RESEARCH ARTICLE
doi: 10.2306/scienceasia1513-1874.2023.070

ScienceAsia 49 (2023): **726–736**

Detection and classification of mutational field between Omicron BA.5 and other SARS-CoV-2 variants of concern with support vector machine

Kabin Kanjamapornkul[a], **Thanyada Rungrotmongkol**[b,c], **Supot Hannongbua**[a,*]

[a] Center of Excellence in Computational Chemistry (CECC), Department of Chemistry, Faculty of Science, Chulalongkorn University, Bangkok 10330 Thailand

[b] Center of Excellence in Biocatalyst and Sustainable Biotechnology, Department of Biochemistry, Faculty of Science, Chulalongkorn University, Bangkok 10330 Thailand

[c] Program in Bioinformatics and Computational Biology, Graduate School, Chulalongkorn University, Bangkok 10330 Thailand

[*]Corresponding author, e-mail: supot.h@chula.ac.th

Received 1 Sep 2022, Accepted 24 May 2023
Available online 12 Oct 2023

ABSTRACT: A theoretical investigation into the hidden source field of mutation affecting the curvature change in the spike (S) protein of Omicron BA.5 and related variants is reported. The curvature in the open string shape of S protein is defined using the Yang-Mills field over a new type of connection known as quantum genotype. By adding more invariant properties of curvature two-forms to the adaptive tensor fields in DNA, RNA, and protein molecules, we redefined the hidden quantum biological co-state of the mutational field between the virus and the host cell. The new algorithm was applied to classify mutations in S protein with a support vector machine. The results showed that the average performance of the prediction of unknown amino acids in 14 variants including Omicron BA.1–BA.5 is 97.79%. Additionally, we demonstrated a new approach for the quantitative measurement of changing curvature of S protein mutations in each amino acid. The empirical analysis of the probability distribution of time series data showed the evolution of the quantum genotype over time, revealing a new direction of evolution in SARS-CoV-2's quantum genotype opposite to the period between 2020 and 2021. This work can be applied to detect new incoming novel variants of SARS-CoV-2 in the future and provide insight into the coupling between the passive and active sides of communication in the biomolecular layer.

KEYWORDS: Omicron BA.5, Chern-Simons current, support vector machine, mutational field

INTRODUCTION

SARS-CoV-2 variants Omicron BA.5 and BA.5.1–BA.5.5 [1, 2] are currently considered new variants of concern by the World Health Organization [3]. Omicron BA.5 can evade the host cell immune system and fight the vaccine [4]. Recent studies have shown metabolic characteristics of Omicron variant infection that are different from Delta variant infection [5]. All variants of SARS-CoV-2 are composed of enveloped proteins with single-strand (+) sense RNA viruses of 25–30 kb length surrounded by nucleocapsid protein (N) [6]. The virus surface is covered by the membrane (M), envelope (E), and spike (S) protein [7]. The S protein is highly variable, and its mutations are essential because it binds with receptor proteins. The RBD domain of S protein is located between about amino acid numbers 450–520 [8]. The mutations in the S protein of Omicron BA.5 and BA5.5 are unique and related to feedback evolution down to the Delta variant while infected with the same host cell [9]. These mutations significantly affect the curvature of the S protein folding in the RBD location, causing changes in the antibody binding [10] and the effectiveness of the immune system [11]. There is a considerable number of mutations in Omicron BA.5 that make it different from other variants of concern [12], and it spreads faster than Delta and other previous known variants of SARS-CoV-2 [13]. Understanding the similarity between Omicron BA.5 and Delta variant is still an open question. The aim of this study is to investigate the hidden source field of mutation that affects the curvature change in the spike (S) protein of Omicron BA.5 and related variants and compare them with Delta variant.

The typical way to detect the mutation in the S protein is through an alignment method over the alphabet code of the genome [14]. There are many problems with this method such as the time computation being too slow. Moreover, the sorting algorithm over the string alphabet gives all statistical parameters that are limited for quantitative measurement. However, in this work, we used a new approach by applying the approach of fiber bundle and Chern-Simons current to visualize the genetic code, which is based on the cohomology theory for quantum biology [15] that can be used for measuring the invariant properties of quantum genotype in the curvature of S protein mutations. All statistical parameters can be computed into a histogram to detect histogram change in near future mutations using this approach with similar results to the previous method.

Several studies have utilized machine learning techniques to predict mutations in the S protein of SARS-CoV-2 variants of concern [16, 17]. For instance, Ali et al [18] predicted mutations in 5 variants of concern without Omicron using a support vector machine approach [19]. Most of these studies such as that by Sathipati et al [20] have relied on alignment-based methods for machine learning. Others, like Qiang et al [21], have utilized machine learning based on molecular dynamics [22] to detect genetic variation in the Omicron variant and compare their results with experimental data. Kumar et al [23] also studied the effect of mutations between Omicron BA.1 and Delta using an alignment-based approach but did not include BA.2–BA.5 variants in their analysis.

MATERIALS AND METHODS

What is a mutational field in the genetic code of SARS-CoV-2

The mutational field in the genetic code of SARS-CoV-2 is a hidden behavior field stored in a quantum genotype as a source that distinguishes living objects from non-living ones. The mathematical structure of this field is based on the properties of photon and instanton as a source of massless spin integer electromagnetic wave as well as the conscious hidden state in the behavior field of social interaction between biomolecules. In this study, we used a new approach that applies transactional analysis theory [24] and Chern-Simons theory to measure the invariant property of 2 fiber bundles between mutated and non-mutated amino acids as a curvature of protein folding in a simple biological model of a one-dimensional curve shape of an open string of protein. We also investigated the hidden relationship between the change of curvature in protein folding and the mutational field in the underlying genetic code of the S protein of the Omicron variant of SARS-CoV-2. Changes in the invariant properties of adaptive fields in genetic code are induced by a quantum transition in the (d, r, p)-layer coordinate system [25], which is a source of Chern-Simons current in differential triple mutations in living organisms, including viruses [26]. We used Yang-Mills theory to define and measure the curvature in the form of mutation in the genetic code of amino acids with an invariant polynomial [27]. This new approach allows us to study the curvature change in S protein folding directly from the underlying empirical genetic code and can be further analyzed with a machine learning algorithm [28] for detecting mutations in the SARS-CoV-2 variant of concern.

Transactional analysis is used to map interpersonal relationships between 3 ego-states of the human psychological state, so-called the parent (P), adult (A), and child (C) states [29]. These 3 states are found in humans, the highest evolutionally living organisms in the biosphere. If we trace back to the source and origin of these 3 states in the smallest undivided subunit of a living cell, these 3 stages are analogous to three biomolecule DNA, RNA, and protein types. We redefine the adaptive fields P, A, and C in the genetic code of 3 kinds of biomolecules by the primary instinct field, immaturity state, and maturing state. It is a spinor field in connection with one form $A_\mu(g)$. The definition of the evolutional path is a phase transition in the eigenvalue λ of the mutational field, which we call the geodesic path of parallel transport of quantum genotype g_{ij}, from the path of adaptive field induced by natural selection to survive and transmit the invariance properties of mutation from the past to the present in the form of a mutational field on the Caesar ring of the biological clock (Fig. 1). It is the adaptative feedback system between living organisms and the environment according to the response system of the behavior field in genetic code with constraints from the environment. From a living organism, adaptation becomes a function of one molecule on a sequence of proteins to change the curvature of the docking system into a non-equilibrium state. The change in curvature of the genetic code is the source of the mutational field $A_\mu(g)$ over the underlying genotype g. The question is, why is the rate of S protein mutation in Omicron higher than the rate of mutation in ACE2 protein and antibody in the memory B-cell of their host cell? What determines the source of the S protein mutation in SARS-CoV-2? One of the answers is the geodesic path, as the least action principle over amino acid sequence needs to be newly appropriately defined and introduced to quantum biology. In differential geometry, the source of the geodesic path is the curvature of space with the connection. The connection is a Jacobian flow and analogy with gravitation in the D-brane theory. That implies the source of protein folding is a gravitational effect in quantum genotype [30]. The change of moving frame in the (d, r, p)-layer of behavior is a parallel transport of mutational field defined by Wilson loop of connection one form on principle G-bundle of a four-dimensional manifold with curvature [31]. In general relativity, $g = g_{ij}$ a metric tensor is used to explain the gravitational field as a spacetime curvature by using the geodesic equation. It is a function of coordinates $g_{ij}(x)$ equipped with a local measure of the deformed distance $\mathrm{d}s^2 = g_{ij}\,\mathrm{d}q_i\,\mathrm{d}q_j$ between 2 separated points in the dual tangent manifold of a living organism.

Definition 1 Let g be the cocycle of quantum genotype with behavior field $A(g)$. Let g_{ij} be the cocycle of transition between genetic code in host cell (−) RNA q_j and viral (+) RNA g_j during viral replication process. Let g_{jk} be the cocycle of transition between genetic code in (+) RNA q_j and viral protein q_k. Let g_{ki} be the cocycle of transition between genetic code in viral protein q_k and host cell (−) RNA q_j.

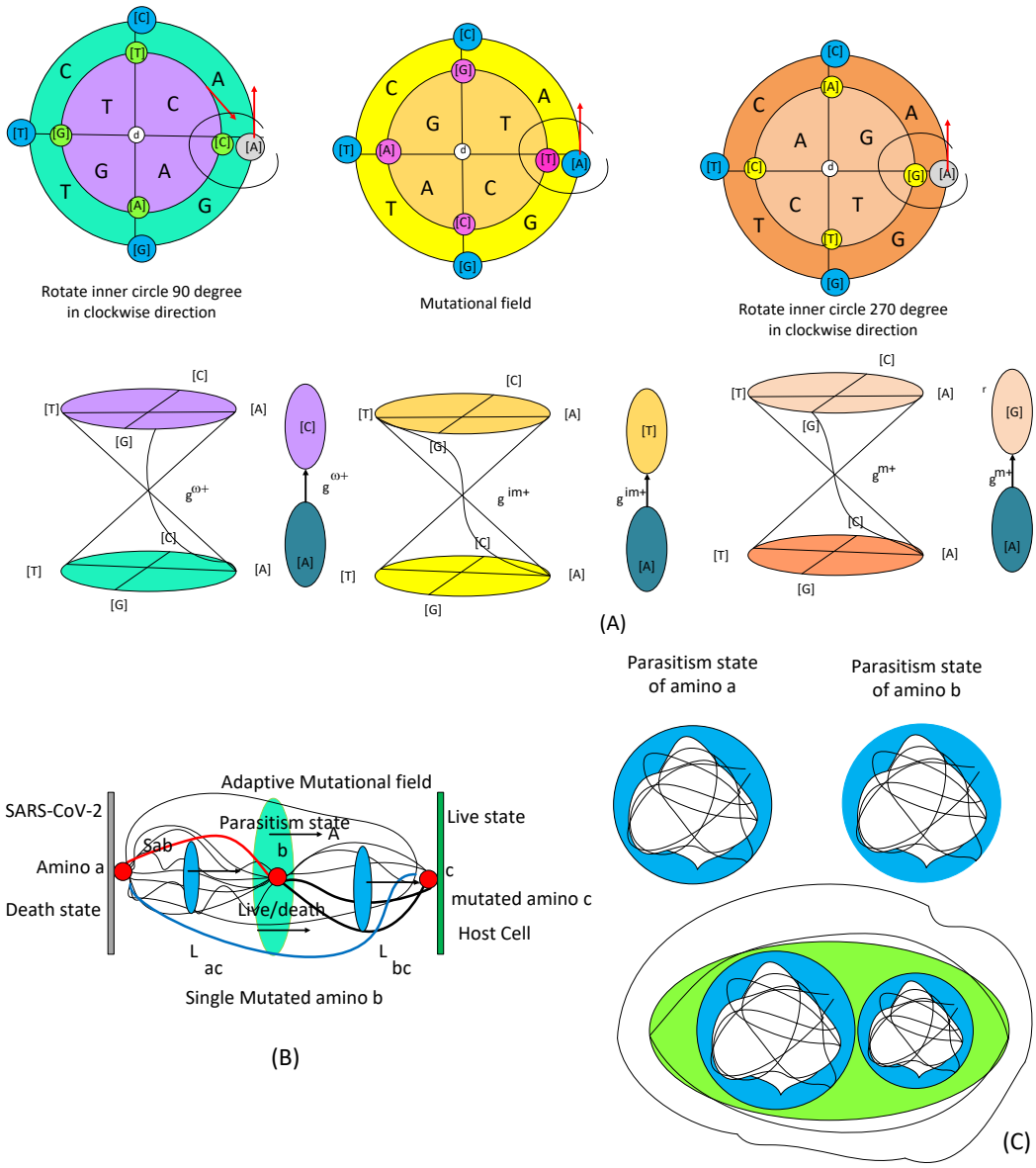

Fig. 1 (A) The mutational field represented in the behavior field of the induced-layer coordinate system of 2 differential forms. (B) Homotopic equivalent class utilized in this model to approximate the amino acid wave properties and shape in an open string shape, where a phase transition occurs between two-pointed space at the path's boundary. (C) The figure depicting the interaction between 2 wave functions of the evolutional field in amino acids.

Let a connection one form $A_{\mu=m}(g) = \frac{1}{2} g^{im}(\frac{\partial g_{ij}}{\partial q_k} + \frac{\partial g_{jk}}{\partial q_i} - \frac{\partial g_{ki}}{\partial q_j})$ over tangent of a manifold of virus. We define Hamiltonian of host cell $H^{(\text{host}-\text{cell})}(p,q)$ by dual space of Lagrangian of virus $L^{\text{virus}}(q,q')$.

In quantum biology, there are 4 genetic codes with unseparated properties of mutation between them. The distance here is not the same as the distance in general relativity. Still, it is the distance of evolution between a mutated genetic code in an excited state and the original one with an unmuted ground state. For example, if the mutation occurs in a viral RNA molecule r^*, the distance is,

$$\begin{aligned} \mathrm{d}s^2 &= A(g)\,\langle p, \mathrm{d}q\rangle = A(g^r_{ij}) p_{i,r^*}\,\mathrm{d}q_{j,r^*} \\ &= H^{\text{host}-\text{cell}}(p,q) + L^{\text{virus}}(q,q'). \end{aligned} \tag{1}$$

A connection in differential geometry $A(g)$ is a source of behavior field in genetic code if we added more properties of 3 types of fields from transactional analysis to the communication behavior of invariant gauge field in DNA, RNA, and protein molecules. Let

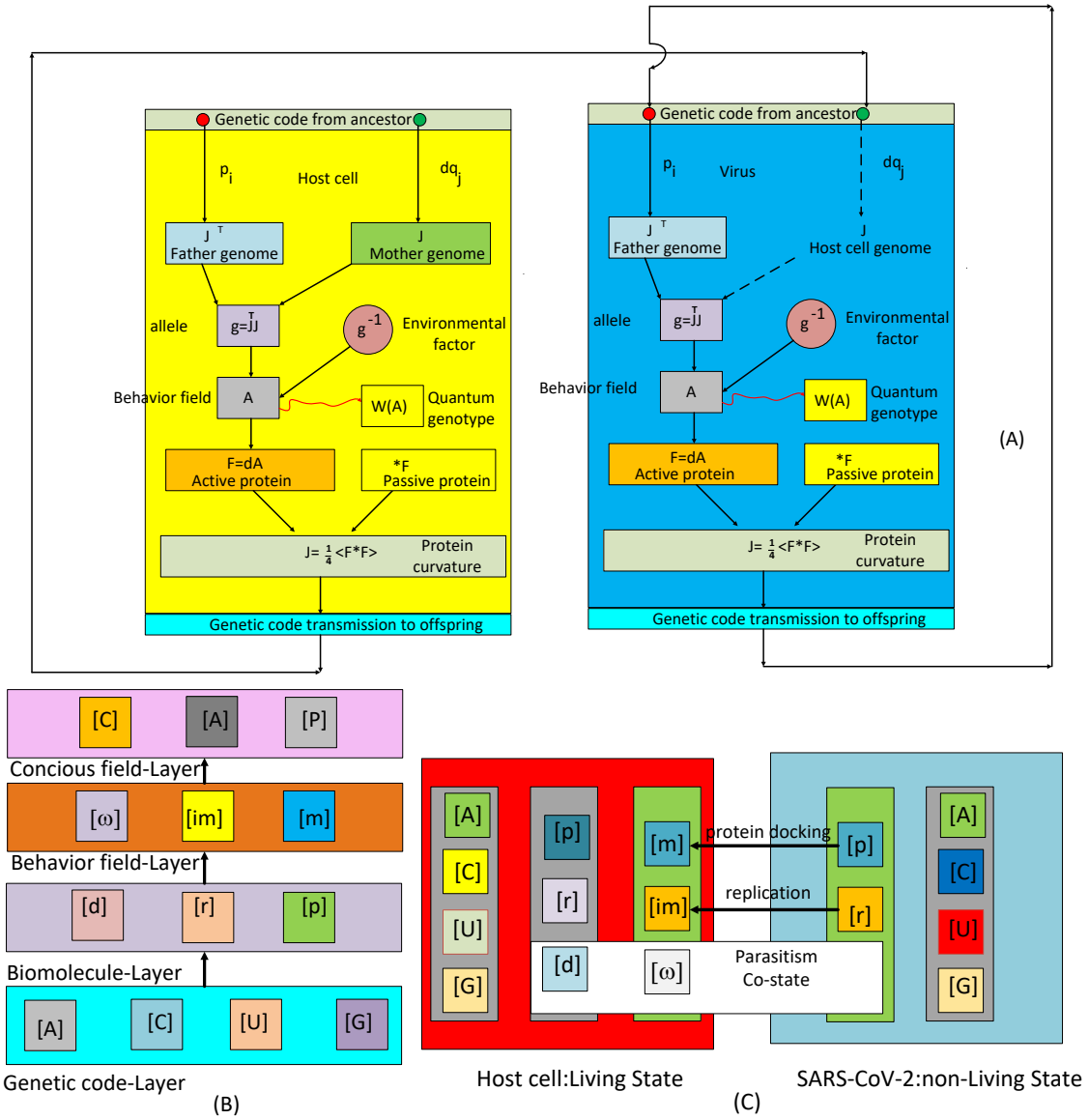

Fig. 2 (A) The diagram depicting the parasitism state between a host cell and virus with a replication loop. In this diagram, W(A) represents the Wilson loop of the behavior field, which acts as a source of quantum genotype. (B) The layer structure of a higher organism being different from that of a virus and single-cell organism as the top layer is a conscious field consisting of various states of behavior fields. The [C] state represents the child state, [A] state is the adult state, and [P] is the parent state. (C) The diagram illustrating the communication between the behavior field in the virus and host cell during parasitism. A virus consists of a genetic code layer and biomolecule layer with only 2 types from 3 types of biomolecules. In SARS-CoV-2, viruses must reproduce by communicating with the host cell through 2 levels of transactional communication. In internal reproduction, SARS-CoV-2 sends viral RNA to host cell RNA, and in the external reproduction process, the virus uses S protein to dock with the ACE2 protein in the host cell.

$A(g) = A^{\text{Behavior}}_{\text{host}}$ be a behavior field in host cell with gene expression into RNA state r and protein state p. The induced field in the host cell is $A^{\text{Behavior}}_{\text{host}} = (A^{\omega^{\pm}}_{\mu}, A^{im^{\pm}}_{\mu}, A^{m^{\pm}}_{\mu})$ On the other hand, the virus also has their own behavior field on biomolecule in which be denoted shortly by (d,r,p)-layer in $A^{(d,r,p)}_{\text{virus}}$. They communicate to each based on the fact that the genetic code will transmit from the previous generation to the next generation in (d,r,p)-layer -communication layer with no mutation. That means, if the first

generation contains start alphabet code in codon with AUG and stop alphabet code in viral RNA, denoted by $A^{(d,r,p)}_{\text{virus}}$, that location cannot be changed also in hidden coordinate of viral protein using the genetic material of host cell (—) RNA during viral replication process. If there exists a mutation in one alphabet code, then the mutational field will affect to invariant gauge field $F = \mathrm{d}A^{(d,r,p)}_{\text{virus}}$. Here $F := F^{(d,r,p)}$ is called the Yang-Mills field in genetic code and denoted with the same signature of (d,r,p)-layer for use in quantum biology with extra properties (Fig. 2). In theoretical physics, this quantity is used for representing the curvature of spacetime. But in quantum biology, we use it to represent the change of curvature of protein folding.

The source of a mutational field in transition between viral RNA and S protein

The source of the mutational field is a Yang-Mills field in genetic code. The effect of hidden Yang-Mills force defines it in biomolecule to localize the spinor field in hydrogen bonding of homochiral molecule for keeping the genetic information from the ancestor on the spinor field A_μ in the form of connection one form of behavior field in genetic code [32]. It is an effect of parallel transport of hydrogen bonding through the geodesic evolutional path of the DNA, RNA, and protein during the viral replication process within variant properties. The differential two forms of connection one form in parallel transport from DNA, RNA, and to protein, $F^{(d,r,p)} = \mathrm{d}A$. We write the differential two forms $F^{(d,r,p)} = \sum F_{\mu\nu}\,\mathrm{d}p_\mu \wedge \mathrm{d}q_\nu$. The induced relation of interaction of behavior field between virus and host cell is

$$\nabla \times A^{(d,r,p)}_{\text{virus}} = \frac{\partial A^{\text{Behavior}}_{\text{host}}}{\partial t} + J^{\mu=k}. \tag{2}$$

The quantity $J^{\mu=k}$ is the Chern-Simons current in the rate of change of adaptive behavior field of interaction between two-part, $A^{(d,r,p)}_{\text{virus}}$. In quantum biology, there exists a new definition of the Chern-Simons current in genetic code by using a differential three form over 3 alphabet codes in a codon that one can use for visualizing and quantitative analysis of the curvature of protein directly from the genetic code in biological time series [33] by using the formula that can plot the curvature of all proteins in living organism now from underlying genetic code,

$$J^{\mu=k} = \frac{1}{4}\left\langle F^{(d,r,p)} * F^{(d,r,p)} \right\rangle, \tag{3}$$

$$J^{\mu=k} = \sqrt{\frac{2}{k+2}}\sin\frac{\pi}{k+2}, \qquad k = 1,2,\ldots,64. \tag{4}$$

where k is a coupling constant representing the transition state in the gauge group of quantum genotype. The flowchart of the algorithm of calculation of the Chern-Simons current in genetic code is shown in Fig. 3.

Table 1 The average values of genetic code in S protein from daily samples collected in each month over the span of one and a half years. The data was obtained from the GISAID database, starting from 15 January 2020 to 30 June 2021 with $y(x) = \beta_1 x^2 + \beta_2 x + \beta_3$.

No.	Month/Year	$\beta_1 \times 10^{-8}$	$\beta_2 \times 10^{-5}$	β_3
1	JAN 2020	−1.1933	1.4553	−0.0091
2	FEB 2020	−1.2125	1.4783	−0.0091
3	MAR 2020	−1.2466	1.5115	−0.0092
4	APR 2020	−1.2346	1.5121	−0.0093
5	MAY 2020	−1.1982	1.4693	−0.0092
6	JUN 2020	−1.2135	1.4889	−0.0092
7	JUL 2020	−1.2077	1.4535	−0.0090
8	AUG 2020	−1.1472	1.3797	−0.0088
9	SEP 2020	−1.0107	1.1544	−0.0081
10	OCT 2020	−0.9441	1.0112	−0.0074
11	NOV 2020	−0.9490	1.0404	−0.0076
12	DEC 2020	−1.0301	1.1336	−0.0078
13	JAN 2021	−1.2159	1.4442	−0.0091
14	FEB 2021	−1.3298	1.6259	−0.0097
15	MAR 2021	−1.3839	1.7004	−0.0098
16	APR 2021	−1.3285	1.6005	−0.0095
17	MAY 2021	−1.3982	1.7362	−0.0101
18	JUN 2021	−1.3383	1.6691	−0.0100

Dataset

Empirical analysis was performed on 5 groups of data consisting of a total of 1740 selected samples. The first group included 504 randomly selected samples of time series of genetic code, covering the period from 17 January 2020 to 30 June 2021. The second group consisted of 326 randomly selected samples of S protein, covering the period from 1 November 2021 to 31 December 2021. The third group included 490 randomly selected samples of S protein, covering the period from 1 December 2022 to 31 January 2023. The fourth group comprised 270 samples of 9 variants of concern with each variant containing 30 randomly selected samples obtained from GISIAD. The 9 variants were Delta, Gamma, Beta, Alpha, Mu, Iota, Lambda, Kappa, and Eta, and the samples were randomly selected from the S protein sequence. The fifth group consisted of 150 samples of Omicron BA.1–BA.5 with each variant containing 30vrandomly selected samples.

RESULTS

The computational results of the Chern-Simons current in the genetic code of the whole genome of SARS-CoV-2 compared with the HIV genome are shown in Fig. 4. The plot of the trend of the Chern-Simons current was fitted with quadratic polynomials, and the difference between the regression coefficients was measured to allow a quantitative measurement of the transition state from the Omicron to SARS-CoV-2 as a complementary approach to the qualitative sequence alignments. The result of the fitting parameters is summarized in Table 1. The average values of β_1 and β_2 of

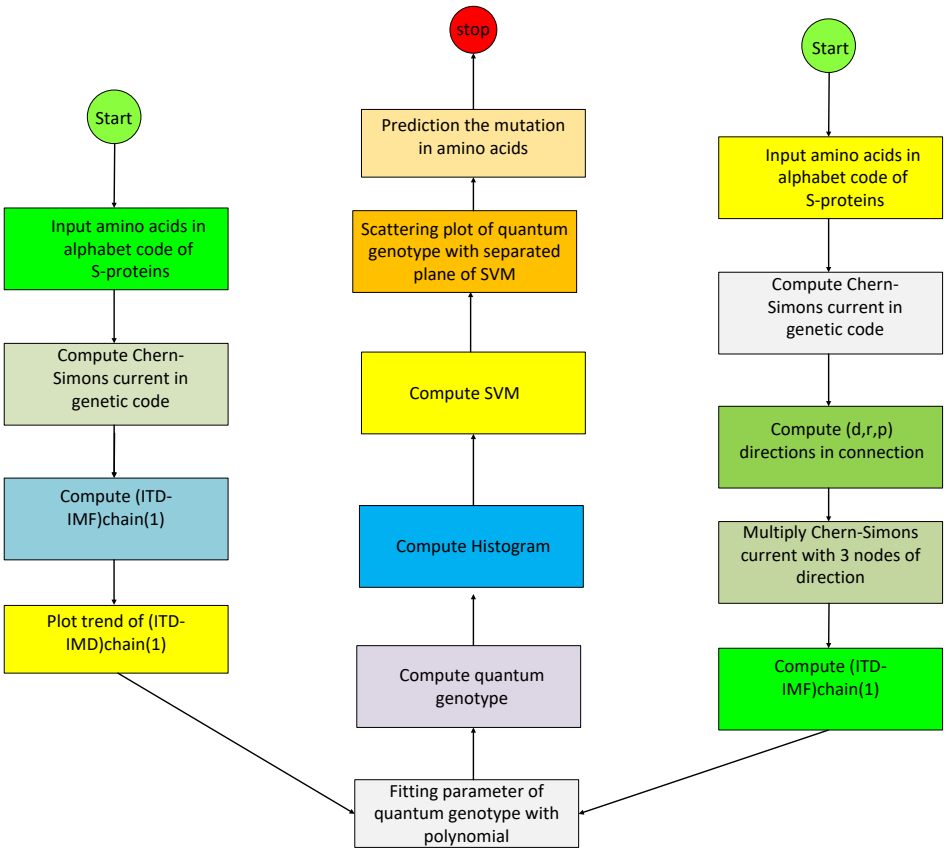

Fig. 3 A schematic diagram illustrating the algorithm used to compute quantum genotype and predict the mutation of S protein using a support vector machine.

Omicron BA.5 are -2.3953×10^{-8} and 2.5745×10^{-5}, which are significantly different from the first SARS-CoV-2 empirical data collected in January 2020. It is also noteworthy that Omicron has a lower quantum genotype than the original SARS-CoV-2. Therefore, we conclude from this empirical analysis of time series data that the evolutional path of the SARS-CoV-2 genome tends to decrease the quantum genotype β_1.

We did a further empirical analysis to detect the shape of the histogram of quantum genotype β_1 in which 3 peaks of mutation appeared in the histogram frequency distribution. The empirical analysis of the probability distribution of time series data revealed that the upcoming new variant of concern in December 2021 is expected to be located on the left shift of the frequency of quantum genotype as compared to the location of the highest peak observed in the year 2020. Omicron is on the most left peak of histogram (Fig. 5A). The scattering plot and fitting equation for the quantum genotype β_2 and β_3 of S protein of S clade with other 14 variants with support vector over separated hyperplane (Fig. 6A). The classification using a separate plane revealed different clusters between 2 groups. The first group consisted of Iota, Alpha, Mu, Omicron BA.1–BA.5, and Kappa, while the second group included Delta, Beta, Lambda, Eta, and Gamma. Additionally, a scattering plot of the quantitative regression algorithm was utilized to detect the mutation and measure the quantum genotype of 30 samples of Omicron BA.5 and other Omicron variants (Fig. 6B). The result of scattering plot of 30 samples of Omicron BA.1-4 and BA.5 showed that BA.5 is the most similar cluster to BA.4.

Prediction of mutation in S Protein

The calculation of quantum genotype was performed on 420 selected sequences of S proteins in 14 variants of concern (Fig. 3). Support vector machine with radial base function was used as the chosen kernel for training with various ratios of the number of samples to detect sensitivity. The training was performed with 50%, 60%, 70%, 80%, and 90% of random amino acids 401–520 in S clade and variants of concern. The training was done with 30 samples of 14 variants of concern with 10-fold cross-validation. The percentage of correction counting from 120 amino acids in unknown test samples was used to evaluate the performance of forecasting unknown test samples for

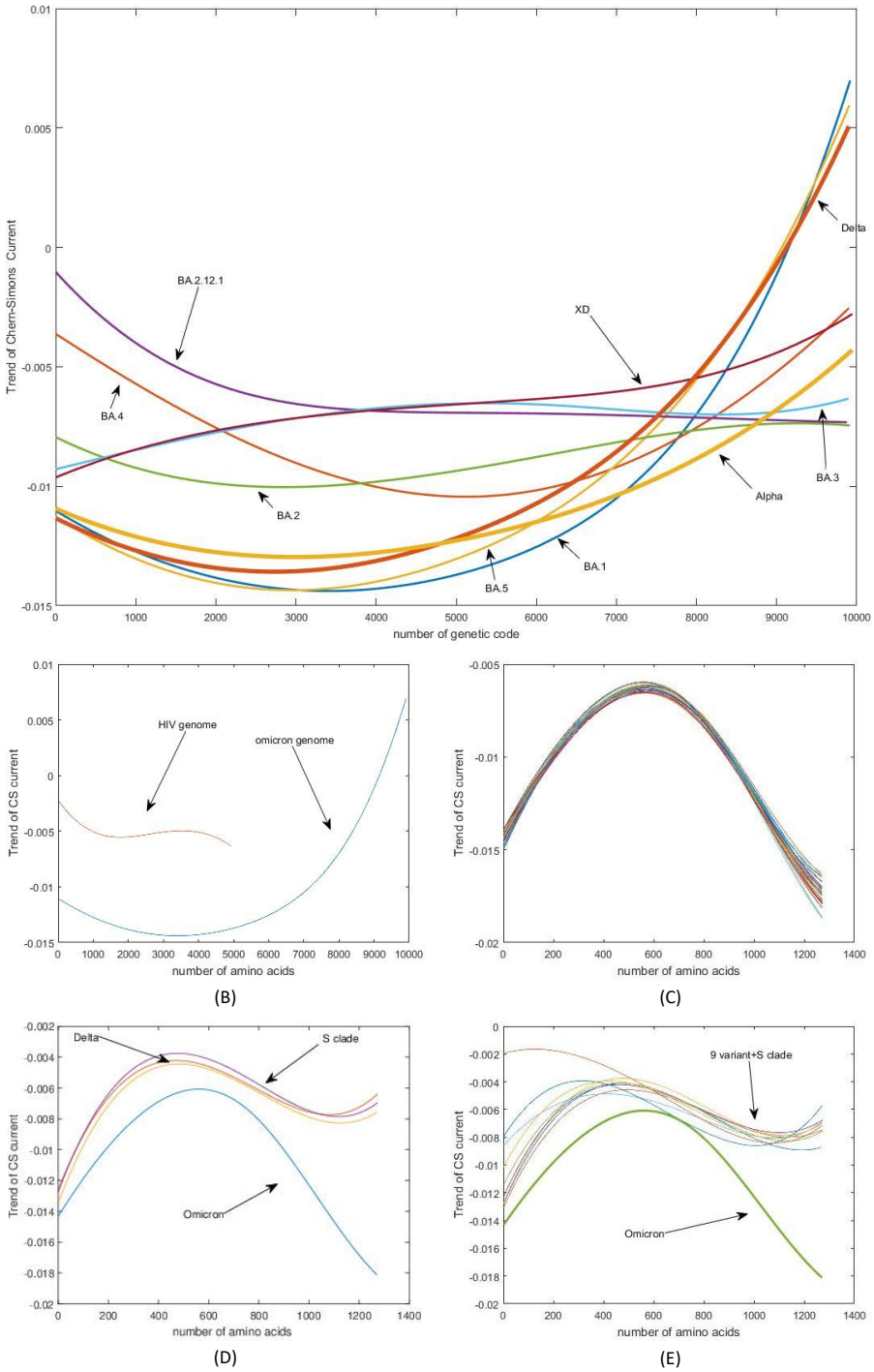

Fig. 4 (A) The comparative plot of the trend of Chern-Simons current of the whole genome of Delta SARS-CoV-2 variants with Omicron variants BA.1–BA.5. In the plot of Deltacron, XD is similar to BA.3. Additionally, BA.5 shows a similar curve to the Delta variant as the other Omicron variants. (B) Comparative plot of the whole genome between HIV and the Omicron variant, showing that Omicron's genome size is longer than that of HIV with a parabolic shape. The initial part of the curve indicates the similarity in the curvature of the graph between these 2 viruses, suggesting that some genes in Omicron's ORF1ab share similar genes with HIV. (C) Plot of 30 samples of S protein of Omicron BA.5 with the trend of Chern-Simons current, which has a parabolic shape with high symmetry and positive curvature. (D) Plot comparing S protein of Omicron with Delta and Alpha variants. (E) Comparative plots of all variants with Omicron, indicating that Omicron's graph is in the ground state and differs from other variants.

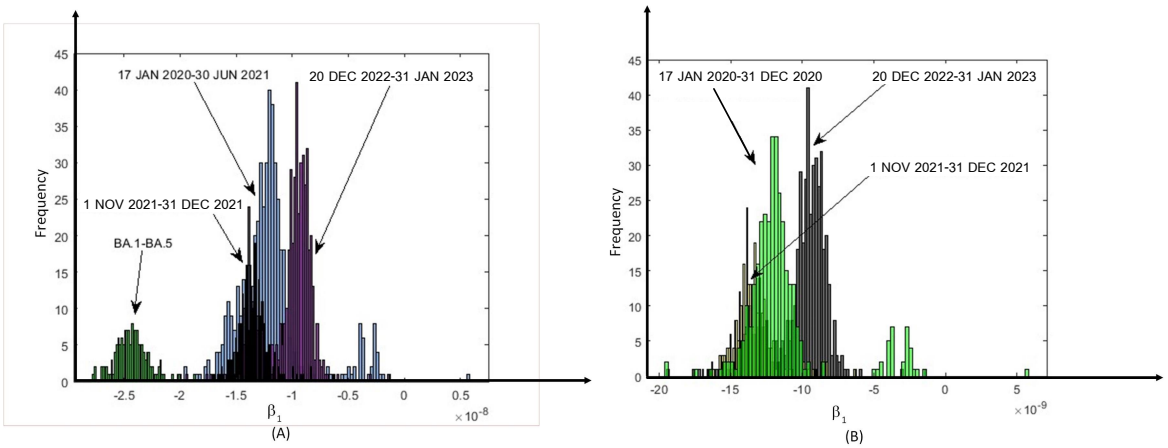

Fig. 5 (A) The plot of the histogram of Omicron BA.1-BA.5 (green color) with the time series of S protein from 17 January 2020 to 30 June 2021 in blue color. The black color in the plot shows the histogram of the time series of S protein from 1 November 2021 to 31 December 2021. The magenta color is the time series of S protein from 20 December 2022 to 31 January 2023.(B) The plot of the histogram of the time series of S protein from 17 January 2020 to 31 December 2022 in green color. The black color is the time series of S protein from 20 December 2022 to 31 January 2023.

Table 2 The prediction performance of S protein mutation in 120 amino acids located between amino acids number 401–520, using support vector machine. Evaluation is based on 14 variants of concern, and the input comprises 30 samples from each variant with 10 folds of cross validation. All values are shown as percentage of correction.

Variant	% Training					
	50	60	70	80	90	Average
BA.1	98.13	98.85	98.75	98.81	98.61	98.63
BA.2	97.77	98.64	98.47	98.47	99.02	98.47
BA.3	98.69	98.19	98.10	98.26	99.02	98.45
BA.4	97.38	97.53	97.77	97.91	98.33	97.78
BA.5	97.83	98.19	98.42	98.26	99.02	98.34
Delta	92.72	92.84	93.19	91.80	95.13	93.13
Iota	97.91	97.98	97.73	97.98	97.77	97.87
Beta	97.77	98.22	98.37	98.19	97.77	98.06
Kappa	97.30	97.95	97.45	97.91	97.63	97.64
Eta	97.58	98.19	97.40	98.26	98.61	98.00
Alpha	98.52	98.61	98.37	98.54	98.33	98.47
Lambda	97.08	97.46	97.40	97.63	96.80	97.27
Gamma	98.58	98.75	99.21	98.89	98.19	98.72
Mu	98.06	97.84	98.51	98.47	98.75	98.32
Average	97.52	97.80	97.79	97.81	98.07	97.79
% Change	–	0.28	0.01	0.02	0.26	0.14

residue numbers 401–520, as presented in Table 2. The results of training with 30 samples were as follows: 93.13% for the Delta variant, 98.47% for the Alpha variant, 98.06% for the Beta, 98.72% for the Gamma variant, 98.32% for the Mu variant, 98.00% for the Eta variant, 97.64% for the Kappa variant, 97.87% for the Iota variant, 97.27% for the Lambda variant, 98.63% for BA.1, 98.47% for BA.2, 98.45% for BA.3, 97.78% for BA.4, and 98.34% for BA.5.

DISCUSSION

From the analysis of the shape of the histogram of quantum genotype in Fig. 5B, it is apparent that there has been a shift in the mean value of the highest peak of the frequency to the left side of the histogram from 2020 to 2021, followed by a shift back to the right in January 2023. This suggests that the evolutional path of SARS-CoV-2 in terms of quantum genotype has been in a new direction that is opposite to the period between 2020 and 2021. The shift observed in January 2023 reveals different states of quantum genotype in the years 2020–2022, likely resulting from the recombination of the viral genome and the COVID-19 vaccine.

Regarding the performance measurement of the prediction of unknown amino acids in 14 variants, our study achieved an average performance of 97.79%, which is slightly lower than the previous study by Ali et al [18] that used an alignment method and achieved a performance of 99.4%. However, our study was able to cover the Omicron variants BA.1–BA.5, which were not included in the previous study. Furthermore, we focused on the quantitative measurement of the sensitivity of the forecasting model when increasing the sample size. Our results showed that when the sample size was increased from 3 to 30 samples, the performance of prediction was significantly improved from 87.88% to 97.79%. This highlights the importance of using a larger sample size for training the model to improve its accuracy. Overall, our empirical analysis provides insights into the evolutional path of SARS-CoV-2 in terms of quantum genotype and demonstrates the potential of our method for predicting mutations in SARS-CoV-2 variants.

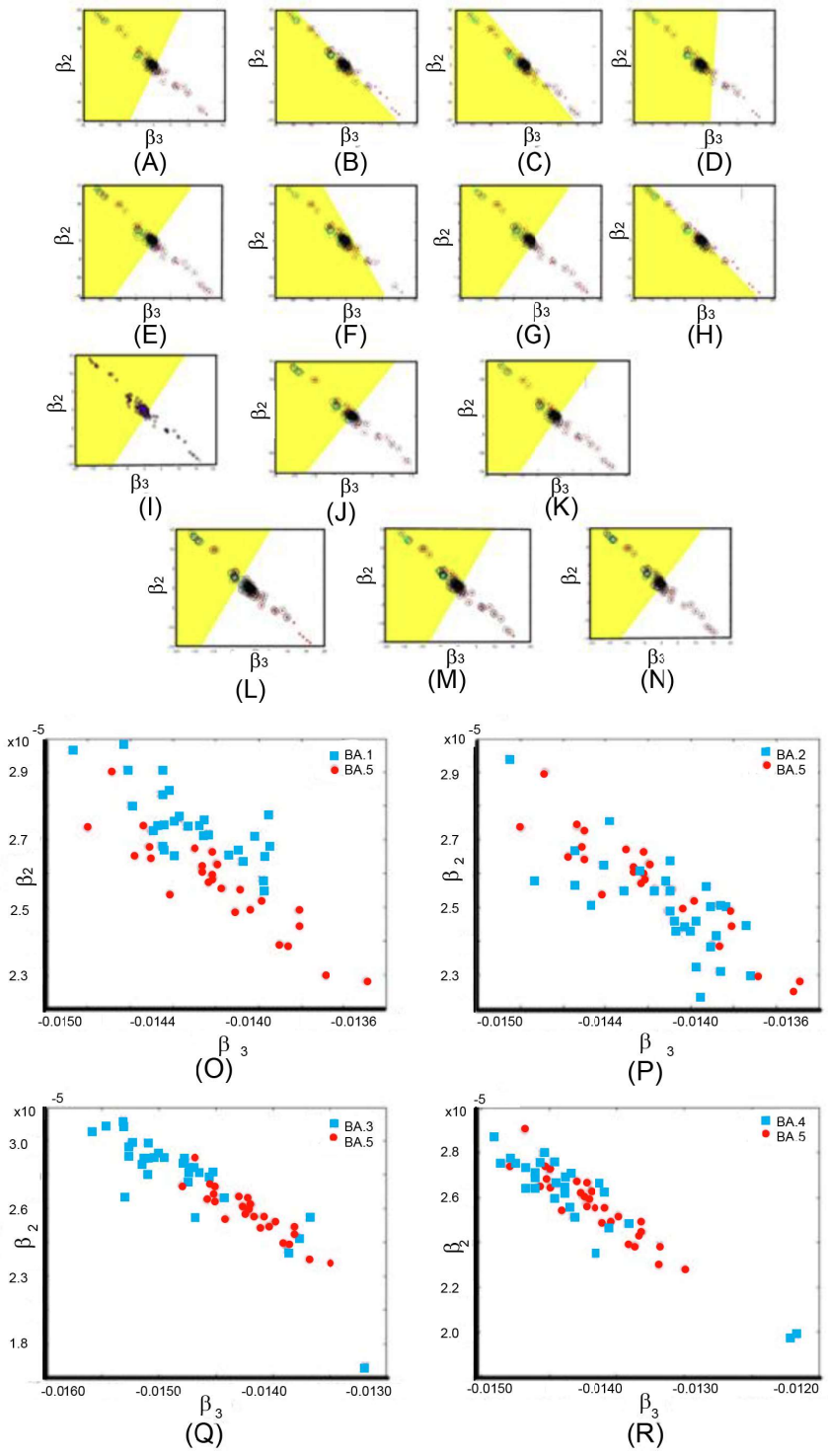

Fig. 6 The scattering plot of quantum genotype β_2 and β_3 from the fitting equation $y = \beta_1 x^2 + \beta_2 x + \beta_3$ of S protein of S clade with other variants with support vector over separated hyperplane. (A) The blue color representing unmuted amino acids in the S clade. The red color representing mutated amino acids in Alpha, (B) Beta, (C) Delta, (D) Eta, (E) Iota, (F) Lambda, (G) Mu, (H) Gamma, (I) Kappa, (J) Omicron BA.1, (K) Omicron BA.2, (L) Omicron BA.3, (M) Omicron BA.4, and (N) Omicron BA.5. We notice from the slope of the hyperplane that Omicron variants are similar to Alpha and Mu rather than Delta variant. The scattering plot of quantum genotype β_2 and β_3 from the fitting equation $y = \beta_1 x^2 + \beta_2 x + \beta_3$ of S protein of Omicron BA.5 with Omicron (O) BA.1, (P) BA.2, (Q) BA.3, and (R) BA.4. In the plot, BA.5 is the most similar cluster to BA.4.

CONCLUSION

This study introduced the concept of quantum genotype and measured it using empirical data analysis of the Chern-Simons current in Omicron BA.5 and related variants. The invariant properties of mutation in quantum genotype were investigated, and a new mathematical model of the quantum biological coordinate system was proposed. The results demonstrated a new approach for the quantitative measurement of the changing curvature of S protein mutation in each amino acid. The performance of the prediction model for unknown amino acids in 14 variants was 97.79%, which is slightly lower than the alignment method but with faster computational time. Additionally, the changing histogram of quantum genotype with time was observed, and it was found that the coming new variant of concern in the year 2022 is located on the left shift of the frequency of the quantum genotype more than the location of the highest peak of the year 2021. This result can be applied to the quantitative measurement of the quantum genotype of future unknown variants. In summary, this research work provides a new perspective on the study of SARS-CoV-2 and a new tool for detecting and predicting new variants of concern.

Acknowledgements: This research was supported by the National Research Council of Thailand (NRC N42A650231). KK thanks the Chulalongkorn University Graduate Scholarship to commemorate the 72nd Anniversary of his majesty King Bhumibol Adulyadej and the 90th Anniversary of Chulalongkorn University, Ratchadapisek Sompote Fund (GCUGR1125651033D).

REFERENCES

1. Callaway E (2022) What Omicron's BA.4 and BA.5 variants mean for the pandemic. *Nature* **606**, 848–849.
2. Tegally H, Moir M, Everatt J, Giovanetti M, Scheepers C, Wilkinson E, Subramoney K, Makatini Z, et al (2022) Emergence of SARS-CoV-2 omicron lineages BA.4 and BA.5 in South Africa. *Nat Med* **28**, 1785–1790.
3. Arora P, Kempf A, Nehlmeier I, Schulz SR, Cossmann A, Stankov MV, Jäck HM, Behrens GM, et al (2022) Augmented neutralisation resistance of emerging omicron subvariants BA.2.12.1, BA.4, and BA.5. *Lancet Infect Dis* **22**, 1117–1118.
4. Cao Y, Yisimayi A, Jian F, Song W, Xiao T, Wang L, Du S, Wang J, et al (2022) BA.2.12.1, BA.4 and BA.5 escape antibodies elicited by Omicron infection. *Nature* **608**, 593–602.
5. Geng J, Yang X, Wang K, Wang K, Chen R, Chen ZN, Qin C, Wu G, et al (2023) Immunological and metabolic characteristics of the Omicron variants infection. *Signal Transduct Target Ther* **8**, 42.
6. Andersen KG, Rambaut A, Lipkin WI, Holmes EC, Garry RF (2020) The proximal origin of SARS-CoV-2. *Nature Med* **26**, 450–452.
7. Li F (2016) Structure, function, and evolution of coronavirus spike proteins. *Ann Rev Virol* **3**, 237–261.
8. Chaisawasd P, Ittisoponpisan S (2023) Computational analysis of the novel Thailand-specific mutations in SARS-CoV-2 spike glycoprotein sequences. *ScienceAsia* **49**, 77–84.
9. Karim SSA, Karim QA (2021) Omicron SARS-CoV-2 variant: a new chapter in the COVID-19 pandemic. *Lancet* **398**, 2126–2128.
10. Tsuchiya K, Maeda K, Matsuda K, Takamatsu Y, Kinoshita N, Kutsuna S, Hayashida T, Gatanaga H, et al (2023) Neutralization activity of IgG antibody in COVID-19-convalescent plasma against SARS-CoV-2 variants. *Sci Rep* **13**, 1263.
11. Kannan SR, Spratt AN, Sharma K, Chand HS, Byrareddy SN, Singh K (2022) Omicron SARS-CoV-2 variant: Unique features and their impact on pre-existing antibodies. *J Autoimmun* **126**, 102779.
12. Petersen E, Ntoumi F, Hui DS, Abubakar A, Kramer LD, Obiero C, Tambyah PA, Blumberg L, et al (2022) Emergence of new SARS-CoV-2 variant of concern Omicron (B.1.1.529)-highlights Africa's research capabilities, but exposes major knowledge gaps, inequities of vaccine distribution, inadequacies in global COVID-19 response and control efforts. *Int J Infect Dis* **114**, 268–272.
13. Sun K, Tempia S, Kleynhans J, von Gottberg A, McMorrow ML, Wolter N, Bhiman JN, Moyes J, et al (2023) Rapidly shifting immunologic landscape and severity of SARS-CoV-2 in the Omicron era in South Africa. *Nat Commun* **14**, 246.
14. Delcher AL, Phillippy A, Carlton J, Salzberg SL (2002) Fast algorithms for large-scale genome alignment and comparison. *Nucleic Acids Res* **30**, 2478–2483.
15. Pinčák R, Kanjamapornkul K, Bartoš E (2020) Cohomology theory for biological time series. *Math Methods Appl Sci* **43**, 552–579.
16. Al-Zyoud W, Haddad H (2021) Dynamics prediction of emerging notable spike protein mutations in SARS-CoV-2 implies a need for updated vaccines. *Biochimie* **191**, 91–103.
17. Hossain MS, Pathan ASU, Islam MN, Tonmoy MIQ, Rakib MI, Munim MA, Saha O, Fariha A, et al (2021) Genome-wide identification and prediction of SARS-CoV-2 mutations show an abundance of variants: Integrated study of bioinformatics and deep neural learning. *Inf Med Unlock* **27**, 100798.
18. Ali S, Sahoo B, Ullah N, Zelikovskiy A, Patterson M, Khan I (2021) A *k*-mer based approach for SARS-CoV-2 variant identification. In: Wei Y, Li M, Skums P, Cai Z (eds) *Bioinformatics Research and Applications, ISBRA 2021*, Lecture Notes in Computer Science, vol 13064, Springer, Cham., pp 153–164.
19. Câmara GB, Coutinho MG, Silva LM, Gadelha WV, Torquato MF, Barbosa RM, Fernandes MA (2022) Convolutional neural network applied to SARS-CoV-2 sequence classification. *Sensors* **22**, 5730.
20. Sathipati SY, Shukla SK, Ho SY (2022) Tracking the amino acid changes of spike proteins across diverse host species of severe acute respiratory syndrome coronavirus 2. *Iscience* **25**, 103560.
21. Qiang XL, Xu P, Fang G, Liu WB, Kou Z (2020) Using the spike protein feature to predict infection risk and monitor the evolutionary dynamic of coronavirus. *Infect Dis Pov* **9**, 33.
22. Köchl K, Schopper T, Durmaz V, Parigger L, Singh A,

Krassnigg A, Cespugli M, Wu W, et al (2023) Optimizing variant-specific therapeutic SARS-CoV-2 decoys using deep-learning-guided molecular dynamics simulations. *Sci Rep* **13**, 774.

23. Kumar S, Thambiraja TS, Karuppanan K, Subramaniam G (2022) Omicron and Delta variant of SARS-CoV-2: a comparative computational study of spike protein. *J Med Virol* **94**, 1641–1649.
24. Berne E (1971) Away from a theory of the impact of interpersonal interaction on non-verbal participation. *Trans Anal J* **1**, 6–13.
25. Kanjamapornkul K, Rongrotmongkol T, Hannongbua S (2021) Frenet-Serret formulas for moving frame with (d, r, p)-layer in protein folding. *J Sci Sci Educ* **4**, 38–50.
26. Capozziello S, Pincak R, Kanjamapornkul K, Saridakis EN (2018) The Chern-Simons current in systems of DNA-RNA transcriptions. *Ann Phys* **530**, 1700271.
27. Pinčák R, Kanjamapornkul K, Bartoš E (2019) Forecasting Laurent polynomial in the Chern-Simons current of V3 loop time series. *Ann Phys* **531**, 1800375.
28. Nan BG, Zhang S, Li YC, Kang XP, Chen YH, Li L, Jiang T, Li J (2022) Convolutional neural networks based on sequential spike predict the high human adaptation of SARS-CoV-2 Omicron variants. *Viruses* **14**, 1072.
29. Berne E (1977) *Intuition and ego states: The origins of transactional analysis: a series of papers*. TA Press, San Francisco.
30. Pinčák R, Kanjamapornkul K, Bartoš E (2020) A theoretical investigation on the predictability of genetic patterns. *Chem Phys* **535**, 110764.
31. Kanjamapornkul K, Rongrotmongkol T, Hannongbua S (2022) Detection of spike protein mutation in SARS-CoV-2 variants of concern with support vector classification over (d, r, p)-layer coordinate system. *J Sci Sci Educ* **5**, 58–71.
32. Kanjamapornkul K, Pinčák R, Chunithpaisan S, Bartoš E (2017) Support spinor machine. *Digit Signal Proc* **70**, 59–72.
33. Kanjamapornkul K, Pinčák R (2016) Kolmogorov space in time series data. *Math Methods Appl Sci* **39**, 4463–4483.