IAES International Journal of Artificial Intelligence (IJ-AI)
Vol. 15, No. 4, August 2026, pp. 3189~3201
ISSN: 2252-8938, DOI: 10.11591/ijai.v15.i4.pp3189-3201

Immuno-bioinformatics analysis of progressive alignment and logic learning machine-derived viral conserved sequences

Felza Ridho[1], Mohammad Isa Irawan[1], Nurul Hidayat[1], Awik Puji Dyah Nurhayati[2]
[1]Department of Mathematics, Institut Teknologi Sepuluh Nopember, Surabaya, Indonesia
[2]Department of Biology, Institut Teknologi Sepuluh Nopember, Surabaya, Indonesia

Article Info

Article history:

Received Feb 6, 2025
Revised Mar 18, 2026
Accepted Jul 9, 2026

Keywords:

Conserved sequence
Immuno-bioinformatics
Logic learning machine
Progressive alignment
Switching neural network

ABSTRACT

This study proposes a method to identify conserved sequence positions in viral data, with the primary goal of facilitating their translation into proteins. The approach aims to support the early detection of sub-sequences that hold promise as vaccine candidates. The method involves five key steps. First, mutation analysis using the Kimura model: analyzed mutations in the viral data using the Kimura model to provide insights into sequence variations. Second, alignment method selection to align sequences effectively, the progressive alignment approach with the neighbor-joining algorithm was chosen. Third, hybrid algorithm for identifying unchanged sequences: a hybrid algorithm that combines progressive alignment and a logic learning machine (LLM) was employed to identify unchanged sequences. Fourth, determining conserved sequences: based on the longest unchanged sequence, conserved regions within the severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) data were identified. Fifth, immuno-bioinformatics analysis for vaccine candidate peptides: using immuno-bioinformatics, protein sequences were analyzed to identify potential vaccine candidates based on B cell epitopes. Experimental validation confirmed the algorithm's ability to identify candidate proteins and generate vaccine-worthy peptides from conserved protein sequences. The streamlined approach focuses on conserved protein sequences, ensuring efficient and targeted identification of vaccine candidates. By leveraging these conserved regions, contributions can be made to the expedited development of vaccines.

Corresponding Author:

Mohammad Isa Irawan
Department of Mathematics, Institut Teknologi Sepuluh Nopember
ITS Campus, Sukolilo, Surabaya, Indonesia
Email: mii@its.ac.id

1. INTRODUCTION

Several disease-causing viruses have spread globally. The most recent pandemic originated from pneumonia known as COVID-19, a severe acute respiratory syndrome (SARS), caused by SARS coronavirus 2 (SARS-CoV-2), a novel virus belonging to the *Coronaviridae* family. First reported in December 2019 in Wuhan, China, the virus rapidly spread worldwide. As of February 26, 2021, SARS-CoV-2 has infected approximately 112.20 million people and resulted in around 2.49 million deaths globally [1]. Additionally, the World Health Organization (WHO) provides updates on other viruses [2], suggesting that there remains potential for their global spread.

One solution to prevent the ongoing spread is the vaccination process [3], [4]. To determine vaccine candidates, it is essential to process sequence data from virus variants. An approach to vaccine development

involves utilizing in silico methods. These methods entail the analysis of protein sequences to identify potential epitopes and peptides, which can serve as vaccine candidates for related viruses [5]–[7].

This study combined bioinformatics and immunoinformatics analyses, namely immuno-bioinformatics [8]–[12]. Bioinformatics analysis was performed by combining sequence alignment and Boolean logic operations. The alignment process is performed on existing deoxyribonucleic acid (DNA) sequences; if there are more than two sequences, this is called multiple sequence alignment [13]–[16]. This process determines the relationship of some virus variants, which can be in the form of a phylogenetic tree [17]–[21], mutation sequences in virus variants, or the determination of conserved sequences, namely, sequences that did not change from the beginning of virus emergence. For multiple sequence alignments and conserved sequences, a specialized method is required to determine their positions. In contrast to previous studies [22]–[24] and studies related to the algorithm used [25] proposed the Java pattern finder (JaPaFi) method, in which it is difficult to find a conserved sequence using the algor, this study uses a combination of progressive alignment and Boolean logic to overcome this limitation. In previous studies [26]–[28], Boolean logic was described as a way to test a value as true or false, yes or no, and in numerical terms, as 1 or 0. During its development, Boolean logic operations were closely related to switching neural networks [29], which were later further developed under the name logic learning machine (LLM) [30]–[32]. The authors [30]–[32] showed that LLMs were implemented as formula-based classification using switching neural networks. Therefore, using an LLM, the similarity between the sequences in the alignment results can be tested. The main objective of this research is to assist in the initial stages of vaccine-related research, namely, determining the focus of the sequence section being studied in the laboratory.

This objective needs to be supported by an immunoinformatics approach used to identify vaccine candidates obtained in silico. Based on related research [33]–[36] regarding peptides in conserved sequences, there is no specific method in their research for obtaining conserved sequences for further immunoinformatics. To address these problems, we propose an algorithm that combines progressive alignment and an LLM to determine the position of conserved sequences at both the DNA and protein levels. This approach leverages the strengths of sequence alignment and Boolean logic operations to identify conserved sequences critical for vaccine development. By integrating bioinformatics and immunoinformatics analyses, the method not only enhances the accuracy of conserved region identification but also facilitates the early detection of promising vaccine candidates. This dual approach ensures a more efficient and targeted identification process, significantly contributing to the expedited development of vaccines against rapidly evolving viruses.

The remainder of this paper is structured as follows. Section 2 presents the proposed methodology. Section 3 presents the experimental results, including the evaluation of proposed method on viral sequences and their variants, as well as the immunoinformatics analysis. Finally, section 4 concludes the paper.

2. METHOD

The main aim of this study was to identify conserved sequences that are subsequences of the main sequence and can be processed further for immunoinformatics analysis. The determination of conserved sequences used the proposed method. The immunoinformatics analysis using that method consisted of several stages, as shown in Figure 1. The stages are explained in the following steps:

i) Collecting sequence data for analysis.
ii) Preprocessing data by analyzing the mutation using Needleman-Wunsch alignment and transforming the data into genetic distance values using the Kimura model.
iii) Processing genetic distance values into the form of a phylogenetic tree using the neighbor-joining algorithm as a guide tree and progressive alignment to perform multiple sequence alignment.
iv) Combining progressive alignment and LLM to determine the conserved sequence.
v) Experiments on real virus data using an in silico method to discover peptides as vaccine candidates with immunoinformatics analysis using BepiPred 2.0 and VaxiJen v2.0.

2.1. Bioinformatics analysis: conserved sequence

2.1.1. Needleman-Wunsch alignment

Needleman-Wunsch alignment is a method for obtaining the optimal overall global alignment between two selected sequences [37]. Suppose there are two sequences, $D = d_1, d_2, \cdots, d_n$ and $E = e_1, e_2, \cdots, e_m$. A matrix F of order $(n+1)(m+1)$ is formed. The elements in the i-th row and j-th column, $F(i,j)$, for $1 < i < n$ and $1 < j < m$ are the same as the optimal alignment scores between $d_1, d_2, \cdots, d_i$ and $e_1, e_2, \cdots, e_j$. The element $F(i,0)$ for $1 < i < n$ is the score of the alignment between $d_1, d_2 \cdots, d_i$ and the gap along i. Likewise, for Element $F(0,j), 1 < j < m$ is the score of the alignment

between $e_1, e_2, \cdots, e_j$ and the gap along j. The F matrix is constructed iteratively, starting with $F(0,0) = 0$. Each element of the matrix is filled from the top-left to the bottom-right. If $F(i-1, j-1)$, $F(i-1, j)$, and $F(i, j-1)$ are known, $F(i, j)$ can be computed using (1).

$$F(i,j) = max \begin{Bmatrix} F(i-1,j-1) + M(d_i, e_j), \\ F(i-1,j) + M(d_i, -), \\ F(i,j-1) + M(-, e_j) \end{Bmatrix} \quad (1)$$

When calculating $F(i, j)$, the direction in which $F(i, j)$ is obtained is stored. After obtaining $F(n, m)$, backtracking was performed to obtain the optimal alignment. $F(n, m)$ is represented as the alignment score.

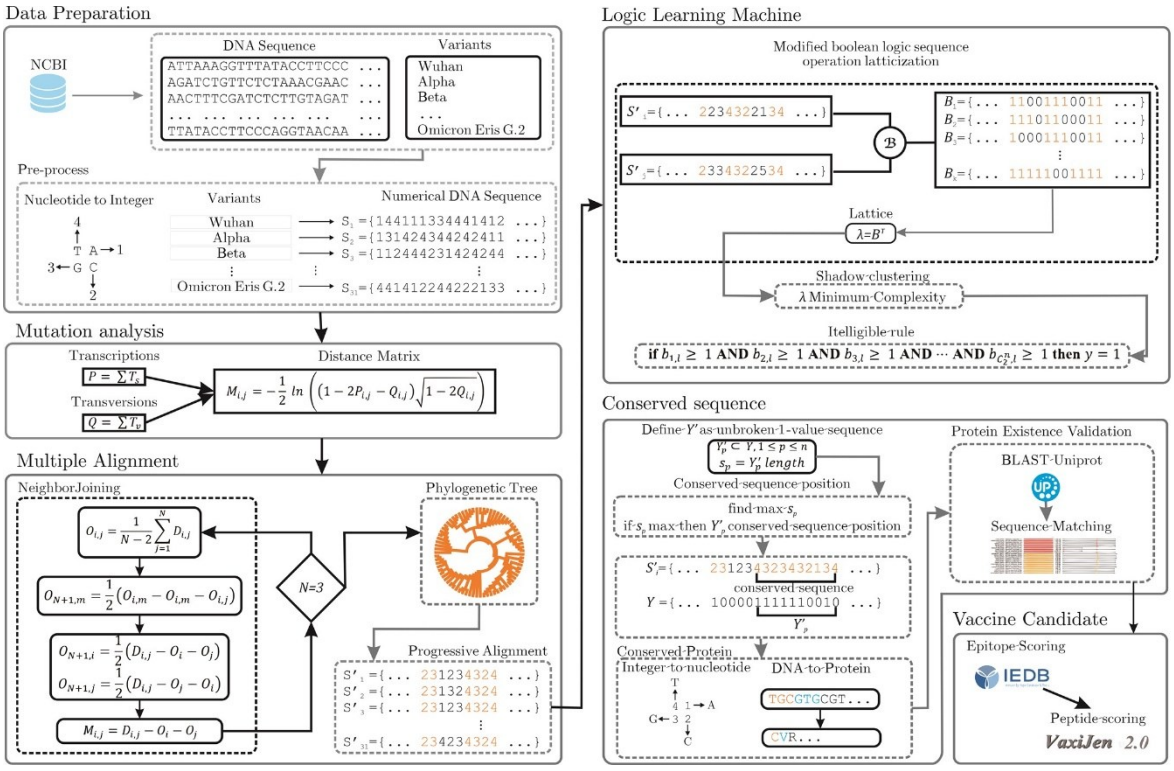

Figure 1. Proposed method for obtaining conserved sequence and immuno-bioinformatics analysis

After alignment, the two aligned sequences are expressed as $D' = d'_1, d'_2, \cdots, d'_h$ and $E' = e'_1, e'_2, \cdots e'_h$. A different part of the sequence was used that changed the nucleotide from the first sequence to the second sequence, which is call a mutation. The mutations used were divided into transitions and transversions [38]. A transition is a change in a nucleotide to another nucleotide in a similar base or the change of A → G, G → A, C → T, and T → C and transversion is a change in a nucleotide to another nucleotide in an unsimilar base; the changes are A → C, A → T, C → A, C → G, G → C, G → T, T → A, and T → G. The process for determining the number of transitions and transversions is presented in Algorithm 1. The number of mutations in Algorithm 1 can then be used to apply Kimura's model.

Algorithm 1. Finding the occurrence of transitions and transversions

```
inputs: Sequence D', Sequence E', Sequence length h
output: Number of transitions P and transversions Q
procedure transition_transversion (D',E')
        Determine T_s and T_v array in h length;
        T_s = t_s1, t_s2, ..., t_sh;
        T_v = t_v1, t_v2, ..., t_vh;
        Convert the elements of T_s and T_v into 0 and 1;
        for d'_i, e'_i, i = 1,2, ..., h do
                if d'_i =A and e'_i =C or d'_i =C and e'_i =A or d'_i =G and e'_i =T or d'_i =T and e'_i =G then
```

```
                    t_si = 1;
                else
                    t_si = 0;
                end
        if d'_i =A and e'_i =G or d'_i =A and e'_i =T or d'_i =G and e'_i =A or d'_i =G and e'_i =C or d'_i =C and
e'_i =G or d'_i=C and e'_i =T or d'_i =T and e'_i =A or d'_i =T and e'_i =C then
                    t_vi = 1;
                else
                    t_vi = 0;
                end
        end
        Determine the number of transitions P and transversions Q;
        P = ∑T_s;
        Q = ∑T_v;
end
```

2.1.2. Kimura model

The Kimura model was applied to obtain the distance matrix between the two sequences. The distance matrix was created using this model. The model for creating the distance matrix is expressed as follows [39]:

– Definition 1. If P number of transitions occurring between two sequences, and Q is the number of transversions that occur. The distance between two sequences is defined as (2).

$$M_{i,j} = -\frac{1}{2}\, ln\, \left(\left(1 - 2P_{i,j} - Q_{i,j}\right)\sqrt{1 - 2Q_{i,j}}\right) \quad (2)$$

After obtaining the M value and forming a distance matrix, the distance matrix was used to form a phylogenetic tree using the neighbor-joining algorithm.

2.1.3. Neighbor-joining algorithm

The neighbor-joining algorithm was used to construct a phylogenetic tree from sequence distances. The method iteratively computes a correction term O_i from the distance matrix, then selects the pair with the smallest $M_{i,j} = D_{i,j} - O_i - O_j$ [39]. That pair is merged into a new operational taxonomic unit (OTU), and distances are recalculated. This process is repeated until only three OTUs remain. The resulting tree then served as a guide tree for progressive alignment.

2.1.4. Progressive alignment

The implementation used progressive alignment, a multiple-sequence alignment heuristic. Starting with two sequences aligned, a third is added and aligned, and the process repeats until all sequences are included, guided by a phylogenetic tree built from Kimura-based pairwise distances [39]. The result is a multiple sequence alignment, formally defined as an expansion of the original sequences with inserted gaps such that all expanded sequences have the same length, and no column consists entirely of gaps. Because alignment alone cannot pinpoint conserved regions, we combined the alignment output with a LLM to identify conserved sequence positions.

2.1.5. Logic learning machine

In this study, a LLM was employed as an efficient implementation of a switching neural network model to determine the position of the conserved sequence [29]. LLMs use latticization, which is closely related to logic testing. The results of this test are based on Boolean logic, and the output of latticization is $\{0,1\}^v$. In Boolean algebra, logic levels are represented by symbols, making it a type of logic algebra. Although any symbol can be used, alphabetic letters are the most frequently used. Any letter used as a variable can have a value of 1 or 0, because logic levels are often represented by symbols 1 and 0. Boolean logic operations are defined as follows [40].

– Definition 2. Boolean algebra is an algebra $\mathcal{B} = (B,\wedge,\vee,\bar{},0,1)$ with type $(2,2,1,0,0)$ which satisfies the following conditions:
  i) $(B,\wedge,\vee)$ is distributive, which has 0 as the smallest element and 1 as the largest element.
  ii) $\bar{}: B \rightarrow B$ is a unary operator so x is the complement of x for $x \in B$.

The $\wedge$ operator is also known as the AND operator, and $\vee$ is known as the OR operator. The proposed algorithm is obtained by performing Boolean algebra on S' sequences. Boolean logic operations are performed for every combination of i,j, where $i = 1,2,3,\cdots,n$ and $j = 1,2,3\cdots,n$ with the condition $j > i$; thus, there are many Boolean logic sequences in the first stage, totaling the combination $C_2^n = \frac{n!}{(n-2)!2!}$.

The Boolean logic results in the initial stage are expressed as $B_k = \{b_{k,1}, b_{k,2}, \cdots, b_{k,l}, \cdots, b_{k,m}\}$, $b_{k,l} = \{0,1\}$, where $k = 1,2,3, \cdots, C_2^n$, l is the sequence position, and m is the sequence length. $b_{k,l}$ has a value of 0 if the two compared $S_i' = s_{i,1}', s_{i,2}', \cdots, s_{i,l}', \cdots, s_{i,m}'$ and $S_j' = s_{j,1}', s_{j,2}', \cdots, s_{j,l}', \cdots, s_{j,m}'$ sequence elements in l position are not the same and $b_{k,l}$ has a value of 1 if the compared S_i' and S_j' sequence elements in l position have the same value. Therefore, in accordance with the latticization of Boolean logic, the result at this stage is $\{0,1\}^{C_2^n}$.

In the context of classification, let $x \in \Re^d$ represent a d-dimensional example that needs to be labeled with a categorical output y, which is assigned to one of q potential classes. The LLM aims to create a model $g(x)$, or classifier, using a training set S of n pairs (x_i, y_i), obtained from prior observations. For most input patterns x, the model $g(x)$ correctly predicts the response $y = g(x)$. When considering element x_j, two distinct scenarios arise: i) for ordered variables, x_j varies within an interval $[a, b]$ on the real axis, with an existing ordering relationship among its values; and ii) for nominal (categorical) variables, x_j can only take values from a finite set, without any inherent ordering relationship. The LLM generates an interpretable model $g(x)$ described by a set of m rules r_k, expressed in if-then form: if <premise> then <consequence>. For ordered variable x_1 in the domain $[100, 300]$ and nominal component x_2 with values $\{D, E, F\}$, a possible rule r_1 can be expressed as follows:

$$\text{if } x_1 > 2000 \text{ AND } x_2 \in \{E, F\} \text{ then } y = 0$$

where 0 denotes one of the q possible assignments (classes).

Therefore, with the results obtained in the previous stage, BS_k' and $BS_k' = \{1,0\}$, the class condition to be taken is the class that expresses the similarities of k-sequences in position l. Thus, if:

$$\begin{aligned} B_1 &= b_{1,1}, b_{1,2}, \cdots, b_{1,l}, \cdots, b_{1,m}, \\ B_2 &= b_{2,1}, b_{2,2}, \cdots, b_{2,l}, \cdots, b_{2,m}, \\ &\vdots \\ B_{C_2^n} &= b_{C_2^n,1}, b_{C_2^n,2}, \cdots, b_{C_2^n,l}, \cdots, b_{C_2^n,m} \end{aligned}$$

The rule used is that if $b_{1,l} \geq 1$ AND $b_{2,l} \geq 1$ AND $b_{3,l} \geq 1$ AND $\cdots$ AND $b_{C_2^n,l} \geq 1$ then $y_l = 1$, which expresses the minimum complexity of the lattice. At this stage, the sequence is expressed in the form $Y = \{y_1, y_2 \cdots, y_i \cdots, y_m\}$ with $y_i = \{0,1\}$, so there are two classes, and class "1" expresses the position that is determined as the conserved sequence position. In the prior process, shadow clustering occurs in the process [41], and with the input being a binary sequence in this study, the results can be directly interpreted as intelligible rules.

2.1.6. Proposed conserved sequence algorithm

The algorithm for determining a conserved sequence is shown in Algorithm 2, which combines progressive alignment and an LLM to create algorithm. To determine the position of conserved sequence, it is assumed that Y' is unbroken 1-value sequence, $Y' = \{Y_1', Y_2', \cdots, Y_n'\}$, where $Y' \subset Y$, $Y_1' \neq Y_2' \neq \cdots \neq Y_n'$, and it is assumed that the length of Y_p' is s, where $1 < s < m$. Thus, the conserved sequence is expressed as Y_p' with the maximum s value.

Algorithm 2. Conserved sequence position

```
inputs: Sequence S_i, number of sequences n
output: Conserved sequence position Y'
procedure conserved_sequence(S_i, n)
        for (S_i, S_j) do
                Align S_i and S_j using Needleman-Wunsch(S_i, S_j);
                Calculate Transition P_{i,j} = ∑T_{si,j};
                Calculate Transversion Q_{i,j} = ∑T_{vi,j};
        end
        for (S_i, S_j) do
                Construct guide tree PG = neighbor_joining (S_i, M_d);
                Align sequences S_i' = progressive_alignment(S_i, PG);
                m = length(S_i');
        end
        Construct temporary sequence t_k;
        t_1 = 0;
        t_2 = 0;
        for t_k, k = 3,4,⋯, n − 1 do
```

```
            t_k = (k − 2) + t_{k−1};
        end
        for (S'_i, n) do
            for S'_i, i = 1,2,⋯,n − 1 do
                for S'_j, j = i + 1, i + 2,⋯,n do
                    for k, k = 1,2,⋯,m do
                        if S'_{i,k} = S'_{j,k} then
                            B_{(i−1)(n−1)+(j−i)−t_i,k} = 1;
                        else if S'_{i,k} ≠ S'_{j,k} then
                            B_{(i−1)(n−1)+(j−i)−t_i,k} = 0;
                        end
                    end
                end
            end
            lattice λ = B^T, λ_l = λ_1, λ_2, ⋯, λ_{C_2^n};
        end
        Perform shadow clustering of λ_l;
        Select the minimum complexity of λ_l as class "1";
        Convert the minimum complexity into Intelligible rules;
        for λ_l do
        Use intelligible rules to classify λ_l;
        Define class "1" as a position of conserved sequence;
            Transform class of λ_l into binary sequence
            Y = {y_1, y_2, ⋯, y_m};
        end
        find Y'_p ⊂ Y unbroken sequence of 1 of Y';
        if length Y'_p is MAX then
            Y'_p conserved sequence position;
        end
end
```

2.1.7. Central dogma

The central dogma of molecular biology outlines the transformation of genetic information from DNA to ribonucleic acid (RNA) and subsequently to proteins. During transcription, the nucleotide base thymine (T) in DNA is replaced by uracil (U) in RNA [42]. A more complex process occurs during the transition from RNA to protein. The change to a protein is based on codons composed of three base groups in RNA; this process is called translation. The central dogma was used to convert conserved sequences into proteins for further analysis.

2.2. Immunoinformatics analysis

Immunoinformatics analysis was used to determine the level of immunogenicity, or the ability of an antigen to stimulate the formation of specific antibodies [43]. The antibodies used are epitope-specific, targeting certain areas of the antigenic molecule that bind to antibodies or receptors on B cells or T cells; B cells were used in this study. B cells or B lymphocytes play an important role in humoral immunity, whereas other lymphocytes, namely T cells, are crucial in cellular immunity. The main function of B cells is to produce antibodies against antigens. To determine the B cell epitope, the following steps were taken:

i) Scoring residues contained in the protein sequence using BepiPred 2.0; the score obtained from the protein was determined to be a B cell epitope if it had an epitope score ≥0.5 [43].
ii) Residues expressed as epitopes in unbroken protein sequences are referred to as peptides [43].

Scoring was performed for each peptide sequence using VaxiJen v2.0. Peptides are declared as vaccine candidates if the peptide score obtained is ≥ 0.4 [44].

3. RESULTS AND DISCUSSION

To evaluate the algorithm, the sequences of 31 SARS-CoV-2 variants were used, as shown in Table 1. The data in Table 1 were obtained from the National Center for Biotechnology Information (NCBI). It is assumed that each variant represents a similar sequence.

3.1. Mutation analysis

In the initial stages, mutations were examined by analyzing FASTA file containing 31 SARS-CoV-2 variants. The Needleman-Wunsch alignment was used between sequences, focusing on the change in nucleotides between two sequences, namely transitions P and transversions Q, as shown in Table 2. Table 2

lists the transitions and transversions of several SARS-CoV-2 variants. Mutation analysis can determine the percentage of mutations in the SARS- CoV-2 variant using (3).

$$Mutation(\%)_{i,j} = \frac{P_{i,j}+Q_{i,j}}{Max\left\{length_{S_i}, length_{S_j}\right\}} 100\% \quad (3)$$

In (3) conclude that the percentage of mutations that occur in the sequence is not more than 1.5%, which means that the homology to the SARS-CoV-2 virus is ≥ 98.5%, with the highest percentage of mutations found in the theta virus variants B.1.617.3 and Omicron BA.1.1 with a mutation percentage of 1.4678%, and the lowest mutation percentage was found in the Omicron BA.1 and Omicron BA.1.1 variants with a percentage of 0.0067%. In addition, the mutation analysis of SARS-CoV-2 showed that, when variant mutation percentages were compared, the highest value was observed between the Wuhan variant and Omicron XBB.1.16, at 0.3511%. The percentage indicates uncertainty in the observed mutations because the number of mutations is not directly proportional to the time at which the new variant appears.

Table 1. SARS-CoV-2 variants

Symbol	Variants/pango lineage	Symbol	Variants/pango lineage	Symbol	Variants/pango lineage
S_1	Wuhan/B	S_{12}	Mu/B.1.621	S_{23}	Omicron/XBC.1
S_2	Alpha/B.1.1.7K	S_{13}	Mu/B.1.621.1	S_{24}	Omicron/BN.1
S_3	Beta/B.1.351	S_{14}	Theta/B.1.617.3	S_{25}	Omicron/XAY
S_4	Gamma/P.1	S_{15}	Zeta/P.2	S_{26}	Omicron/BQ.1
S_5	Delta/B.1.617.2	S_{16}	Omicron/B.1.1.529	S_{27}	Omicron/XBB.1
S_6	Epsilon/B.1.427	S_{17}	Omicron/BA.1	S_{28}	Omicron/CH.1.1
S_7	Epsilon/B.1.429	S_{18}	Omicron/BA.1.1	S_{29}	Omicron/XBB.1.16
S_8	Eta/B.1.525	S_{19}	Omicron/BA.2	S_{30}	Omicron Eris/EG.5
S_9	Iota/B.1.526	S_{20}	Omicron/BA.3	S_{31}	Omicron Eris/EG.2
S_{10}	Kappa/B.1.617.1	S_{21}	Omicron/BA.4		
S_{11}	Lambda/C.37	S_{22}	Omicron/BA.5		

Table 2. SARS-CoV-2 virus transition and transversion

	S_1		S_2		S_3		S_4		⋯	S_{31}	
	P	Q	P	Q	P	Q	P	Q	⋯	P	Q
S_1	0	0	20	12	17	10	21	5		64	35
S_2	20	12	0	0	29	20	32	17		70	40
S_3	17	10	29	20	0	0	30	15		73	43
S_4	21	5	32	17	30	15	0	0		72	38
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
S_{31}	64	35	70	40	73	43	72	38		0	0

3.2. Distance matrix and phylogenetic tree

The Kimura model was applied using (2) to create the M_d-distance matrix. A higher value of $M_{di,j}$ indicates a greater genetic distance between sequences i and j. Table 3 presents the distance matrix M_d. Table 3 shows that the farthest genetic distance between variants was found in the theta B.1.617.3 and Omicron BA.1.1 virus variants with a distance of 5.319, and the shortest genetic distance was found in the Omicron BA.1 and Omicron BA.1.1 variants with a distance of 1.664. Based on the distance matrix in Table 3, a phylogenetic tree is constructed using the neighbor-joining algorithm. The phylogenetic tree is shown in Figure 2. The phylogenetic tree depicted in Figure 2 illustrates the relationships among SARS-CoV-2 variants, with group 1 comprising all Omicron subvariants and group 2 including variants other than Omicron. Notably, closer kinship is observed within the same branch of the tree.

Table 3. Distance of the SARS-CoV-2 virus

	S_1	S_2	S_3	S_4	S_5	S_6	⋯	S_{31}
S_1	0.0000	0.1070	0.0903	0.0869	0.0702	0.0736	⋯	0.3311
S_2	0.0000	0.0000	0.1641	0.1644	0.1475	0.1509	⋯	0.3689
S_3	0.0000	0.0000	0.0000	0.1507	0.1137	0.1205	⋯	0.3882
S_4	0.0000	0.0000	0.0000	0.0000	0.1240	0.1274	⋯	0.3686
S_5	0.0000	0.0000	0.0000	0.0000	0.0000	0.0503	⋯	0.3754
S_6	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	⋯	0.3786
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
S_{31}	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	⋯	0.0000

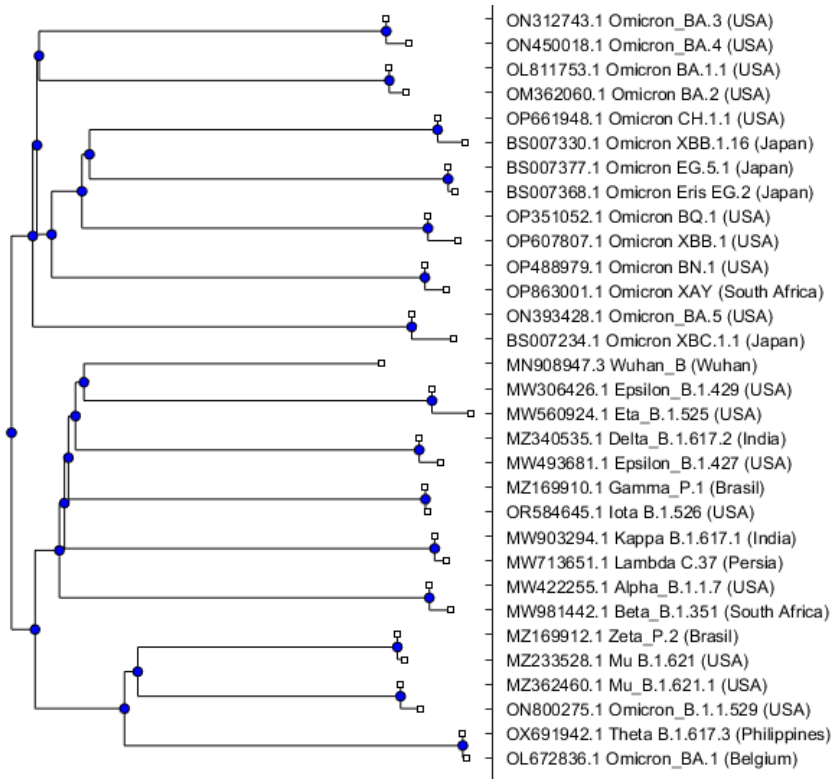

Figure 2. Phylogenetic tree of the SARS-CoV-2 virus

3.3. Conserved sequence

The sequences were aligned using the progressive method based on the phylogenetic tree shown in Figure 2, resulting in the aligned sequence $S' = \{S'_1, S'_2, S'3, \cdots, S'_{31}\}$ shown in Figure 3, with a sequence length of 29,929 bp. In Figure 3, the conserved sequence is shown in the same color from S'_1 to S'_{31}, but we wanted to know the position of the longest unbroken conserved sequence; therefore, an LLM was used to determine its position. Before using the LLM, the sequences were first converted into binary sequences $B_k = \{b_{k,1}, b_{k,2}, \cdots, b_{k,m}\}$, $k = 1,2, \cdots, C_2^n$. The binary sequence B_k was created by comparing the elements of the two sequences using Boolean operations. Then, by using an LLM in the latticization stage, a lattice with length C_2^n is obtained such that, in the shadow clustering process, looking for the minimum complexity of the lattice, a lattice with minimum complexity is obtained, which only contains the number one as a member of the sequence.

Therefore, if this is change to an intelligible rule, the rule obtained is: if $b_{1,l} \geq 1$ AND $b_{2,l} \geq 1$ AND $b_{3,l} \geq 1$ AND $\cdots$ AND $b_{C_2^n,l} \geq 1$ then $y_l = 1$, at this stage, the sequence is expressed in the form $Y = \{y_1, y_2, \cdots, y_i, \cdots, y_l\}$ with $y_i = \{0,1\}$ and $Y'_p \subset Y$ denotes the value of 1 in positions i to $i + s$. The maximum value of s indicates that Y'_p was the position of the conserved sequence. The position of the conserved sequence, based on the longest sequence, is shown in Figure 4. Figure 4 shows the conserved sequence in green, which represents the longest continuous conserved sequence with positions $i = 13{,}861$ to $i + s = 14{,}256$ in sequence S'_i with length $s = 395$. The conserved sequences of SARS-CoV-2 are "ACATTAAAAGAAATACTTGTCACATACAATTGTTGTGATGATGATTATTTCAATAAAAAGGACT
GGTATGATTTTGTAGAAAACCCAGATATATTACGCGTATACGCCAACTTAGGTGAACGTGTACG
CCAAGCTTTGTTAAAAACAGTACAATTCTGTGATGCCATGCGAAATGCTGGTATTGTTGGTGTAC
TGACATTAGATAATCAAGATCTCAATGGTAACTGGTATGATTTCGGTGATTTCATACAAACCACG
CCAGGTAGTGGAGTTCCTGTTGTAGATTCTTATTATTCATTGTTAATGCCTATATTAACCTTGACC
AGGGCTTTAACTGCAGAGTCACATGTTGACACTGACTTAACAAAGCCTTACATTAAGTGGGATTT
GTTAAA." Next, through transcription and translation, which convert the nucleotide sequence of DNA into an amino acid sequence, the conserved protein sequences of SARS-CoV-2 are "TLKEILVTYNCCDDDYFNKKDWYDFVENPDILRVYANLGERVRQALLKTVQFCDAMRNAGIVGVL
TLDNQDLNGNWYDFGDFIQTTPGSGVPVVDSYYSLLMPILTLTRALTAESHVDTDLTKPYIKWDLL."

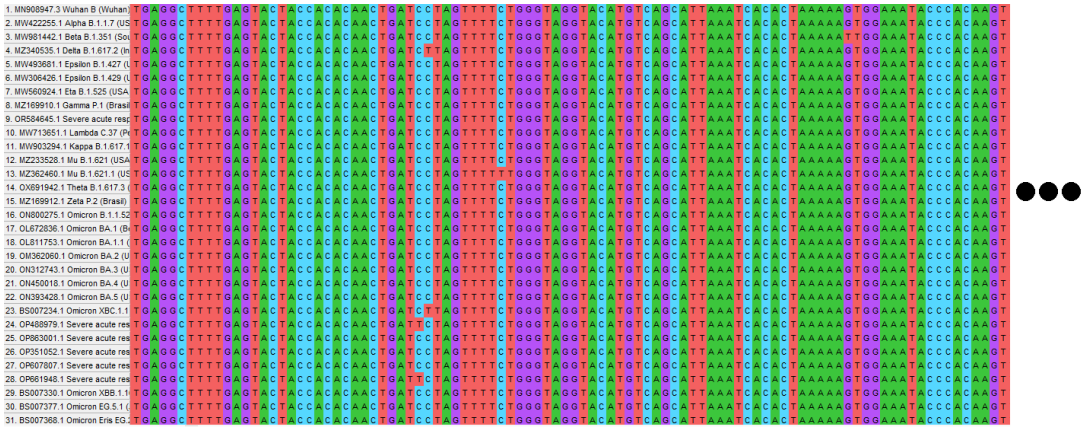

Figure 3. Alignment result of SARS-CoV-2 sequences

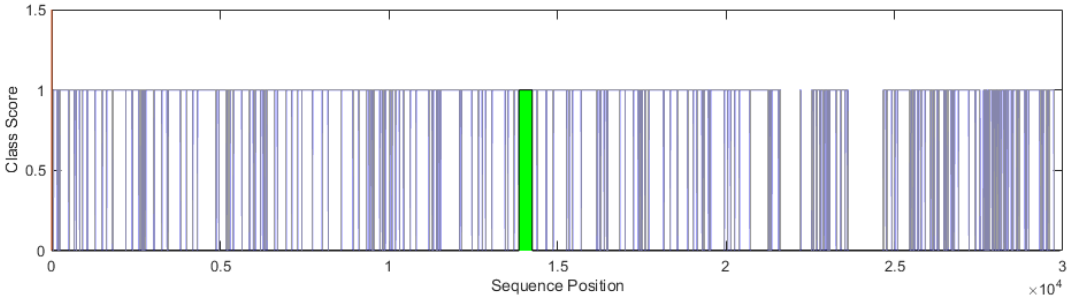

Figure 4. Conserved sequence of the SARS-CoV-2 virus

The conserved protein sequences of SARS-CoV-2 shows the 182-amino-acid conserved protein sequence of SARS-CoV-2. From the conserved protein sequences obtained, their presence in the complete SARS-CoV-2 protein sequence was confirmed using basic local alignment search tool (BLAST). BLAST on the UniProt website was used to test whether the conserved protein sequences found are present in coronaviruses, and the results are shown in Figure 5. Figure 5 shows that the conserved protein sequence obtained was indeed found in the SARS-CoV-2 section, as shown by the colored points on the right of the image, which show the position of the conserved sequence obtained and its position in several coronavirus protein sequences.

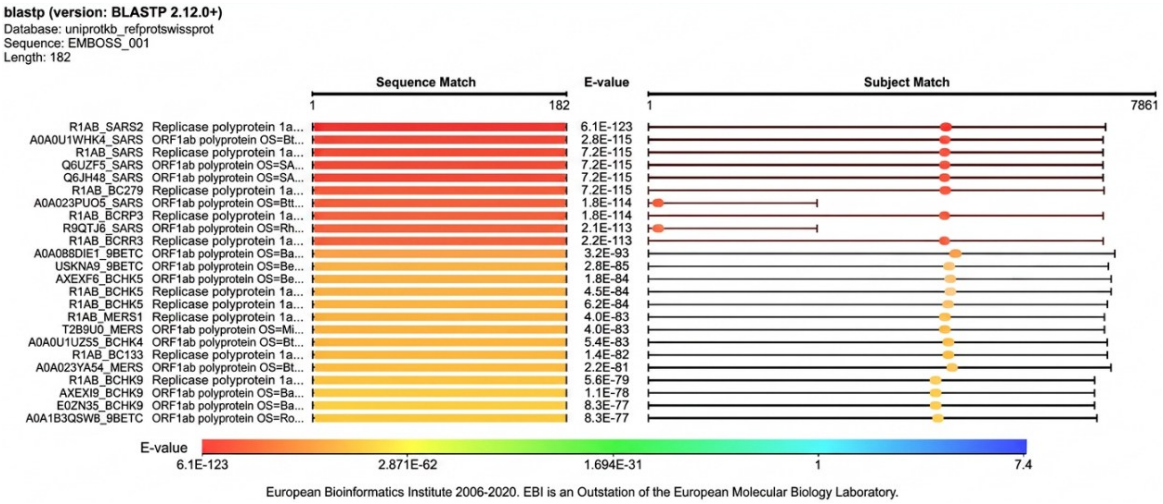

Figure 5. Position of conserved protein

3.4. Vaccine candidate for SARS-CoV-2

BepiPred-2.0, a web server for predicting B cell epitopes from antigen sequences, was employed to analyze the obtained protein sequences. Epitope scores for each amino acid residue in the protein were computed using BepiPred 2.0 [43]. These scores are depicted in Figure 6.

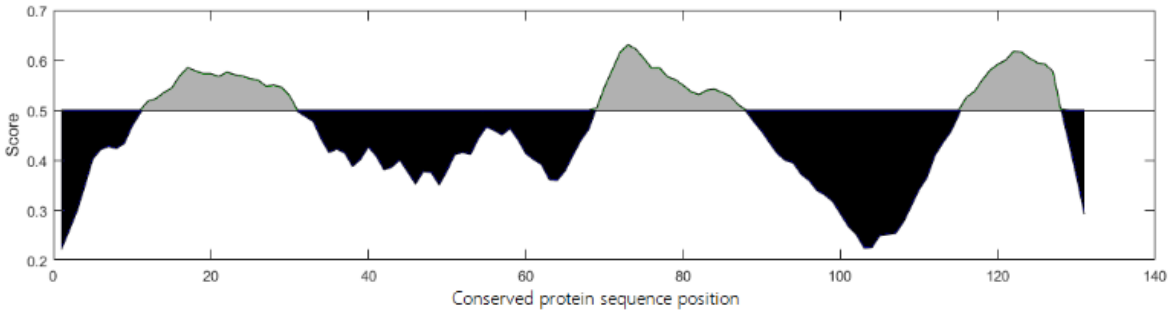

Figure 6. Epitope score of the SARS-CoV-2 conserved protein sequence

Figure 6 shows the score of each residue; the gray graph indicates that the residue score is ≥ 0.5, and the black graph indicates that the residue score is < 0.5. This indicates that the residue's position on the gray graph is an epitope, and the sequence at the unbroken epitope position is called a peptide. From the obtained epitope values, peptides expressing unbroken epitope sequences and their immunogenicity values were determined using VaxiJen v2.0 [44]. The scores for each peptide are listed in Table 4. Ptd3 cannot be used as a vaccine candidate because it has a peptide score < 0.4; however, Ptd1 and Ptd2 can be used as vaccine candidates because they have peptide scores > 0.4.

Table 4. Peptide score

Start	End	Peptide	Immunogenicity	Code
12	30	CDDDYFNKKDW YDFVENPD	0.4053 Immunogen	Ptd1
69	87	NQDLNGNWYDF GDFIQTTPD	0.7213 Immunogen	Ptd2
116	128	HVDTDLTKPYI KW	-0.0612 Non-Immunogen	Ptd3

The results were validated using web-based applications, enabling further analysis to determine whether the peptides could be accepted as in silico vaccine candidates. The applications used include AllerTOP v2.0 (https://www.ddgpharmfac.net/AllerTOP/) for predicting peptide allergenicity and ToxinPred (https://webs.iiitd.edu.in/raghava/-toxinpred/design.php/) for predicting peptide toxicity. Peptides are considered acceptable if they are immunogens, non-toxins, and non-allergens. The prediction results obtained using AllerTOP and ToxinPred applications are listed in Table 5. Table 5 shows that of the four peptides previously obtained, only one, namely Ptd2, is accepted as an in silico vaccine candidate because Ptd2 is an immunogen, non-allergen, and non-toxic.

Table 5. Vaccine candidates status

Peptides	Allergenicity	Toxicity	Status
Ptd1	Allergen	Non-toxin	Rejected
Ptd2	Non-allergen	Non-toxin	Accepted
Ptd3	Allergen	Non-toxin	Rejected

4. CONCLUSION

An experiment on a hybrid algorithm that uses an LLM and progressive alignment was presented, from which several conclusions were drawn. First, the proposed algorithm demonstrated its capability to identify conserved sequences. Employing this algorithm revealed that numerous nucleotide mutations, particularly transitions and transversions, were expressed as mutation percentages, indicating that the mutation rate within the sequence did not exceed 1.5%. This mutation percentage suggests homology to SARS-CoV-2 of approximately ±98.5%, with the highest mutation percentage in the theta B.1.617.3 and Omicron BA.1 variants at 1.4578%, and the lowest in Omicron BA.1 and Omicron BA.1.1 at 0.0067%.

Moreover, the proposed algorithm demonstrated its efficacy in determining proteins suitable for vaccine candidate analysis. Experimental validation using SARS-CoV-2 confirmed its ability to identify proteins for candidate analysis. Through in silico immuno-bioinformatics analysis, conserved protein sequences were used to extract epitopes, yielding four peptides in SARS-CoV-2 experiments, one of which emerged as a promising vaccine candidate.

ACKNOWLEDGMENTS

The authors would like to express gratitude to the Institut Teknologi Sepuluh Nopember for their support.

FUNDING INFORMATION

This work funded by Institut Teknologi Sepuluh Nopember research grant number 1723/PKS/ITS/2023.

AUTHOR CONTRIBUTIONS STATEMENT

This journal uses the Contributor Roles Taxonomy (CRediT) to recognize individual author contributions, reduce authorship disputes, and facilitate collaboration.

Name of Author	C	M	So	Va	Fo	I	R	D	O	E	Vi	Su	P	Fu
Felza Ridho	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓		✓	
Mohammad Isa Irawan	✓	✓			✓	✓	✓	✓		✓		✓	✓	✓
Nurul Hidayat					✓	✓		✓		✓	✓			
Awik Puji Dyah Nurhayati				✓		✓				✓	✓			

C : **C**onceptualization
M : **M**ethodology
So : **So**ftware
Va : **Va**lidation
Fo : **Fo**rmal analysis
I : **I**nvestigation
R : **R**esources
D : **D**ata Curation
O : Writing - **O**riginal Draft
E : Writing - Review & **E**diting
Vi : **Vi**sualization
Su : **Su**pervision
P : **P**roject administration
Fu : **Fu**nding acquisition

CONFLICT OF INTEREST STATEMENT

Authors state no conflict of interest.

DATA AVAILABILITY

The data that support the findings of this study are openly available in NCBI at https://www.ncbi.nlm.nih.gov/.

REFERENCES

[1] A. Kumar *et al.*, "Wuhan to world: the COVID-19 pandemic," *Frontiers in Cellular and Infection Microbiology*, vol. 11, Mar. 2021, doi: 10.3389/fcimb.2021.596201.
[2] World Health Organization, "Mpox (monkeypox)," *who.int*, 2023, Accessed: Jul. 13, 2023. [Online]. Available: https://www.who.int/emergency/situations/monkeypox-oubreak-2022
[3] J. S. Heitmann *et al.*, "A COVID-19 peptide vaccine for the induction of SARS-CoV-2 T cell immunity," *Nature*, vol. 601, no. 7894, pp. 617–622, Jan. 2022, doi: 10.1038/s41586-021-04232-5.
[4] L. A. Grills and A. L. Wagner, "The impact of the COVID-19 pandemic on parental vaccine hesitancy: a cross-sectional survey," *Vaccine*, vol. 41, no. 41, pp. 6127–6133, Sep. 2023, doi: 10.1016/j.vaccine.2023.08.044.
[5] S. S. Koshal *et al.*, "Critical factors in the successful expansion of pneumococcal conjugate vaccine in India during the COVID-19 pandemic," *Vaccine: X*, vol. 14, Aug. 2023, doi: 10.1016/j.jvacx.2023.100328.
[6] J. V. L. C. Novaes, F. M. D. F. Faria, B. S. C. de Bragança, and L. I. dos Santos, "Impacts of the COVID-19 pandemic on immunization with pneumococcal vaccines in children and older adults in Brazil," *Preventive Medicine*, vol. 173, Aug. 2023, doi: 10.1016/j.ypmed.2023.107602.
[7] S. H. Alhashemi, F. Ahmadi, and A. Dehshahri, "Lessons learned from COVID-19 pandemic: vaccine platform is a key player," *Process Biochemistry*, vol. 124, pp. 269–279, Jan. 2023, doi: 10.1016/j.procbio.2022.12.002.
[8] S. Saha, S. Vashishtha, B. Kundu, and M. Ghosh, "In-silico design of an immunoinformatics-based multi-epitope vaccine against Leishmania donovani," *BMC Bioinformatics*, vol. 23, no. 1, Aug. 2022, doi: 10.1186/s12859-022-04816-6.
[9] L. Lu, W. Ma, C. H. Johnson, S. A. Khan, M. L. Irwin, and L. Pusztai, "In silico designed mRNA vaccines targeting CA-125 neoantigen in breast and ovarian cancer," *Vaccine*, vol. 41, no. 12, pp. 2073–2083, Mar. 2023, doi: 10.1016/j.vaccine.2023.02.048.

[10] M. Shafaghi, Z. Bahadori, H. Madanchi, M. M. Ranjbar, A. A. Shabani, and S. F. Mousavi, "Immunoinformatics-aided design of a new multi-epitope vaccine adjuvanted with domain 4 of pneumolysin against Streptococcus pneumoniae strains," *BMC Bioinformatics*, vol. 24, no. 1, Feb. 2023, doi: 10.1186/s12859-023-05175-6.
[11] S. Heidari *et al.*, "Identification of immunodominant proteins of *Leishmania infantum* by immunoproteomics to evaluate a recombinant multi-epitope designed antigen for serodiagnosis of human visceral leishmaniasis," *Experimental Parasitology*, vol. 222, Mar. 2021, doi: 10.1016/j.exppara.2021.108065.
[12] N. Bano and A. Kumar, "Immunoinformatics study to explore dengue (DENV-1) proteome to design multi-epitope vaccine construct by using CD4+ epitopes," *Journal of Genetic Engineering and Biotechnology*, vol. 21, no. 1, Dec. 2023, doi: 10.1186/s43141-023-00592-9.
[13] L. Santus, E. Garriga, S. Deorowicz, A. Gudyś, and C. Notredame, "Towards the accurate alignment of over a million protein sequences: current state of the art," *Current Opinion in Structural Biology*, vol. 80, Jun. 2023, doi: 10.1016/j.sbi.2023.102577.
[14] C. Zheng, X. Chen, T. Zhang, N. V. Sahinidis, and J. J. Siirola, "Learning process patterns via multiple sequence alignment," *Computers & Chemical Engineering*, vol. 159, Mar. 2022, doi: 10.1016/j.compchemeng.2022.107676.
[15] Y. Zhang, Q. Zhang, Y. Liu, M. Lin, and C. Ding, "Multiple sequence alignment based on deep Q network with negative feedback policy," *Computational Biology and Chemistry*, vol. 101, Dec. 2022, doi: 10.1016/j.compbiolchem.2022.107780.
[16] A. Caucchiolo and F. Cicalese, "Hardness and approximation of multiple sequence alignment with column score," *Theoretical Computer Science*, vol. 946, Feb. 2023, doi: 10.1016/j.tcs.2022.12.033.
[17] Y. S. Bukin, A. N. Bondaryuk, N. V. Kulakova, S. V. Balakhonov, Y. P. Dzhioev, and V. I. Zlobin, "Phylogenetic reconstruction of the initial stages of the spread of the SARS-CoV-2 virus in the Eurasian and American continents by analyzing genomic data," *Virus Research*, vol. 305, Nov. 2021, doi: 10.1016/j.virusres.2021.198551.
[18] Y. Pan *et al.*, "Characterisation of SARS-CoV-2 variants in Beijing during 2022: an epidemiological and phylogenetic analysis," *The Lancet*, vol. 401, pp. 664–672, Feb. 2023, doi: 10.1016/S0140-6736(23)00129-0.
[19] M. B. Sassi *et al.*, "Phylogenetic and amino acid signature analysis of the SARS-CoV-2 lineages circulating in Tunisia," *Infection, Genetics and Evolution*, vol. 102, Aug. 2022, doi: 10.1016/j.meegid.2022.105300.
[20] Y.-C. Huang *et al.*, "Outbreak investigation in a COVID-19 designated hospital: The combination of phylogenetic analysis and field epidemiology study suggesting airborne transmission," *Journal of Microbiology, Immunology and Infection*, vol. 56, no. 3, pp. 547–557, Jun. 2023, doi: 10.1016/j.jmii.2023.01.003.
[21] H. N. Kim, S.-Y. Yoon, C. S. Lim, C. K. Lee, and J. Yoon, "Phylogenetic analysis of human parainfluenza type 3 virus strains responsible for the outbreak during the COVID-19 pandemic in Seoul, South Korea," *Journal of Clinical Virology*, vol. 153, Aug. 2022, doi: 10.1016/j.jcv.2022.105213.
[22] M. Partipilo, G. Yang, M. L. Mascotti, H. J. Wijma, D. J. Slotboom, and M. W. Fraaije, "A conserved sequence motif in the *Escherichia coli* soluble FAD-containing pyridine nucleotide transhydrogenase is important for reaction efficiency," *Journal of Biological Chemistry*, vol. 298, no. 9, Sep. 2022, doi: 10.1016/j.jbc.2022.102304.
[23] M. V.-Ozdeniz, A. Akturk, M. Demirdizen, R. Leka, R. Acar, and O. Konu, "CoVrimer: a tool for aligning SARS-CoV-2 primer sequences and selection of conserved/degenerate primers," *Genomics*, vol. 113, no. 5, pp. 3174–3184, Sep. 2021, doi: 10.1016/j.ygeno.2021.07.020.
[24] D. Selvaraj, S. P. Mugunthan, H. M. Chandra, C. Govindasamy, K. S. Al-Numair, and R. Balasubramani, "In silico identification of novel and conserved microRNAs and targets in peppermint (*Mentha piperita*) using expressed sequence tags (ESTs)," *Journal of King Saud University - Science*, vol. 35, no. 4, May 2023, doi: 10.1016/j.jksus.2023.102604.
[25] A. Sadeque, M. Barsky, F. Marass, P. Kruczkiewicz, and C. Upton, "JaPaFi: a novel program for the identification of highly conserved DNA sequences," *Viruses*, vol. 2, no. 9, pp. 1867–1885, Aug. 2010, doi: 10.3390/v2091867.
[26] A. A. Hemedan, A. Niarakis, R. Schneider, and M. Ostaszewski, "Boolean modelling as a logic-based dynamic approach in systems medicine," *Computational and Structural Biotechnology Journal*, vol. 20, pp. 3161–3172, Jan. 2022, doi: 10.1016/j.csbj.2022.06.035.
[27] S. F. Peixoto, A. M. C. Horbe, T. M. Soares, C. A. Freitas, E. M. D. de Sousa, and E. R. H. de F. Iza, "Boolean and fuzzy logic operators and multivariate linear regression applied to airborne gamma-ray spectrometry data for regolith mapping in granite-greenstone terrain in Midwest Brazil," *Journal of South American Earth Sciences*, vol. 112, Dec. 2021, doi: 10.1016/j.jsames.2021.103562.
[28] R. Barbuti, R. Gori, and P. Milazzo, "Encoding Boolean networks into reaction systems for investigating causal dependencies in gene regulation," *Theoretical Computer Science*, vol. 881, pp. 3–24, Aug. 2021, doi: 10.1016/j.tcs.2020.07.031.
[29] M. Muselli, "Switching neural networks: a new connectionist model for classification," in *Neural Nets*, Berlin, Heidelberg: Springer, 2006, pp. 23–30, doi: 10.1007/11731177_4.
[30] A. Gerussi *et al.*, "LLM-PBC: logic learning machine-based explainable rules accurately stratify the genetic risk of primary biliary cholangitis," *Journal of Personalized Medicine*, vol. 12, no. 10, Sep. 2022, doi: 10.3390/jpm12101587.
[31] M. Mongelli, "Design of countermeasure to packet falsification in vehicle platooning by explainable artificial intelligence," *Computer Communications*, vol. 179, pp. 166–174, Nov. 2021, doi: 10.1016/j.comcom.2021.06.026.
[32] D. Verda, S. Parodi, E. Ferrari, and M. Muselli, "Analyzing gene expression data for pediatric and adult cancer diagnosis using logic learning machine and standard supervised methods," *BMC Bioinformatics*, vol. 20, no. S9, Nov. 2019, doi: 10.1186/s12859-019-2953-8.
[33] B. Çakır, B. Okuyan, G. Şener, and T. T.-Akbay, "Investigation of beta-lactoglobulin derived bioactive peptides against SARS-CoV-2 (COVID-19): in silico analysis," *European Journal of Pharmacology*, vol. 891, Jan. 2021, doi: 10.1016/j.ejphar.2020.173781.
[34] C. Chakraborty, A. R. Sharma, M. Bhattacharya, G. Sharma, and S.-S. Lee, "Immunoinformatics approach for the identification and characterization of T cell and B cell epitopes towards the peptide-based vaccine against SARS-CoV-2," *Archives of Medical Research*, vol. 52, no. 4, pp. 362–370, May 2021, doi: 10.1016/j.arcmed.2021.01.004.
[35] J. Tobias *et al.*, "Active immunization with a Her-2/neu-targeting multi-peptide B cell vaccine prevents lung metastases formation from Her-2/neu breast cancer in a mouse model," *Translational Oncology*, vol. 19, May 2022, doi: 10.1016/j.tranon.2022.101378.
[36] M. S. Hossen *et al.*, "Immunoinformatics-aided rational design of multiepitope-based peptide vaccine (MEBV) targeting human parainfluenza virus 3 (HPIV-3) stable proteins," *Journal of Genetic Engineering and Biotechnology*, vol. 21, no. 1, Dec. 2023, doi: 10.1186/s43141-023-00623-5.
[37] M. Čavojský, M. Drozda, and Z. Balogh, "Analysis and experimental evaluation of the Needleman-Wunsch algorithm for trajectory comparison," *Expert Systems with Applications*, vol. 165, Mar. 2021, doi: 10.1016/j.eswa.2020.114068.
[38] B. T. Thomas, L. A. Ogunkanmi, B. A. Iwalokun, and O. D. Popoola, "Transition-transversion mutations in the polyketide synthase gene of *Aspergillus* section *Nigri*," *Heliyon*, vol. 5, no. 6, Jun. 2019, doi: 10.1016/j.heliyon.2019.e01881.

[39] A. Isaev, *Introduction to mathematical methods in bioinformatics*, 1st ed. Berlin, Heidelberg: Springer, 2006, doi: 10.1007/978-3-540-48426-4.
[40] D. A. Simovici and C. Djeraba, *Mathematical tools for data mining*, 2nd ed. London: Springer London, 2014, doi: 10.1007/978-1-4471-6407-4.
[41] M. Muselli and E. Ferrari, "Coupling logical analysis of data and shadow clustering for partially defined positive Boolean function reconstruction," *IEEE Transactions on Knowledge and Data Engineering*, vol. 23, no. 1, pp. 37–50, Jan. 2011, doi: 10.1109/TKDE.2009.206.
[42] W. C. Summers, "Central dogma of molecular biology," in *Reference Module in Life Sciences*, Elsevier, 2022, doi: 10.1016/B978-0-12-822563-9.00002-0.
[43] M. C. Jespersen, B. Peters, M. Nielsen, and P. Marcatili, "BepiPred-2.0: improving sequence-based B-cell epitope prediction using conformational epitopes," *Nucleic Acids Research*, vol. 45, no. W1, pp. W24–W29, Jul. 2017, doi: 10.1093/nar/gkx346.
[44] I. A. Doytchinova and D. R. Flower, "VaxiJen: a server for prediction of protective antigens, tumour antigens and subunit vaccines," *BMC Bioinformatics*, vol. 8, no. 1, Dec. 2007, doi: 10.1186/1471-2105-8-4.

BIOGRAPHIES OF AUTHORS

Felza Ridho received a bachelor's degree in Mathematics from the University of Bengkulu, Indonesia, in 2020 and a master's degree in Mathematics from Institut Teknologi Sepuluh Nopember, Indonesia, in 2023. Currently, he is a student at the Department of Mathematics, Institut Teknologi Sepuluh Nopember, Surabaya, Indonesia. His research interests include computational mathematics, computer science, data science, and bioinformatics. He can be contacted at email: 7002231003@student.its.ac.id.

Mohammad Isa Irawan received bachelor degree in Applied Mathematics from Universitas Airlangga, Indonesia (1989), and Master of Electrical Engineering from Institut Teknologi Bandung, Indonesia (1994) in field of Control System Engineering. His Ph.D. in Computer Science from Vienna University of Technology, Austria (1999). He is currently lecturing in the Department of Mathematics at Institut Teknologi Sepuluh Nopember, Indonesia. He wrote introduction to machine learning, lecture notes (2022) and bioinformatics, lecture notes (2019) both in Indonesian language. His interest researches are bioinformatics and artificial intelligence for health. He can be contacted at email: mii@its.ac.id.

Nurul Hidayat holds a bachelor degree in Mathematics from Institut Teknologi Sepuluh Nopember, Indonesia, in 1987, and Master of Computer Science from University of Indonesia, Indonesia, in 1999. He is currently lecturing in the Department of Mathematics at Institut Teknologi Sepuluh Nopember, Surabaya, East Java, Indonesia. He is a member of Indonesian Mathematical Society (INDOMS). His research interests include data science, computer vision, computational algorithm, and biometrics. He can be contacted at email: nurul_hdy@yahoo.com.

Awik Puji Dyah Nurhayati holds a bachelor degree in Biology from Universitas Gadjah Mada, Indonesia, in 1994, Master of Science in Biology from Universitas Gadjah Mada, Indonesia, in 2003, and doctor in Biology from Universitas Gadjah Mada, Indonesia, in 2015. She is currently lecturing in the Department of Biology at Institut Teknologi Sepuluh Nopember, Surabaya, East Java, Indonesia. She wrote Animal development (2020), Animal structure (2019), and Marine natural products (2019) in the Indonesian language. Her research interest include immunology, anticancer, and natural products. She can be contacted at email: awiknurhayati@gmail.com.