmam *et al*., 2022), and it has been reported to share a common receptor with SARS-CoV (ACE-2). Subsequently, Beta coronaviruses from pangolins were identified as close relatives to SARS-CoV-2. The virus spreads via the droplet released from an infected person's mouth or nasal cavity. COVID-19 affects different people in different ways and the most common symptoms include fever, cough, tiredness, loss of taste or smell, sore throat and red or irritated eyes.
The word corona means crown and refers to the appearance that these viruses get from the spike proteins on their surface. The spike protein is the part of the virus that attaches to a human cell to infect it, allowing it to replicate inside of the cell and spread to other cells. Some antibodies can protect us from SARS-CoV-2 by targeting these spike proteins. Because of the importance of this specific part of the virus, scientists who sequence the virus for research constantly monitor mutations causing changes to the spike protein

through a process called genomic surveillance. As genetic changes to the virus happen over time, the SARS-CoV-2 virus begins to form genetic lineages. Just as a family has a family tree, the SARS-CoV-2 virus can be similarly mapped out. Sometimes branches of that tree have different attributes that change how fast the virus spreads, or the severity of illness it causes, or the effectiveness of treatments against it. Scientists call the viruses with these changes' "variants". They are still SARS-CoV-2 but may act differently (Houtman *et al*., 2022).

A mutation also referred to as viral mutation or genetic mutation of the severe acute respiratory syndrome coronavirus 2 (SARS – CoV-2) is a change in the genetic sequence of the SARS – CoV-2 virus when compared with a reference genome from Wuhan – Hubei (The first genetic sequence identified). A new variant (Virus variant or genetic variant or sub-lineage of SARS – CoV-2) may have one or more mutations that is differentiated form the reference sequence and these variants of SARS – CoV-2 can have different characteristics. All viruses, including SARS-CoV-2, the virus that causes COVID-19, change over time. Most changes have little to no impact on the virus' properties. However, some changes may affect the virus's properties, such as how easily it spreads, the associated disease severity, or the performance of vaccines, therapeutic medicines, diagnostic tools, or other public health and social measures.

Given the continuous evolution of the virus that leads to SARS-CoV-2 and the constant developments in our understanding of the impacts of variants, these working definitions may be periodically adjusted. When necessary, variants not otherwise meeting all criteria outlined in these definitions may be designated as Variants of concern (VOC), variant of interest (VOI), variants under monitor (VUM), and those posing a diminishing risk relative to other circulating variants may be reclassified, in consultation with the Technical Advisory Group on Virus Evolution.

This study aims to present an in-silico analysis of the SARS-CoV-2 genomes of the variants of concern (VOC) observed in India with respect to the reference genome obtained from Wuhan, China, at the nucleotide sequence level by performing pairwise sequence alignments.

The objective of this study is to compare and analyse the genomes of the SARS-CoV-2 variants of concern (VOC) with respect to the reference genome at nucleotide level by performing pairwise sequence alignment to determine their percent identity i.e., their degree of similarity with the reference genome which depicts their evolutionary relationship and also to determine the positions of mutations in the nucleotide sequences of the variants of concern, which has a great influence on their structural and functional properties like structure of their receptor binding domain, spike glycoprotein, their transmissibility, disease severity, risk of reinfection, and impacts on diagnostics and vaccine performance etc.

MATERIALS AND METHODS

The NCBI virus database was used for the retrieval of the nucleotide sequences of the reference genome and the genomes of the variants of concern (VOC) of SARS-CoV –2 taken for analysis in this study. It is an integrative, value-added resource supporting retrieval, display and analysis of a curated collection of virus sequences and large sequence datasets.

The NCBI BLAST (basic local alignment search tool) an algorithm and program for comparing primary biological sequence information, such as the amino-acid sequences of proteins or the nucleotides of DNA and/or RNA sequences was used to perform the pairwise sequence alignment of the nucleotide sequences of the genomes of VOC with that of the reference genome.

The nucleotide sequence of the reference genome and the genome sequences of the SARS-CoV-2 VOC observed to be prevalent in India was retrieved from the NCBI Virus database in FASTA format.

Reference genome -Accession id: NC_045512.2

The SARS-CoV-2 variants of concern (VOC) taken for analysis in this study are:

1. Delta variant-B.1.617.2

Accession id: OM918219

i. Omicron variant sub-lineages: BA.1

Accession id: ON063252

ii.BA.2
Accession id: ON060017
iii.BA.1.1
Accession id: ON063244

The data about the percent identity of the genome of each variant with the reference genome, the query cover percentage, the number of nucleotide identities, number of gaps in the alignment, number of mismatches of nucleotide base pairs (due to point mutations) and the positions or coordinates of the mutation in the nucleotide sequences of the genome of each variant of concern was obtained by performing the pairwise sequence alignment using the tool NCBI Blast.

The nucleotide sequences of the reference genome of SARS-CoV-2 and the variants of concern – B.1.617.2, BA.1, BA.2, BA.1.1 were retrieved from the NCBI Virus database as a FASTA file and the tool NCBI BLAST was used for performing the pairwise sequence alignment between the two genome sequences taken at a time. The 'Blastn' program is selected. The option 'Align two or more nucleotide sequences' is selected, which then provides two entry fields namely, 'query sequence' and 'subject sequence' to upload the nucleotide sequences of the two genomes to be aligned. The genomes of the VOCs are taken as query sequences and the reference genome of SARS-CoV-2 as the subject sequence. The FASTA file of the reference genome and the each of the VOC taken at a time is uploaded and the BLAST program is run.

RESULTS AND DISCUSSION

The following are the results for the nucleotide sequence pairwise alignment of the variants of concern with the reference genome:

1. B.1.617.2:

The percent identity obtained for the nucleotide sequence pairwise-alignment of the genome of delta variant (B.1.617.2) with the reference genome is 99.85%.

- Query cover: 100%
- Identities: 29785/29831
- Gaps: 13/29831
- Mismatches: 33

2. BA.2:

The percent identity obtained for the nucleotide sequence pairwise-alignment of the genome of omicron variant-BA.2 with the reference genome is 99.88%.

- Query cover: 99%
- Identities: 20656/20680
- Gaps: 0/20680
- Mismatches: 24

3.BA.1:

The percent identity obtained for the nucleotide sequence pairwise-alignment of the genome of omicron variant-BA.1 with the reference genome is 99.81%

- Query cover: 100%
- Identities: 29682/29738
- Gaps: 0/29738
- Mismatches: 56

4. BA.1.1:

The percent identity obtained for the nucleotide sequence pairwise-alignment of the genome of omicron variant-BA.1.1 with the reference genome is 99.81%.

- Query cover: 100%
- Identities: 29723/29780
- Gaps: 0/29738
- Mismatches: 57

Table 1: Summary table of the pairwise sequence alignment performed for variants of concern with respect to the reference genome at nucleotide level

VOC	No. of nucleotide Identities	Total Aligned Nucleotides	Gaps	Mismatches
B.1.617.2	29785	29831	13	33
BA.2	20656	20680	0	24
BA.1	29682	29738	0	56
BA.1.1	29723	29780	0	57

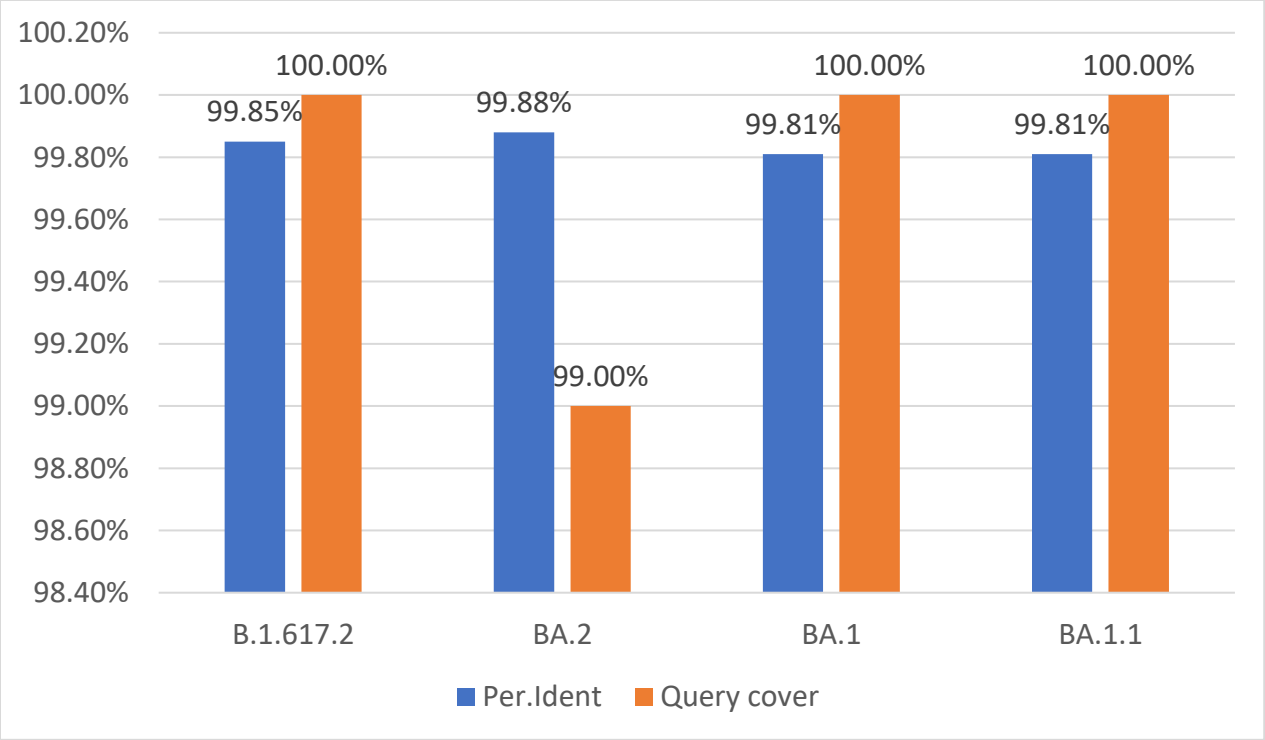

Figure 1: Graph representing the query coverage and percent identity of the VOCs with the reference genome.

From this in silico analysis of the genomes of SARS-CoV-2 variants of concern with respect to the reference genome, we are able to determine the degree of similarity of their nucleotide sequences which reflects about their evolutionary relationship. We are also able to determine the SNPs in the genome of these variants (Koyama *et al*., 2020). Such genomic analysis of VOC helps us to monitor the changes in their genetic code and have a better understanding of how these variants might impact public health.

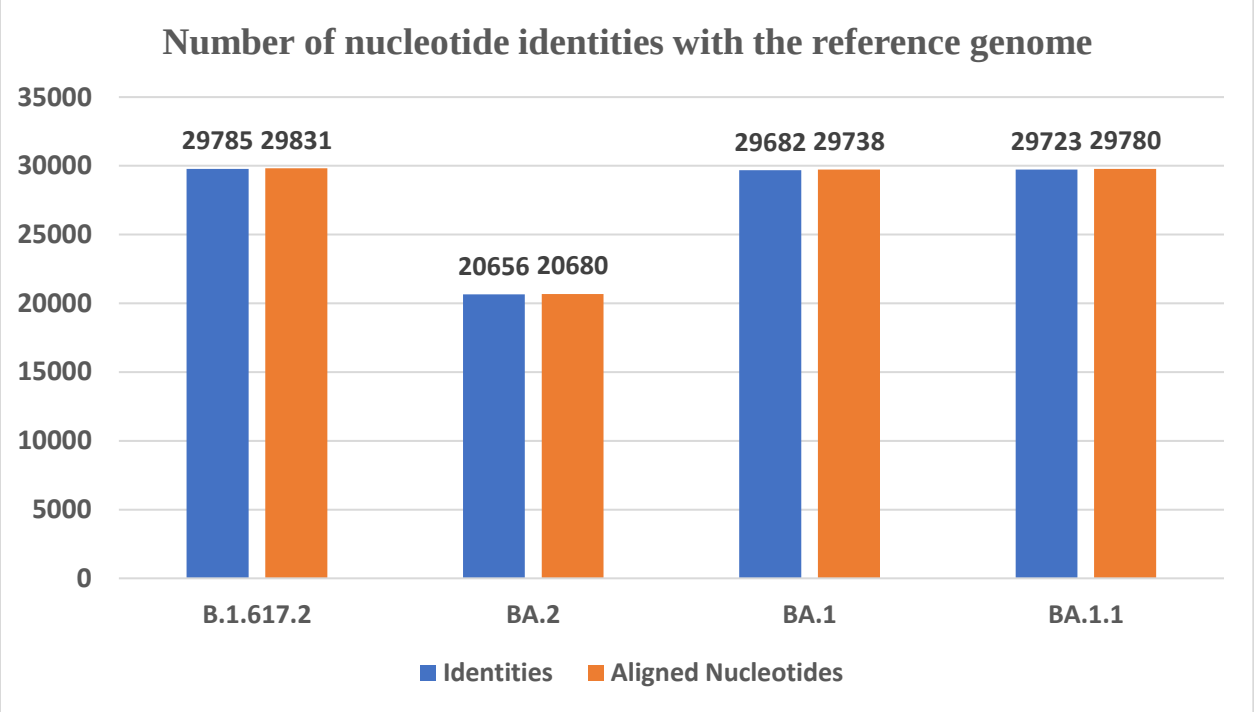

Figure 2: Graph representing the total aligned nucleotides and number of nucleotide identities of each of the VOCs with the reference genome.

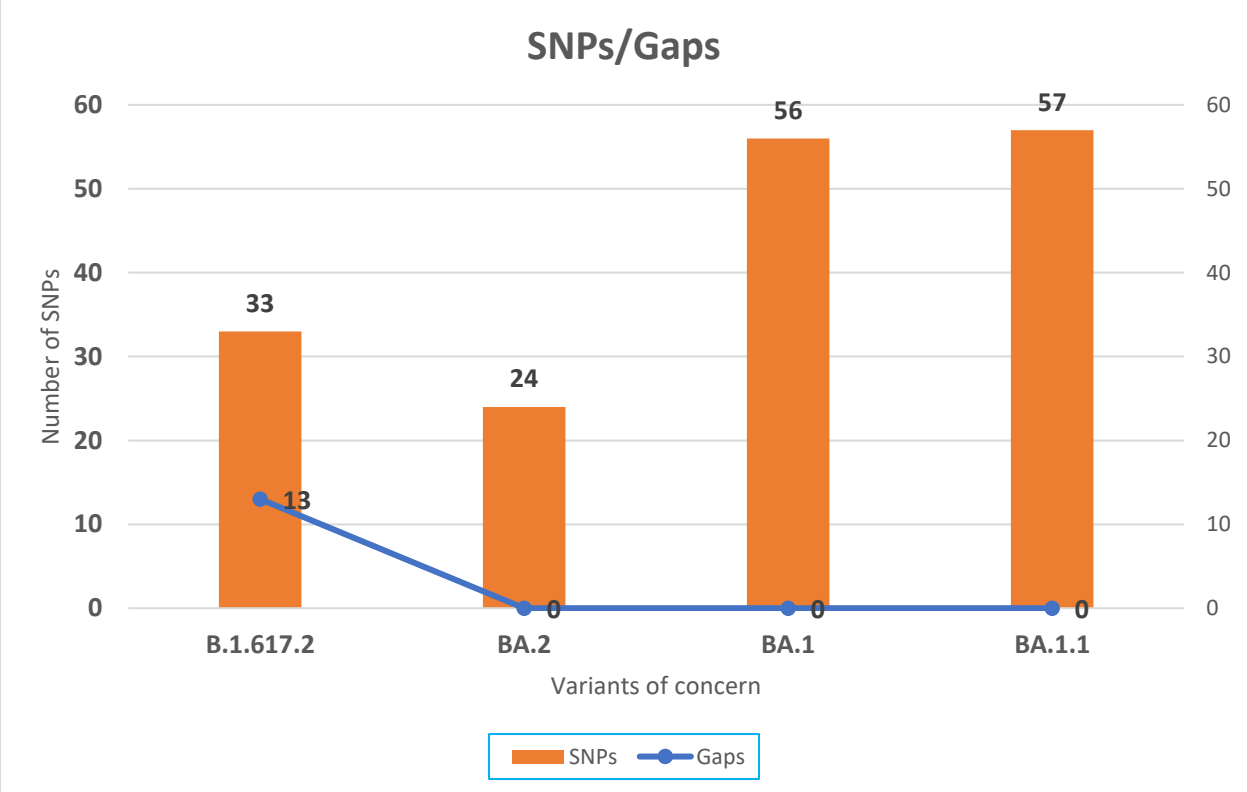

Figure 3: Graph showing comparison of the number of SNPs and gaps of the VOCs with respect to the reference genome.

CONCLUSION

By interpreting the results from the pairwise sequence alignment of the variants with the reference genome, it can be concluded that the omicron variant – BA.2 has the closest nucleotide sequence homology with the reference genome of SARS-CoV-2, as it has highest percent identity of 99.88%, with 20656/20680 nucleotide identities and 0/20680 gap region. Hence it is concluded to be the most closely related genome to the reference genome of SARS-CoV-2 than the other variants taken for analysis in this study.

The omicron sub-variant BA.1.1 has comparatively lesser percent identity of 99.81% with the reference genome than the other variants taken for analysis in this study. It has 29723/29780 identities, 0/29780 gaps and maximum number of mismatches of 57 with respect to the reference genome, due to point mutations. Therefore, we conclude that the omicron sub-variant BA.1.1 is quite distantly related to the reference genome than the other variants of concern of SARS-CoV-2 taken for study.

RECOMMENDATIONS

In silico analysis of the genome of these variants at amino acid sequence level can reveal to us about the mutations in their protein sequences from which we can understand their structural and functional properties which can hint towards the susceptible antigen targets of SARS-CoV-2, and this can help us come up with potential therapeutics and prophylactic interventions for the prevention of this public threat of SARS-CoV-2.

ACKNOWLEDGEMENT

We sincerely thank Dr. Amthul Azeez, Principal, the Head of the Department of Zoology, Justice Basheer Ahmed Sayeed College for Women, for her support and encouragement in conducting this study. We also extend our sincere and heartfelt obligation to our friends and family who provided us moral support in completing this study.

REFERENCES

Alkhatib M, Salpini R, Carioti L, Ambrosio FA, Anna S, Duca L, Costa G, Bellocchi MC, Piermatteo L, Artese A, Santoro MM, Alcaro S, Svicher V, & Ceccherini-Silberstein F (2022). Update on SARS-COV-2 omicron variant of concern and its peculiar mutational profile. *Microbiology Spectrum*, **10**(2).

Chakraborty S, Devendran R, & Kumar M (2020). Genome analysis of SARS-COV-2 isolates occurring in India: Present scenario. *Indian Journal of Public Health*, **64**(6), 147.

Chang TJ, Yang DM, Wang ML, Liang KH, Tsai PH, Chiou SH, Lin TH, & Wang CT (2020). Genomic analysis and comparative multiple sequences of SARS-cov2. *Journal of the Chinese Medical Association*, **83**(6), 537–543.

Jaroszewski L, Iyer M, Alisoltani A, Sedova M, & Godzik A (2020). The interplay of SARS-COV-2 evolution and constraints imposed by the structure and functionality of its proteins.

Jia Y, Shen G, Nguyen S, Zhang Y, Huang KS, Ho HY, Hor WS, Yang CH, Bruning JB, Li C, & Wang WL (2020). Analysis of the mutation dynamics of SARS-COV-2 reveals the spread history and emergence of RBD mutant with lower Ace2 binding affinity.

Kim JS, Jang JH, Kim JM, Chung YS, Yoo CK, & Han MG (2020). Genome-wide identification and characterization of point mutations in the SARS-COV-2 genome. *Osong Public Health and Research Perspectives*, **11**(3), 101–111.

Koyama T, Platt D, & Parida L (2020). Variant analysis of SARS-COV-2 genomes. *Bulletin of the World Health Organization*, **98**(7), 495–504.

Kumar A, Asghar A, Singh HN, Faiq MA, Kumar S, Narayan RK, Kumar G, Dwivedi P, Sahni C, Jha RK, Kulandhasamy M, Prasoon P, Sesham K, Kant K, & Pandey SN (2021). An in-silico analysis of early SARS-COV-2 variant B.1.1.529 (omicron) genomic sequences and their epidemiological correlates.

Padhan K, Parvez MK, & Al-Dosari MS (2021). Comparative sequence analysis of SARS-COV-2 suggests its high transmissibility and pathogenicity. *Future Virology*, **16**(3), 245–254.
Rahman MS, Islam MR, Hoque MN, Alam AS, Akther M, Puspo JA, Akter S, Anwar A, Sultana M, & Hossain MA (2020). Comprehensive annotations of the mutational spectra of SARS-COV-2 spike protein: A fast and accurate pipeline.
Saha I, Ghosh N, Maity D, Sharma N, & Mitra K (2020). Inferring the genetic variability in Indian sars-COV-2 genomes using consensus of multiple sequence alignment techniques. *Infection, Genetics and Evolution*, **85**, 104522.
Said KB, Alsolami A, Fathuldeen A, Alshammari F, Alhiraabi W, Alaamer S, Alrmaly H, Aldamadi F, Aldakheel DF, Moussa S, Jadani AA, & Bashir A (2021). In-silico pangenomics of SARS-COV-2 isolates reveal evidence for subtle adaptive expression strategies, continued clonal evolution, and sub-clonal emergences, despite genome stability. *Microbiology Research*, **12**(1), 204–233.
Tang X, Wu C, Li X, Song Y, Yao X, Wu X, Duan Y, Zhang H, Wang Y, Qian Z, Cui J, & Lu J (2020). On the origin and continuing evolution of SARS-COV-2. *National Science Revie*w, **7**(6), 1012–1023.
Yamasoba D, Kimura I, Nasser H, Morioka Y, Nao N, Ito J, Uriu K, Tsuda M, Zahradnik J, Shirakawa K, Suzuki R, Kishimoto M, Kosugi Y, Kobiyama K, Hara T, Toyoda M, Tanaka YL, Butlertanaka EP, Shimizu R, Sato K (2022). Virological characteristics of SARS-COV-2 ba.2 variant.
Yang HC, Chen C, Wang JH, Liao HC, Yang CT, Chen CW, Lin YC, Kao CH, Lu MYJ, & Liao JC (2020). Analysis of genomic distributions of SARS-COV-2 reveals a dominant strain type with strong allelic associations. *Proceedings of the National Academy of Sciences*, **117**(48), 30679–30686.