Frontiers in Multidisciplinary Research and Development

Journal Homepage: https://samjohnparr.github.io/Frontiers-Journal

Topological Characterization and Phylogenetic Analysis of SARS-CoV-2 Genomic RNA Using Chaos Geometry and Persistent Homology

Jun Nathaniel G. Ratilla[a,*], Perie T. Miego[a], Nilmae Love P. Ngujo[a]

[a] *Senior High School Department, Padada National High School, Padada, Davao del Sur, 8007, Philippines*
[*]Corresponding author: ratillajunnathaniel@gmail.com

Article Info

Article history:

Received May 13, 2025
Revised July 24, 2025
Accepted August 1, 2025

Keywords:

SARS-CoV-2
Persistent homology
Chaos geometry
Phylogenetic analysis
Variants of concern

ABSTRACT

The SARS-CoV-2 pandemic produced multiple Variants of Concern that continue to complicate surveillance and control. While sequence-based pipelines identify mutations with precision, they may miss mesoscopic structural patterns along the genome. This study applies topological data analysis (TDA) to evaluate whether topology-derived features add complementary information for distinguishing variants. Complete RNA sequences for Alpha, Beta, Gamma, Delta, and Omicron were retrieved from GenBank and embedded as four-dimensional point clouds using a chaos-geometry mapping. Persistent homology was computed on these point clouds to obtain persistence diagrams that summarize zero- and one-dimensional topological features, and pairwise Wasserstein distances between diagrams were used to quantify structural similarity. A topology-informed phylogeny derived from the distance matrix was compared with a tree produced by a standard alignment method. The topology-based analysis grouped B.1.351 (Beta) with B.1.617.2 (Delta), indicating substantial structural similarity, and suggested that B.1.1.7 (Alpha) diverged earliest; several branching orders differed from the alignment-based tree, implying that topology captures geometric and correlation structure not reflected by character-wise alignments. These findings show that TDA provides a complementary lens on genomic organization, with potential to inform vaccine design, guide variant-specific diagnostics, and enhance genomic surveillance when integrated with conventional sequence analytics.

1. INTRODUCTION

Recent advancements in methods for analyzing genomic sequence similarities have significantly enhanced our understanding of genetic information (Pereira et al., 2020). These methods are categorized into graphical and numeric representations. Graphical methods map each nucleotide base (A, C, G, T) to a point in Euclidean space $\mathbb{R}^n$, where n ranges from 2 to 5, creating a collection of points representing the genomic sequence (Hoang et al., 2016). Techniques like signal processing use measures such as Euclidean distances or correlation angles to compare sequence similarities (Akhtar et al., 2008). Numeric representations convert genomic sequences into numeric sequences, facilitating various comparison techniques.

Graphical methods face challenges such as geometric inaccuracies and poor performance with varying sequence lengths (Brzinsky-Fay, 2014). To address these, this paper introduces a novel method combining graphical representation with algebraic topology tools, specifically persistent homology (PH). This method transforms each genomic sequence into a collection of points in Euclidean space $\mathbb{R}^n$, viewed as point clouds.

Persistent homology extracts topological signatures, represented by persistence diagrams that encode features like the number of connected components or 1-dimensional holes. The Wasserstein distance between persistence diagrams reflects similarities among genomic sequences.

Topological data analysis (TDA) offers powerful tools for understanding the shape and structure of data (Wasserman, 2018; Edelsbrunner & Harer, 2022). Persistent homology can uncover these hidden structural relationships by identifying topological features like loops and voids in genomic data. This approach allows us to identify variants with similar structural characteristics, potentially offering insights into transmissibility, immune escape, and other functional traits beyond genetic mutations. Persistent homology captures topological features across multiple scales, making it useful for analyzing complex biological data. It has been successfully applied to various biological and non-biological datasets, demonstrating its versatility and effectiveness (Wei, 2017; Meng et al., 2020; Aggarwal & Periwal, 2023; De Lara, 2023; Parreño & Anter, 2024).

This study applies these methodologies to analyze the positive-sense single-stranded genomic RNA (ssRNA) of SARS-CoV-2, focusing specifically on the Variants of Concern (VOC) as defined by the World Health Organization (WHO) (Parums, 2021; Gowrisankar et al., 2022) due to their increased transmissibility or severity. The dataset used in this study is retrieved from the NCBI Genbank, ensuring a comprehensive and authoritative collection of SARS-CoV-2 VOC sequences. Positive-sense ssRNA viruses possess genomes that serve as mRNA for protein synthesis, making their analysis crucial for understanding viral replication and pathogenicity (Poltronieri et al., 2015). Given the global impact of COVID-19, advanced methods for analyzing SARS-CoV-2 genomic sequences are essential for developing effective diagnostics, treatments, and vaccines.

This paper presents a topological characterization of SARS-CoV-2 genomic RNA sequences using chaos geometry in 4-dimensional space and persistent homology. Turning each RNA sequence into a collection of points in 4-dimensional Euclidean space creates point clouds that capture topological properties. Persistent homology is applied to these point clouds to obtain persistence diagrams encoding topological signatures. The Wasserstein distance between persistence diagrams measures genomic similarity, allowing the construction of phylogenetic trees and comparing SARS-CoV-2 sequences effectively.

This study addresses existing methods' limitations and provides a comprehensive framework for genomic analysis. By integrating chaos geometry and persistent homology, this study offers a novel perspective on the topological characteristics of SARS-CoV-2 RNA sequences, contributing to understanding and combating COVID-19.

The rest of the paper is organized as follows: introduce a new 4-dimensional representation of positive-sense single-stranded genomic RNA sequences based on chaos geometry, review basic concepts related to persistent homology, formally introduce the method, apply it to analyze SARS-CoV-2 genomic sequences, produce a phylogenetic tree, and comparing the results with other state-of-the-art methods such as Clustal Omega.

2. METHOD

2.1 Dataset

This study analyzes SARS-CoV-2 Variants of Concern (VOCs), identified by the World Health Organization (WHO) due to their increased transmissibility, severity, or potential for immune escape. The dataset consists of complete genomic sequences of these variants retrieved from the NCBI GenBank database.

Although SARS-CoV-2 has a positive-sense single-stranded RNA genome, the sequences used in this study are represented using the nucleotide bases A (Adenine), C (Cytosine), G (Guanine), and T (Thymine), which correspond to DNA sequences. This is because, for many viral studies, RNA sequences are often reverse-transcribed into complementary DNA (cDNA) before being sequenced and stored in databases like NCBI. The reverse transcription process converts RNA nucleotides into their DNA counterparts, where Uracil (U) in RNA is replaced by Thymine (T) in the resulting cDNA. This cDNA is then used for storage, analysis, and alignment

purposes, which is why the sequences in this study are represented using ACGT (Wu et al., 2020). The variants examined in this study include Alpha (B.1.1.7), Beta (B.1.351), Gamma (P.1), Delta (B.1.617.2), and Omicron (B.1.1.529). Multiple sequence entries were collected for each variant to ensure comprehensive analysis. The specific accession numbers for these sequences, obtained from the NCBI Genbank, are as follows: Alpha (B.1.1.7) [OV054768], Beta (B.1.351) [OR278055], Gamma (P.1) [OR578388], Delta (B.1.617.2) [OK091006], and Omicron (B.1.1.529) [OQ905474].

2.2 Chaos 4-Dimensional Representation

This paper utilized a 4-dimensional chaotic system to convert an RNA sequence into a finite set of points within $\mathbb{R}^4$, creating a 4-dimensional representation of the sequence (Feldman, 2012). Specifically, let α represent an RNA sequence of length n, written as $\beta_1, \beta_2, \dots, \beta_n$, where β_i refers to one of the four nucleotide bases A, C, G, or T. The next step is to map $\beta_1, \beta_2, \dots, \beta_n$ to a sequence of integers $(a_1, a_2, \dots, a_n)$, where

$$a_i = \begin{cases} 1 & \text{if } \beta_i \text{ is nucleotide A,} \\ 2 & \text{if } \beta_i \text{ is nucleotide C,} \\ 3 & \text{if } \beta_i \text{ is nucleotide G,} \\ 4 & \text{if } \beta_i \text{ is nucleotide T,} \end{cases}$$

Next, the standard unit vectors in $\mathbb{R}^4$ are defined, $e_1 = (1,0,0,0)$, $e_2 = (0,1,0,0)$, $e_3 = (0,0,1,0)$, and $e_4 = (0,0,0,1)$. Using this 4-dimensional chaotic approach, a finite set of points X_α in $\mathbb{R}^4$ is generated, which is composed of points $b_1, \dots, b_n \in \mathbb{R}^4$ as follows:

1. $b_1 = e_{a_i}$; and
2. For each $2 \leq k \leq n$,
   $b_k = \frac{1}{2} b_{k-1} + \frac{1}{2} e_{a_k}$.

This mapping, which transforms each RNA sequence $\alpha = \beta_1 \beta_2 \dots \beta_n$ into a finite set of points $X_\alpha = \{\beta_1, \dots, \beta_n\}$ in $\mathbb{R}^4$, is referred to as the Chaos 4-dimensional Representation (C4DR).

2.3 Point Clouds and Simplices

A point cloud is a set of data points in space. In this study, each RNA sequence is represented as a point cloud in $\mathbb{R}^4$, using the C4DR, which transforms the sequence into points that capture the sequence's geometric structure.

Once a point cloud is generated, the next step is to build a simplicial complex. A simplicial complex is a mathematical structure comprising vertices, edges, triangles, and higher-dimensional analogs called simplices. Each simplex is a set of points that form a convex hull (Virk, 2022). For instance: a 0-simplex is a point; a 1-simplex is an edge connecting two points; a 2-simplex is a triangle formed by three points; a 3-simplex is a tetrahedron, and so on.

Figure 1 illustrates examples of simplices. The simplicial complex allows us to capture individual points, relationships, and higher-dimensional structures systematically.

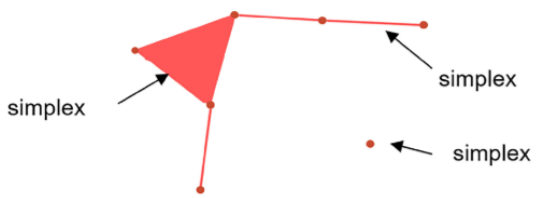


Figure 1. Examples of simplices (Riñon & Sambayan, 2024).

2.4 Filtration and Rips Complexes

Filtration is a process of gradually building a simplicial complex by varying a parameter (denoted ϵ) that controls how we connect the points in the point cloud. As ϵ increases, more points connect, and higher-dimensional simplices are formed (Munkres, 1975). This gradual construction of the simplicial complex reveals important topological features of the data.

In the method of this study, the Vietoris-Rips complexes (also called Rips complexes) are used, where simplices are formed whenever points are close enough to each other (i.e. when their pairwise distance is less than or equal to ϵ). For each value of ϵ, a new simplicial complex is created, and this process is repeated over a range of ϵ values (Zomorodian, 2010). Figure 2 shows an example of a filtration.

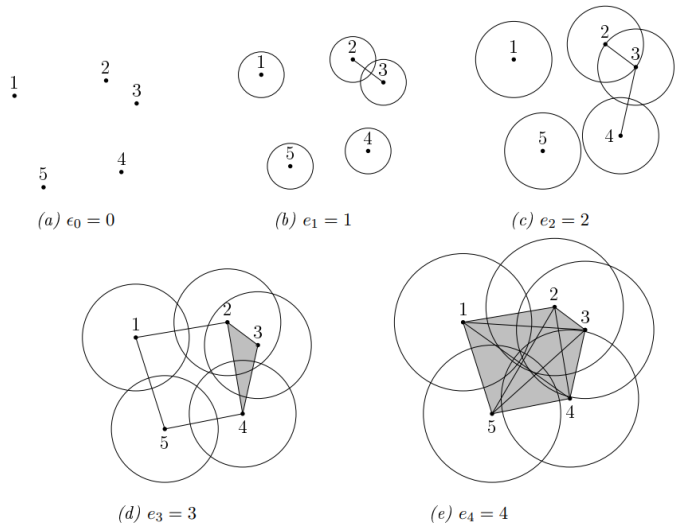


Figure 2. Forming simplices using filtration.

2.5 Persistent Homology

Persistent homology quantifies the persistence of topological features (such as connected components, loops, and voids) as the filtration progresses. It tracks explicitly when topological features appear (birth) and disappear (death) as ϵ increases.

To compute the persistent homology of a dataset, we examine how features evolve through filtration (Virk, 2022). For instance: 0-dimensional homology (H_0) measures connected components; 1-dimensional homology (H_1) captures loops or holes; and 2-dimensional homology (H_2) detects enclosed voids or cavities.

Each topological feature can be represented by its birth and death times in the filtration, and these are summarized in a persistence diagram.

2.6 Persistence Diagram

A persistence diagram is a plot that records the birth and death of each topological feature in a filtration. Each point on the diagram corresponds to a feature, with the x-coordinate representing the birth time and the y-coordinate representing the death time. The significance of a topological feature is indicated by its distance from the diagonal line in the persistence diagram – the farther a point is from the diagonal, the more critical the feature. Features that last longer are deemed significant, whereas those that appear and vanish quickly are typically considered noise (Edelsbrunner & Harer, 2022). Figure 3 presents an illustration of a persistence diagram.

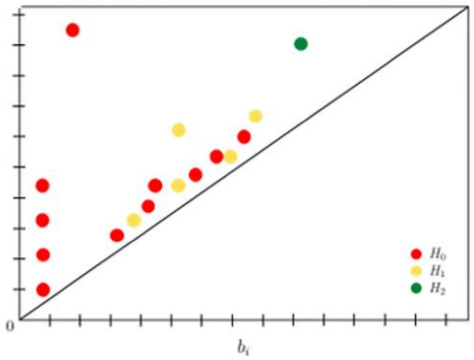


Figure 3. Persistence diagram example.

2.7 Wasserstein Distance

The Wasserstein distance compares the topological features of two-point clouds (representing two RNA sequences). This metric measures the dissimilarity between their persistence diagrams by calculating the minimum cost required to match the points in one diagram with those in the other. This study uses the 1st Wasserstein distance (W_1) for the persistence diagrams of the 1-dimensional homology (H_1), which captures the loops and cycles in the data.

The Wasserstein distance between two persistence diagrams $PD_1(X)$ and $PD_2(Y)$ is given by:

$$W_1(PD_1(X), PD_2(Y)) = \inf_{\delta: PD_1(X) \to PD_2(Y)} \sum_{(b,d) \in PD_1(X)} \|(b,d) - \delta(b,d)\|_\infty$$

where $\|\cdot\|_\infty$ denotes the L_∞-distance between two points in $\mathbb{R}^2$ (Nguyen et al., 2022).

In other words, the Wasserstein distance is a way to compare two persistence diagrams by figuring out how much "effort" or "cost" it would take to match the features from one diagram to the other. The smaller the Wasserstein distance, the more similar the two RNA sequences are in their topological features.

2.8 Proposed Method

This research adapted the steps for reconstructing a phylogenetic tree of RNA sequences from the study by Nguyen et al. (2022). The method is divided into three phases:

PHASE I - Data Collection and Preparation

Input. Begin with n RNA sequences labeled $\alpha_1, \alpha_2, \ldots, \alpha_n$. The RNA sequences represent the Variants of Concern (VOCs) retrieved from the NCBI GenBank database.

PHASE II - Topological Representation and Feature Extraction

Chaos 4-Dimensional Representation (C4DR). For each RNA sequence α_i, apply the C4DR method to map it into a point cloud X_{α_i} in $\mathbb{R}^4$. This converts each RNA sequence into a geometric structure that captures its features in a 4-dimensional space.

Persistent Homology Computation. Compute the 1st persistence diagram $PD_1(X_{\alpha_i})$ for each sequence using the TDA package in R. This step captures the topological features of the sequence, such as loops and cycles, in four-dimensional space.

PHASE III – Phylogenetic Tree Construction and Comparison.

Distance Matrix Construction. Next, calculate the Wasserstein distance W_1 between the persistence diagrams of each pair of sequences, creating a distance matrix of size $n \times n$, where each entry reflects the topological similarity between the two sequences.

Phylogenetic Tree Construction: Use the UPGMA algorithm in R to build a phylogenetic tree from the distance matrix, clustering RNA sequences based on their topological features.

Comparison. Finally, the results will be compared by analyzing both the persistent homology-based phylogenetic tree and the Clustal Omega-based phylogenetic tree, allowing verification of the findings.

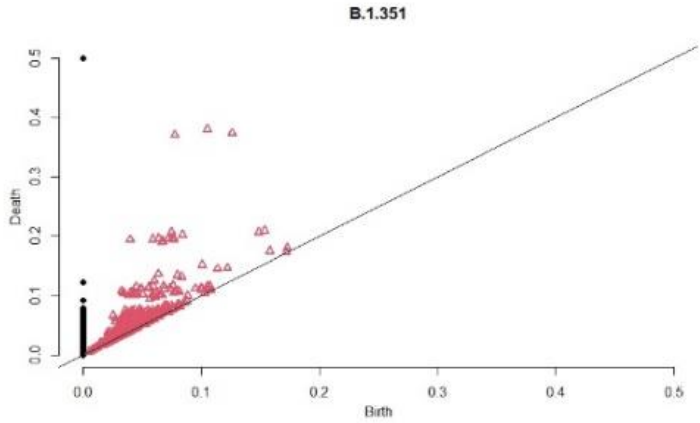
2.9 Computational Environment and Libraries

The analysis was conducted in the Posit Cloud environment using R version 4.4.1. The computational resources allocated for this study included 32 GB of RAM and an 8 GB CPU. The following R libraries were used in the analysis: *TDA* for topological data analysis and computation of persistent homology; *phangorn* for phylogenetic analysis and tree construction; *stats* for standard statistical analysis; *ape* for handling phylogenetic trees; and *msa* for multiple sequence alignment.

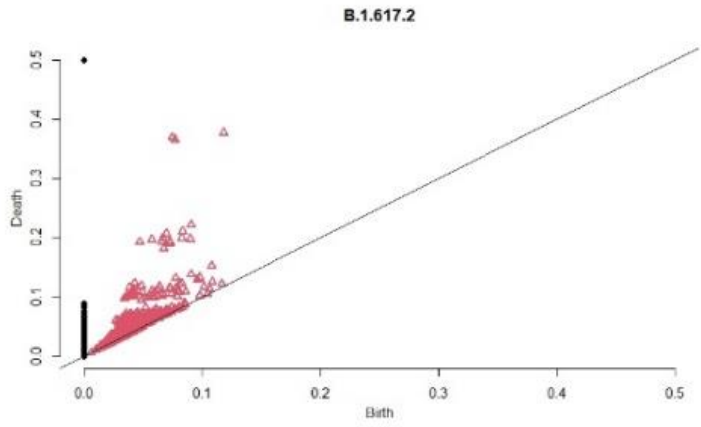
3. RESULTS AND DISCUSSION

Persistent homology was used to compute the persistence diagrams for each SARS-CoV-2 variant, which allowed us to capture their topological features. Figure 4 shows the persistence diagrams for B.1.351 (Beta) and B.1.617.2 (Delta). The diagrams reveal similar distributions of topological features, particularly in the 1-dimensional homology (H1), which reflects the loops and cycles in the data. The Wasserstein distance between these two variants was small, indicating a strong topological similarity between B.1.351 and B.1.617.2.

The topological similarities observed between the Beta (B.1.351) and Delta (B.1.617.2) variants align with shared mutations known to enhance immune escape and transmission (Zhou et al., 2021; Liu et al., 2021). Both variants exhibit mutations in the spike protein that alter its structure, reducing antibody binding and allowing for immune evasion (Garcia-Beltran et al., 2021). For instance, the E484K mutation in Beta and the L452R mutation in Delta both result in structural adaptations that confer immune escape advantages (Wang et al., 2021; Motozono et al., 2021). In persistent homology, these structural similarities are reflected by prominent features in the 1-dimensional homology, suggesting loops that correspond to these functionally significant genome regions. Thus, topological analysis captures the structural impacts of these mutations, revealing functional similarities even when precise genetic sequences differ.



(a)



(b)

Figure 4. Persistence diagrams of SARS-CoV-2 genomes with the least Wasserstein distance with (a) B.1.351 and (b) B.1.617.2

Then, the pairwise Wasserstein distances between the persistence diagrams were used to construct a phylogenetic tree (Figure 5). The PH-based tree shows that B.1.1.7 (Alpha) was the first to diverge, followed by B.1.1.529 (Omicron). P.1 (Gamma), B.1.351 (Beta), and B.1.617.2 (Delta) clustered closely, with B.1.351 and B.1.617.2 sharing the same node, reflecting their topological similarity.

This placement of B.1.1.7 (Alpha) as the first to diverge aligns with epidemiological and genomic data, indicating that Alpha was among the earliest Variants of Concern to emerge during the pandemic. Studies have shown that the Alpha variant, first identified in the United Kingdom in September 2020, accumulated many mutations distinguishing it from the ancestral strain, suggesting an early divergence (Andrew, 2020; Volz et al., 2021). The early emergence and rapid spread of Alpha support our topological findings, confirming that Alpha was indeed the first among the studied VOCs to diverge.

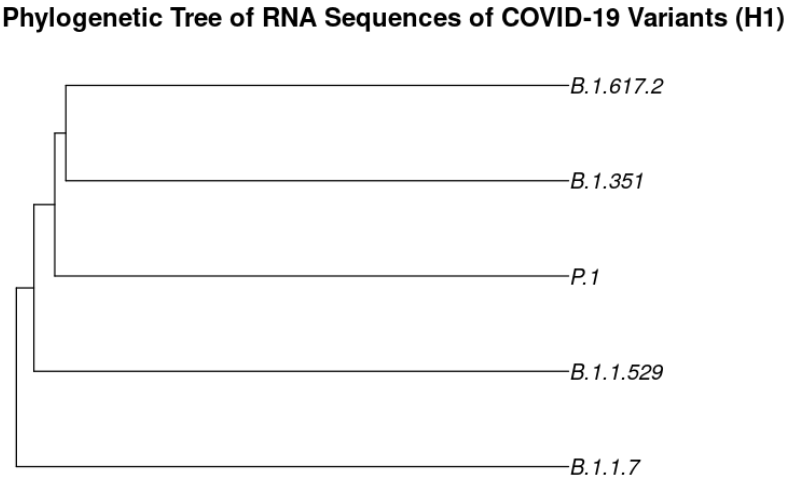


Figure 5. Phylogenetic tree of SARS-CoV-2 based on persistent homology.

To validate the topological approach, the PH-based phylogenetic tree was compared to a phylogenetic tree generated using Clustal Omega for multiple sequence alignment (Figure 6). Clustal Omega is widely regarded as a standard tool for phylogenetic analysis because it efficiently generates accurate multiple sequence alignments, mainly when working with large datasets (Sievers & Higgins, 2018). The Clustal Omega tree was

largely consistent with the results, particularly in clustering P.1 (Gamma), B.1.351 (Beta), and B.1.617.2 (Delta). However, a key difference was that Clustal Omega placed B.1.351 as the first to diverge, with B.1.1.7 (Alpha) and B.1.617.2 (Delta) clustering under the same node. In contrast, the PH-based phylogenetic tree had B.1.1.7 diverging first, while B.1.351 and B.1.617.2 were grouped.

Our findings suggest that the topological method may capture structural relationships that differ from those inferred by sequence alignment methods. The placement of Alpha as the first to diverge in our PH-based tree is supported by its early detection and spread, reinforcing the validity of our approach (Davies et al., 2021).

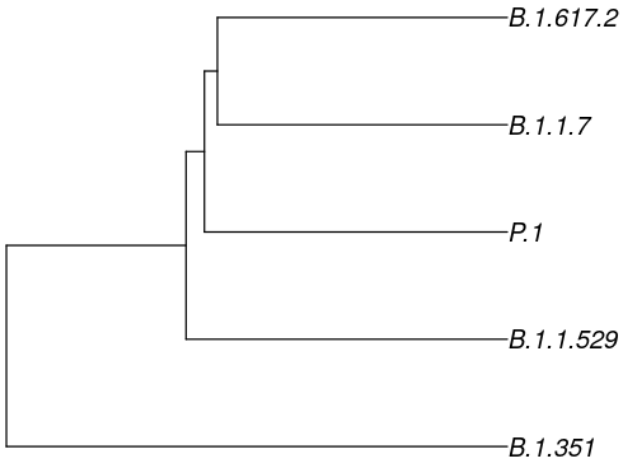


Figure 6. Phylogenetic tree of SARS-CoV-2 based on Clustal omega.

Utilizing chaos geometry and persistent homology, the topological method effectively captures structural relationships between SARS-CoV-2 variants. Nguyen et al. (2022) demonstrated that topological methods, such as persistent homology and Wasserstein distances, can reveal structural similarities in genomic sequences. In this study, the close clustering of B.1.351 (Beta) and B.1.617.2 (Delta) in the phylogenetic tree, supported by similar persistence diagrams and small Wasserstein distances, reflects the significant topological similarities between these variants. This is consistent with findings from studies highlighting similarities in their spike protein mutations, which contribute to increased transmissibility and immune evasion (Tegally et al., 2021; Harvey et al., 2021; Chaudhuri et al., 2024). As Zomorodian (2010) and Lum et al. (2013) explain, topological analysis primarily focuses on the geometric properties of data rather than tracking mutation history. In this context, the structural similarities identified in this study suggest shared functional characteristics between the variants rather than indicating a direct evolutionary lineage.

The advantage of the topological approach lies in its ability to capture these broader structural patterns that traditional sequence alignment methods might miss. For example, while Clustal Omega placed B.1.351 as the first to diverge, clustering B.1.1.7 (Alpha) with B.1.617.2 (Delta), the PH-based method places B.1.1.7 as the first to diverge and highlights a close relationship between Beta and Delta. Epidemiological findings further support this difference. A model-inference system analyzing pandemic dynamics in South Africa found that both Beta and Delta variants shared functional similarities, particularly regarding immune erosion and transmissibility. Beta eroded immunity among 63.4% of previously infected individuals, while Delta eroded 24.5%, and Delta was 47.5% more transmissible than the ancestral strain (Yang et al., 2022). These findings align with the structural similarities identified in the topological analysis, suggesting that both approaches capture critical functional aspects of the virus that are not easily detected through traditional sequence-based methods. This difference suggests that the method is sensitive to structural features that may not be apparent from sequence alignment alone, offering a complementary perspective on how different variants may function similarly. Moreover, the confirmation that Alpha was the first to diverge highlights the potential of our topological method to reflect the actual evolutionary pathways of SARS-CoV-2 variants, complementing traditional phylogenetic analyses.

Persistent homology provides an intuitive and quantitative way to analyze the structural complexity of genomic sequences. By tracking topological features like loops and voids, we better understand how genomic structures compare across variants. In the analysis, the structural similarities between Beta and Delta variants are supported by their shared topological features, which align with studies reporting similar functional behaviors, such as immune escape (Planas et al., 2021; Afroze et al., 2024). These similarities could reflect critical functional aspects of the virus that impact its behavior and spread, though they do not directly indicate evolutionary relationships.

While the method successfully captures these structural relationships, some discrepancies between topological and sequence-based approaches indicate that further refinement may be necessary. The difference in branching orders between the PH-based tree and Clustal Omega highlights the distinct types of relationships each approach is designed to capture. Combining topological and sequence-based approaches could provide a more holistic view of variant relationships, offering insights into structural similarities and genetic mutations.

4. CONCLUSION

This study demonstrates the effectiveness of topological methods, such as chaos geometry and persistent homology, in uncovering structural relationships between SARS-CoV-2 variants. Unlike traditional sequence-based approaches, which focus solely on genetic mutations, the method offers deeper insights by capturing and quantifying topological features. By focusing on structural similarities, this novel approach provides a new way to classify viral variants, with significant implications for public health and scientific research. One key finding from this study is the close topological relationship between B.1.351 (Beta) and B.1.617.2 (Delta), supported by their shared spike protein mutations, which are known to influence transmissibility and immune escape. These similarities demonstrate the value of topological analysis in identifying structural patterns that may not be evident through traditional sequence alignment methods.

Data availability statement

The dataset consists of complete genomic sequences of these variants retrieved from the NCBI GenBank database. The specific accession numbers for these sequences, obtained from the NCBI Genbank, are as follows: Alpha (B.1.1.7) [OV054768], Beta (B.1.351) [OR278055], Gamma (P.1) [OR578388], Delta (B.1.617.2) [OK091006], and Omicron (B.1.1.529) [OQ905474].

Ethics statement

This research only involved the computational analysis of genomic data from public databases, there are no personal or sensitive data handled.

Funding

The study received no funding.

CRediT authorship contribution statement

J.R. led Conceptualization and Project administration; co-designed the Methodology; performed the Formal analysis; produced the Visualization; and wrote the Writing – original draft. P.M. co-designed the Methodology; implemented the Software and ran experiments; performed Validation; contributed to Formal analysis; and carried out Writing – review & editing. N.N. led Data curation; performed Validation; assisted with Investigation and Visualization; and contributed to Writing – review & editing. All authors contributed to Investigation, secured Resources (GenBank retrieval and computing environment), jointly interpreted results, and approved the final manuscript.

Declaration of generative ai and ai-assisted technologies in the writing process

During manuscript preparation, the authors used ChatGPT-4o for language editing. The authors reviewed and revised the text and accept full responsibility for the content. No generative AI system was credited as an author, and no confidential or personally identifying data were entered into AI tools.

Declaration of competing interests

The authors declare no competing financial or personal interests relevant to the work reported in this article.

Acknowledgments

The authors thank Dr. Mark Lexter D. De Lara for the technical assistance.

REFERENCES

Afroze, F., Begum, M. N., Ahmed, T., El Arifeen, S., Rahman, M. Z., Rahman, A. E., ... & Chisti, M. J. (2024). Clinical characterisation, treatment outcomes, and case fatality risk of patients with different SARS-CoV-2 variants in Bangladesh. *Journal of Global Health*, *14*. https://doi.org/10.7189/jogh.14.05009

Aggarwal, M., & Periwal, V. (2023). Tight basis cycle representatives for persistent homology of large biological data sets. *PLOS Computational Biology*, *19*(5), e1010341. https://doi.org/10.1371/journal.pcbi.1010341

Akhtar, M., Epps, J., & Ambikairajah, E. (2008). Signal processing in sequence analysis: advances in eukaryotic gene prediction. *IEEE Journal of Selected Topics in Signal Processing*, *2*(3), 310-321. https://doi.org/10.1109/JSTSP.2 008.923854

Andrew, R. (2020). Preliminary genomic characterisation of an emergent SARS-CoV-2 lineage in the UK defined by a novel set of spike mutations. *Virological*.

Brzinsky-Fay, C. (2014). Graphical representation of transitions and sequences. In *Advances in sequence analysis: Theory, method, applications* (pp. 265-284). Cham: Springer International Publishing. https://doi.org/10.1007/978-3- 319-04969-4_14

Chaudhuri, A., Das, S., & Chakrabarti, S. (2024). Mutational and evolutionary dynamics of non- structural and spike proteins from variants of concern (VOC) of SARS-CoV-2 in India. *International Journal of Biological Macromolecules*, 137154. https://doi.org/10.1016/j.ijbiomac.2024.137154

Davies, N. G., Abbott, S., Barnard, R. C., Jarvis, C. I., Kucharski, A. J., Munday, J. D., ... & Edmunds, W. J. (2021). Estimated transmissibility and impact of SARS-CoV-2 lineage B.1.1.7 in England. *Science*, *372*(6538), eabg3055.

De Lara, M. L. D. (2023). Persistent homology classification algorithm. *PeerJ Computer Science*, *9*, e1195. https://doi.org/10.7717/peerj-cs.1195

Edelsbrunner, H., & Harer, J. L. (2022). *Computational topology: an introduction*. American Mathematical Society.

Feldman, D. P. (2012). *Chaos and fractals: an elementary introduction*. Oxford University Press.

Garcia-Beltran, W. F., Lam, E. C., Denis, K. S., Nitido, A. D., Garcia, Z. H., Hauser, B. M., ... & Balazs, A. B. (2021). Multiple SARS-CoV-2 variants escape neutralization by vaccine-induced humoral immunity. Cell, 184(9), 2372-2383. https://doi.org/10.1016/j.cell.2021.03.013

Gowrisankar, A., Priyanka, T. M. C., & Banerjee, S. (2022). Omicron: a mysterious variant of concern. *The European Physical Journal Plus*, *137*(1), 1-8. https://doi.org/10.1140/epjp/s13360-021-02321-y

Harvey, W. T., Carabelli, A. M., Jackson, B., Gupta, R. K., Thomson, E. C., Harrison, E. M., ... & Robertson, D. L. (2021). SARS-CoV-2 variants, spike mutations and immune escape. *Nature Reviews Microbiology*, *19*(7), 409-424. https://doi.org/10.1038/s41579-021-00573-0

Hoang, T., Yin, C., & Yau, S. S. T. (2016). Numerical encoding of DNA sequences by chaos game representation with application in similarity comparison. *Genomics*, *108*(3-4), 134-142. https://doi.org/10.1016/j.ygeno.2016.08.002

Liu, C., Ginn, H. M., Dejnirattisai, W., Supasa, P., Wang, B., Tuekprakhon, A., ... & Screaton, G. R. (2021). Reduced neutralization of SARS-CoV-2 B. 1.617 by vaccine and convalescent serum. *Cell*, *184*(16), 4220-4236. https://doi.org/10.1016/j.cell.2021.06.020

Lum, P. Y., Singh, G., Lehman, A., Ishkanov, T., Vejdemo-Johansson, M., Alagappan, M., ... & Carlsson, G. (2013). Extracting insights from the shape of complex data using topology. *Scientific reports*, *3*(1), 1236. https://doi.org/10.1038/srep01236

Meng, Z., Anand, D. V., Lu, Y., Wu, J., & Xia, K. (2020). Weighted persistent homology for biomolecular data analysis. *Scientific reports*, *10*(1), 2079. https://doi.org/10.1038/s41598-019-55660-3

Motozono, C., Toyoda, M., Zahradnik, J., Saito, A., Nasser, H., Tan, T. S., ... & Sato, K. (2021). SARS-CoV-2 spike L452R variant evades cellular immunity and increases infectivity. *Cell host & microbe*, *29*(7), 1124-1136. https://doi.org/10.1016/j.chom.2021.06.006

Munkres, J. R. (1975). *Topology: A First Course.* Prentice Hall. Eaglewood Cliffs, New Jersey.

Nguyen, D. Q. N., Le, P. D. T., Xing, L., & Lin, L. (2022, December). A topological characterization of dna sequences based on chaos geometry and persistent homology. In *2022 International Conference on Computational Science and Computational Intelligence (CSCI)* (pp.1591-1597).IEEE. https://doi.org/10.1109/CSCI58124.2022.00285

Parreño, S. J., & Anter, M. C. J. (2024). A novel approach to analyzing Philippine inflation dynamics through persistent homology. *Multidisciplinary Science Journal*, *6*(12), 2024270-2024270.

Parums, D. V. (2021). Revised World Health Organization (WHO) terminology for variants of concern and variants of interest of SARS-CoV-2. *Medical science monitor: international medical journal of experimental and clinical research*, *27*, e933622-1. https://doi.org/10.12659/MSM.933622

Pereira, R., Oliveira, J., & Sousa, M. (2020). Bioinformatics and computational tools for next-generation sequencing analysis in clinical genetics. *Journal of clinical medicine*, *9*(1), 132. https://doi.org/10.3390/jcm9010 132

Planas, D., Veyer, D., Baidaliuk, A., Staropoli, I., Guivel-Benhassine, F., Rajah, M. M., ... & Schwartz, O. (2021). Reduced sensitivity of SARS-CoV-2 variant Delta to antibody neutralization. *Nature*, *596*(7871), 276-280. https://doi.org/10.1038/s41586-021-03777-9

Poltronieri, P., Sun, B., and Mallardo, M. (2015). RNA viruses: RNA roles in pathogenesis, coreplication and viral load. *Current genomics, 16*(5), 327-335.

Riñon, E. M. A., & Sambayan, R. R. (2024). Topological Approach for Detection of Structural Breakpoints in Philippine Stock Price Data Surrounding the COVID-19 Pandemic. *Philippine Journal of Science*, *153*(3).

Sievers, F., & Higgins, D. G. (2018). Clustal Omega for making accurate alignments of many protein sequences. *Protein Science, 27*(1), 135-145. https://doi.org/10.1002/pro.3290

Tegally, H., Wilkinson, E., Giovanetti, M., Iranzadeh, A., Fonseca, V., Giandhari, J., ... & de Oliveira, T. (2021). Detection of a SARS-CoV-2 variant of concern in South Africa. *Nature, 592*(7854), 438-443. https://doi.org/10.1038/s41586- 021-03402-9

Virk, Ž. (2022). *Introduction to Persistent Homology*. Založba UL FRI.

Wang, P., Nair, M. S., Liu, L., Iketani, S., Luo, Y., Guo, Y., ... & Ho, D. D. (2021). Antibody resistance of SARS-CoV-2 variants B. 1.351 and B. 1.1. 7. *Nature*, *593*(7857), 130-135. https://doi.org/10.1038/s41586- 021-03398-2

Wasserman, L. (2018). Topological data analysis. *Annual Review of Statistics and Its Application, 5*(1), 501-532. https://doi.org/10.1146/annurev-statistics-031017-100045

Wei, G. W. (2017). Persistent homology analysis of biomolecular data. *Society for Industrial and Applied Mathematics.*

Wu, T., Ge, Y., Zhao, K., Zhu, X., Chen, Y., Wu, B., ... & Cui, L. (2020). A reverse-transcription recombinase-aided amplification assay for the rapid detection of N gene of severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2). *Virology*, *549*, 1-4. https://doi.org/10.1016/j.virol.2020.07.006

Yang, W., & Shaman, J. L. (2022). COVID-19 pandemic dynamics in South Africa and epidemiological characteristics of three variants of concern (Beta, Delta, and Omicron). *Elife, 11*. https://doi.org/10.7554/eLife.78933

Zhou, D., Dejnirattisai, W., Supasa, P., Liu, C., Mentzer, A. J., Ginn, H. M., ... & Screaton, G. R. (2021). Evidence of escape of SARS-CoV-2 variant B.1.351 from natural and vaccine-induced sera. *Cell*, *184*(9), 2348-2361. https://doi.org/10.1016/j.cell.2021.02.037

Zomorodian, A. (2010). Fast construction of the Vietoris-Rips complex. *Computers and Graphics, 34*(3), 263-271. https://doi.org/10.1016/j.cag.2010.03.007

Volz, E., Mishra, S., Chand, M., Barrett, J. C., Johnson, R., Geidelberg, L., ... & Ferguson, N. M. (2021). Assessing transmissibility of SARS-CoV-2 lineage B. 1.1. 7 in England. *Nature*, *593*(7858), 266-269. https://doi.org/10.1038/s41586-021-03470-x