Open Access Article

Identification of the Spread of the Influenza Virus Type A/H9N2 in Indonesia Using the Neighbor-Joining Algorithm with Felsenstein Models

Siti Amiroch[1,2], Mohammad Isa Irawan[1,*], Imam Mukhlash[1], Arif Nur Muhammad Ansori[3], Chairul A. Nidom[3,4,*]

[1] Department of Mathematics, Faculty of Science and Data Analytics, Institut Teknologi Sepuluh Nopember, Surabaya, Indonesia

[2] Department of Mathematics, Faculty of Mathematics and Natural Sciences, Universitas Islam Darul ‘Ulum, Lamongan, Indonesia

[3] Professor Nidom Foundation, Surabaya, Indonesia

[4] Faculty of Veterinary Medicine, Universitas Airlangga, Surabaya, Indonesia

Abstract: The H9N2 virus is an avian influenza disease that does not harm humans but causes respiratory infections and replicates in the tract reproduction in chickens, reducing egg production by 50% to 80%. This case is related to a bioinformatics review from a mathematical perspective, namely, analyzing the H9N2 virus spread using phylogenetic trees with other methods. The phylogenetic tree is a visualization of multiple alignment analysis obtained from the evolution analysis between each sequence pair based on the pairwise alignment method. As a way to identify the spread of the H9N2 virus, the formation of this phylogenetic tree uses pairwise alignment with the Needleman-Wunsch algorithm as the first step in processing the multiple alignments of the H9N2 virus. In this study, phylogenetic trees were constructed using a distance-based method, namely the neighbor-joining algorithm with the Felsenstein model, and simulated in MATLAB®. With this method, we can arrange the phylogenetic tree views with different colors according to the cluster and each branch's spacing shown in more detail. The pairwise alignment results also show mutations from the initial sequence to the entire series in the first cluster.

Keywords: avian influenza, H9N2, phylogenetics tree, neighbor-joining algorithm, Felsenstein models.

使用费尔森施泰因模型的邻居加入算法识别印度尼西亚一种 /我们的权利型流感病毒的传播

摘要：我们的权利病毒是一种禽流感疾病，不会伤害人类，但会引起呼吸道感染，并在鸡的生殖道中复制，从而使产蛋量减少 50％至 80％。这种情况与从数学角度对生物信息学的回顾有关，即使用系统进化树和其他方法分析我们的权利病毒的传播。系统发育树是从多序列比对分析中可视化的结果，该比对分析是基于成对比对方法从每个序列对之间的进化分析中获得的。作为鉴定我们的权利病毒传播的一种方法，该系统发育树的形成使用针线愿望算法的成对比对作为处理我们的权利病毒多重比对的第一步。在这项研究中，使用基于距离的方法（即具有费尔森施泰因模型的邻居连接算法）构建了系统树，并在 MATLAB®中进行了仿真。通过这种方法，我们可以根据群集和每个分支的间距更详细地显示不同颜色的系统树视图。成对的比对结果还显示在第一簇中从初始序列到整个序列的突变。

Received: February 8, 2021 / Revised: March 10, 2021 / Accepted: April 15, 2021 / Published: May 28, 2021

About the authors: Siti Amiroch, Department of Mathematics, Faculty of Science and Data Analytics, Institut Teknologi Sepuluh Nopember, Surabaya, Indonesia; Department of Mathematics, Faculty of Mathematics and Natural Sciences, Universitas Islam Darul ‘Ulum, Lamongan, Indonesia; Mohammad Isa Irawan, Imam Mukhlash, Department of Mathematics, Faculty of Science and Data Analytics, Institut Teknologi Sepuluh Nopember, Surabaya, Indonesia; Arif Nur Muhammad Ansori, Professor Nidom Foundation, Surabaya, Indonesia; Chairul A. Nidom, Professor Nidom Foundation, Surabaya, Indonesia; Faculty of Veterinary Medicine, Universitas Airlangga, Surabaya, Indonesia

Corresponding author Mohammad Isa Irawan, mii@its.ac.id; Chairul A. Nidom, nidomca@fkh.unair.ac.id

关键词：禽流感，我们的权利，系统树，邻居加入算法，费尔森施泰因模型。

1. Introduction

Influenza A virus, a new type of avian influenza virus in Indonesia from subtype H9N2, was first discovered in South Sulawesi in December 2016 [1]. Since the virus is low pathogenic avian influenza or avian influenza, it does not harm humans [2-7]. However, the H9N2 virus causes respiratory tract infections and replicates in the reproductive tract in chickens, reducing egg production by 50% to 80% [8, 9]. In Indonesia, the H9N2 virus was identified in chicken farms from December 2016 to May 2017 by reverse transcriptase PCR. The method used to identify the phylogeny was the Neighbor-Joining method in the MEGA 6 package [10, 11].

Using the pairwise alignment method to process multiple alignments is very important in bioinformatics as a rough analysis of multiple sequences' general structure. In this case, the study of hereditary relationships and evolution between each pair of sequences can be done based on the pairwise alignment method, suggesting that pairwise alignment for multiple sequences is essential to consider [12, 13]. The formation of a phylogenetic tree to identify the spread of a virus also uses pairwise alignment as the first step in processing the virus's multiple alignments, namely the H9N2 virus. The dynamic programming method used in pairwise alignment and the Needleman-Wunsch algorithm as a global alignment algorithm [14], and the Smith-Waterman algorithm as a local alignment algorithm [15], have a significant role in the alignment process.

Meanwhile, the phylogenetic tree can be formed in various methods, one of which is the distance-based method using the neighbor-joining algorithm [16]. The distance in question is a genetic distance, which is the difference between two alignment sequences (mutations). Since mutation is a molecular event, the matrix obtained from genetic distances is converted into an evolutionary distance matrix using the Felsenstein model. The Felsenstein model is a development of the Jukes-Cantor model [17]. We can say that the Jukes-Cantor model is a particular case of the Felsenstein Model because, in Jukes-Cantor, the probability of each element is considered has the same value. In contrast, Felsenstein has a different matter.

Previous studies have analyzed the phylogenetic tree using readily available software packages. For example, the study of H9N2 cases in China, Korea, and the Middle East analyzes phylogenetic trees globally using the MEGA 5.2 package [18]. Analysis of phylogenetic influenza A viruses of the H9 Hemagglutinin subtype [19] from gulls was also analyzed using the PHYLIP 3.57 package. Likewise, the study conducted by [20] uses the MEGA 6 package to analyze phylogenetic cyanobacterial glutaredoxins. [20]. However, in this study, the phylogenetic tree used to identify the spread of the H9N2 viruses in Indonesia was constructed using a neighbor-joining algorithm with the Felsenstein models simulated in MATLAB® and analyzed mathematically. The data used were 75 H9N2 DNA sequences from various provinces in Indonesia in 2016-2018.

2. Methods

To facilitate more understanding of the methodology that we used, the following presented in a framework: (1) The data of 75 DNA sequences revealed from NCBI, (2) Pairwise alignment process: dynamic programming algorithm using Needleman-Wunsch algorithm, (3) Distance matrix, (4) Evolutionary models: Felsenstein, (5) Construction of phylogenetic tree using the neighbor-joining algorithm, (6) Identification of the spread of H9N2 virus in chickens in Indonesia using a phylogenetic tree based on the neighbor-joining algorithm with Felsenstein models, and (7) Multiple alignment analysis.

2.1. Data Retrieval

Data used for this research were from NCBI [21] on April 13, 2020, in the form of 72 sequences from HA protein Subtype H9 with GenBank® access code MK559797 (Sequence 1) to MK559868 (sequence 72) from several provinces in Indonesia. This research also utilized three sequence data from South Sulawesi, Indonesia, and Central Sulawesi, Indonesia, in 2016 (which only appeared in GenBank early September 2020). The three sequence data are GenBank® access code MF164879 to MF164881, HA protein subtype H9.

2.2. Pairwise Alignment

The pairwise alignment method used is the Needleman-Wunsch method. The Needleman-Wunsch algorithm is a global alignment algorithm for paired sequences [12]. The steps in this algorithm are as follows:

If given sequence $A = (a_1, a_2, \dots, a_n)$, and $B = (b_1, b_2, \dots, b_n)$. Table 1 shows the two sequences with $s(i, j)$ obtained from the next step.

Table 1 Two-dimensional table of Sequence A, B

		a_1	a_2	…..	a_n
	s(0,0)	s(1,0)	s(2,0)	…..	s(n,0)
b_1	s(0,1)	s(1,1)	s(2,1)	…..	s(n,1)
b_2	s(0,2)	s(1,2)	s(2,2)	…..	s(n,2)

…	…..	…..	…..	…..	…..
b_m	s(0,m)	s(1,m)	s(2,m)	…..	s(n,m)

For the elements $s(i,j)$, it can be calculated using the following formula:

$s(i,j) = \max\{(i-1,j-1) + s(a_i, b_j),\ s(i1,j) - d, s(i,j-1) - d\}$.

Then, we obtain the alignment of the sequence as follows: (1) Denote the pair of the sequence as a_i, b_i if the reverse-path starts from a_i, b_i to the top left corner, (2) Insert a virtual symbol in a vertical sequence and denote it as $(a_i, -)$ if the reverse path is horizontal, (3) Insert a virtual symbol in a horizontal sequence and denote it as $(-, a_i)$ if the reverse path is vertical and (4) Finally, the two sequences' optimal alignment is obtained.

2.3. Evolutionary Models

The result of pairwise alignment is a distance matrix that is then converted into an evolutionary distance matrix using the Felsenstein model. The Felsenstein model is a development of the Jukes-Cantor model [17]. In other words, the Jukes-Cantor model is a particular case of the Felsenstein Model because, in Jukes-Cantor, the probability of each element is considered the same. Given the matrix mean:

$$Q = \begin{pmatrix} * & \pi_G & \pi_C\pi_T \\ \pi_A & * & \pi_C\pi_T \\ \pi_A & \pi_G & *\ \pi_T \\ \pi_A & \pi_G & \pi_C\ * \end{pmatrix}$$

The existing formula shows that:

$$\beta = \frac{1}{(1 - {\pi_A}^2 - {\pi_C}^2 - {\pi_G}^2 - {\pi_T}^2)} \quad (1)$$

$$P_{ij}(v) = \begin{cases} e^{-\beta v} + \pi_j\left(1 - e^{-\beta v}\right) & if\ i = j \\ \pi_j\left(1 - e^{-\beta v}\right) & if\ i \neq j \end{cases} \quad (2)$$

Note:

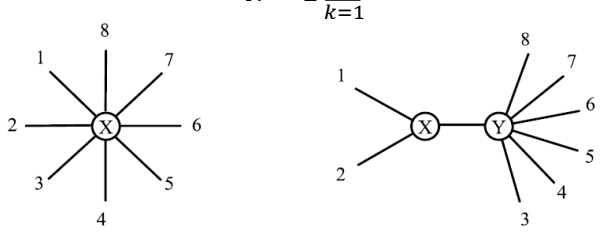
v is the branch's length, calculated from the expected per-site changes $P_{ij}(v)$ where $P_{ij}(v)$ is the probability of branch length i to j. The branch length v calculated using the derivative of the formula described in the Results section.

2.4. Neighbor-Joining Algorithm

The Neighbor-Joining method is a distance-based method used to build phylogenetic trees [14]. This algorithm requires input in a distance matrix obtained by aligning each sequence to find its similarity. In contrast, distance is the dissimilarity of the sequence in question. The Neighbor-Joining method starts from a star-like structure and gathers all the "neighbors" together to form a tree without roots as output. For the set of N sequences, the computational steps are as follows:

1. Determine the N sequence distance matrix.
2. Assume a tree with all OTUs in the matrix as branches from the center. After that, create a star-like pattern as in the schematic representation in Fig. 1a.
3. For each OTU, calculate S, where S is the sum of the distance (D) between one OTU and another, divided by (N-2), where N is the total number of OTUs.

$$S_i = \frac{1}{N-2}\sum_{k=1}^{N} D_{ik}$$



(a) (b)

Fig. 1 Neighbor-joining: a. Starting structure star-like; b. Tree-like structure

3. Results and Discussion

3.1. Sequence Alignment

The alignment between 2 sequences simultaneously starts from sequence-1 with sequence-2, sequence-1 with sequence-3, etc., until sequence-75. The alignment results are genetic differences or genetic distances between sequences implemented in the penalty matrix. If the penalty matrix fulfills the distance function, it is automatically also called the distance matrix. The following is the simulation result of the sequence alignment process in MATLAB® displayed in a distance matrix.

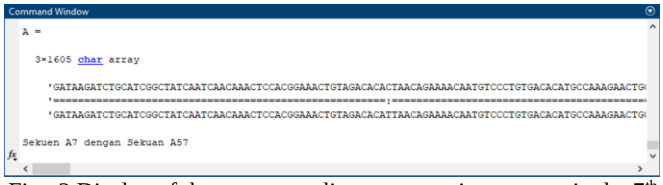


Fig. 2 Display of the sequence alignment running process in the 7th sequence with the 56th sequence

The "=" sign indicates the nucleotides aligned together are the same, while the ":" symbol indicates a different nucleotide.

The following is the distance matrix of the H9N2 sequence alignment results. The matrix size is 75x75, where each matrix element represents the genetic distance from the corresponding sequence. Row 1, Column 1 represents the alignment between Sequence 1 and Sequence 1; Row 3, Column 2 represents the alignment between Sequence 3 and Sequence 2, and so on.

	Seq1	Seq2	Seq3	.....	Seq75
Seq1	0	0,00246	0,000615	.....	0,021525
Seq2	0,00246	0	0,003075	.....	0,02337
Seq3	0,000615	0,003075	0	.....	0,02214
Seq4	0,00492	0,00738	0,005535	.....	0,024922
Seq5	0,000615	0,003075	0,00123	.....	0,02243
Seq6	0,009225	0,011685	0,00984	.....	0,026791
Seq7	0,01107	0,01353	0,011685	.....	0,029907
⋮	⋮	⋮	⋮	⋱	⋮
⋮	⋮	⋮	⋮	⋱	⋮

⋮	⋮	⋮	⋮	⋱	⋮
Seq75	0,021525	0,02337	0,02214	.....	0

Since the 75x75 matrix cannot display the whole matrix, the matrix view represents only a few elements.

3.2. Derivation of the Felsenstein Formula

It is necessary to derive the Felsenstein model formula to find the evolutionary distance denoted as the branch length v. Since, in this case, using molecular events (sequences), v must be defined. The notation v is the distance between sequences, and this distance directly relates to the amount of difference between the observed sequences. By considering equation (2), it is necessary to derive the formula to get the length of the branch v.

Suppose $\{x_1, x_2, x_3, \dots, x_n\}$ and $\{y_1, y_2, y_3, \dots, y_n\}$ are two nucleotide sequences with position n. We calculate the distance between them using the maximum likelihood method. The likelihood we meant here is the likelihood of one sequence and ends with another sequence. The formula for likelihood starts with sequence x and ends with sequence y, i.e.

$$L = p(x_1)p(x_1 \to y_1|t)p(x_2)p(x_2 \to y_2|t) \dots \dots p(x_n)p(x_n \to y_n|t) \quad (3)$$

where $p(x_i)$ is the probability of starting with the nucleotide x_i, and $p(x_i \to y_i|t)$ is the probability of changing from x_i to y_i, given by t.

Equation (3) results in the following equation

$$\ln L = \ln (p(x_1)\ p(x_1 \to y_1|t)p(x_2)p(x_2 \to y_2|t) \dots \dots \dots p(x_n)p(x_n \to y_n|t)$$

$$\ln L = \ln (p(x_1) + ln\, p(x_1 \to y_1|t) + \ln p(x_2) + \ln p(x_2 \to y_2|t) + \cdots + ln\, p(x_n) + ln\, p(x_n \to y_n|t) \quad (4)$$

$$\ln L = \ln (p(x_1) + ln\, p(x_2) + \cdots + ln\, p(x_n) + ln\ p(x_1 \to y_1|t) + ln\, p(x_2 \to y_2|t) + \cdots + ln\, p(x_n \to y_n|t) \quad (5)$$

Suppose:

$\ln (p(x_1) + ln\, p(x_2) + \cdots + ln\, p(x_n) = constanta = C$, then:

$$\ln L = C + ln\ p(x_1 \to y_1|t) + ln\, p(x_2 \to y_2|t) + \cdots + ln\, p(x_n \to y_n|t) \quad (6)$$

The probability of substituting a nucleotide x to y where $x \neq y$ is the same for all x and y and replacing a nucleotide x to x is the same for all x. Thus, the equation becomes:

$$\ln L = C + m_1 ln\, p_{ij}(t) + m_2 ln\, p_{ii}(t) \quad (7)$$

where:

$$P_{ij}(t) = p(x_k \to y_k)\ if\ x_k \neq y_k$$
$$P_{ii}(t) = p(x_k \to y_k)\ if\ x_k = y_k$$

m_1 is the number of positions where the substitution to a nucleotide is different,

m_2 is the number of positions where the substitution occurs to the same nucleotide.

To find the maximum likelihood for t, then ln L is differentiated and obtained:

$$\frac{d(\ln L)}{dt} = \frac{m_1}{P_{ij}(t)}P_{ij}'(t) + \frac{m_2}{P_{ii}(t)}P_{ii}'(t) \quad (8)$$

If equal to 0, then we get:

$$\frac{m_1}{P_{ij}(t)}P_{ij}'(t) + \frac{m_2}{P_{ii}(t)}P_{ii}'(t) = \frac{m_1}{\pi_j(1-e^{-\beta v})}\beta\pi_j e^{-\beta v} + \frac{m_2}{e^{-\beta v}+\pi_j(1-e^{-\beta v})}(\pi_j - 1)\beta e^{-\beta v} \quad (9)$$

$$\frac{m_1}{\pi_j(1-e^{-\beta v})}\beta\pi_j e^{-\beta v} + \frac{m_2}{e^{-\beta v}+\pi_j(1-e^{-\beta v})}(\pi_j - 1)\beta e^{-\beta v} = 0 \quad (10)$$

$$\left(\frac{m_1\pi_j}{\pi_j(1-e^{-\beta v})} + \frac{m_2(\pi_j - 1)}{e^{-\beta v}+\pi_j(1-e^{-\beta v})}\right)\beta e^{-\beta v} = 0 \quad (11)$$

Then:

$$\left(\frac{m_1}{(1-e^{-\beta v})} + \frac{m_2(\pi_j - 1)}{e^{-\beta v}+\pi_j(1-e^{-\beta v})}\right) = 0 \quad or \quad \beta e^{-\beta v} \neq 0; so\ that \neq 0 \quad (12)$$

Suppose: $x = e^{-\beta v}$ then:

$$\frac{m_1}{1-x} + \frac{m_2\pi_j - m_2}{x + \pi_j - \pi_j x} = 0 \quad (13)$$

$$\frac{m_1}{1-x} + \frac{m_2(\pi_j - 1)}{x + \pi_j(1-x)} = 0 \quad (14)$$

$$m_1 x + m_1\pi_j - m_1\pi_j x + m_2\pi_j - m_2\pi_j x - m_2 + m_2 x = 0 \quad (15)$$

$$(m_1 - m_1\pi_j - m_2\pi_j + m_2)x + (m_1\pi_j + m_2\pi_j - m_2) = 0 \quad (16)$$

$$(m_1 + m_2 - \pi_j(m_1 + m_2)x + ((m_1 + m_2)\pi_j - m_2) = 0 \quad (17)$$

$$(1 - \pi_j)(m_1 + m_2)x + ((m_1 + m_2)\pi_j - m_2) = 0 \quad (18)$$

$$(1 - \pi_j)(m_1 + m_2)x = m_2 - \pi_j(m_1 + m_2) \quad (19)$$

$$x = \frac{m_2 - \pi_j(m_1 + m_2)}{(1 - \pi_j)(m_1 + m_2)} \quad (20)$$

If we returnx to its original value, then:

$$e^{-\beta v} = \frac{m_2 - \pi_j(m_1 + m_2)}{(1 - \pi_j)(m_1 + m_2)} \quad (21)$$

$$-\beta v = ln\left(\frac{m_2 - \pi_j(m_1 + m_2)}{(1 - \pi_j)(m_1 + m_2)}\right) \quad (22)$$

$$v = -\frac{1}{\beta} ln\left(\frac{m_2}{(1 - \pi_j)(m_1 + m_2)} - \frac{\pi_j}{(1 - \pi_j)}\right) \quad (23)$$

Suppose $m_1 + m_2 = n$, where n is the number of positions in the sequence.

$$v = -\frac{1}{\beta} ln\left(\frac{1}{(1 - \pi_j)}\frac{m_2}{n} - \frac{\pi_j}{(1 - \pi_j)}\right) \quad (24)$$

Defined P_{ij} as

$$P_{ij} = \frac{m_2}{n}$$
$$= \frac{number\ of\ nucleotides\ that\ are\ the\ same}{n}$$

So we obtain the value of v:

$$v = -\frac{1}{\beta} ln\left(\frac{P_{ij} - \pi_j}{1 - \pi_j}\right) \quad (25)$$

We used the formula in equation (25) to convert each element in the distance matrix so that the distance matrix turns into a Felsenstein evolutionary distance matrix like below:

	Seq1	Seq2	Seq3	.....	Seq75
Seq1	0	0,002469	0,000616	.....	0,035832
Seq2	0,002469	0	0,003089	.....	0,037861
Seq3	0,000616	0,003089	0	.....	0,036507
Seq4	0,018086	0,020658	0,018727	.....	0,025891
Seq5	0,018007	0,020568	0,018646	.....	0,027789
Seq6	0,0226	0,025206	0,02325	.....	0,027915
Seq7	0,024553	0,027174	0,025206	.....	0,031314
⋮	⋮	⋮	⋮	⋱	⋮
⋮	⋮	⋮	⋮	⋱	⋮
⋮	⋮	⋮	⋮	⋱	⋮
Seq75	0,035832	0,037861	0,036507	.....	0

3.3. Phylogenetic Tree of H9N2 Spread in Indonesia

The phylogenetic tree formation process using the neighbor-joining algorithm with the Felsenstein evolutionary model led to the following results. The first data result as the initial benchmark for alignment was a sequence with the closest branching being S-1 and S-49. This means that sequence 1 was more comparable to sequence 49 than the other sequences. S-1 represents MF 164879 from South Sulawesi, Indonesia, while S-49 represents MK 559842 from West Java, Indonesia.

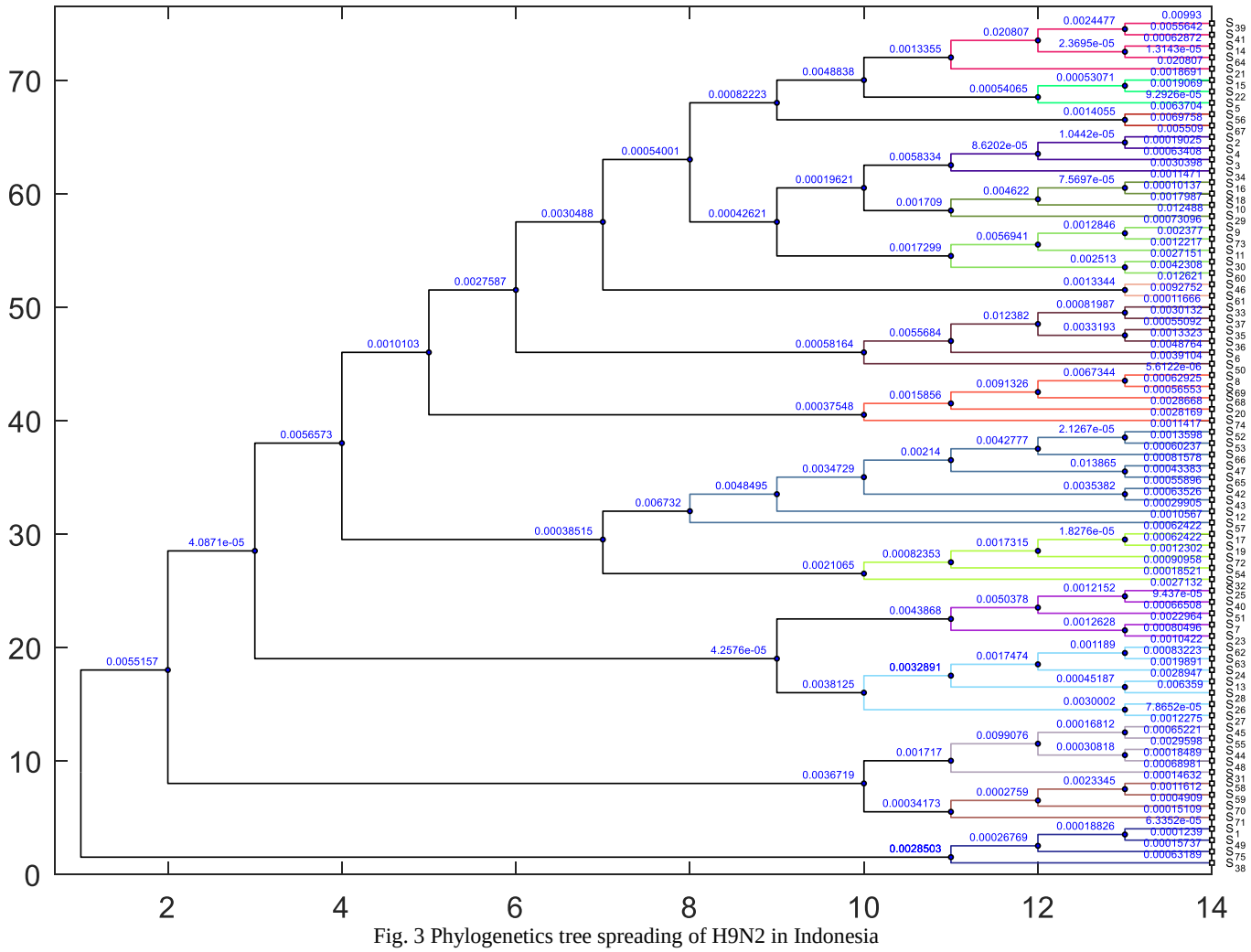


Fig. 3 Phylogenetics tree spreading of H9N2 in Indonesia

In Fig. 3, the visualization also shows the spread of the H9N2 virus forming 16 clusters, where each cluster offers the proximity of genes based on genetic distance, namely:

3.3.1. Cluster 1

Table 2 shows that sequence from S-1 and S-49 then spread to S-75 and then to S-38. Suppose cluster 1 is A. We represent A in a set based on the same branching. So that $A = \{((S_1, S_{49}), S_{75}), S_{38})\}$. Cluster 1 was chosen as an outgroup because there was a sequence MF 164879, which was the initial sequence for spreading the H9N2 virus into Indonesia.

Cluster 1 in Table 2 shows the spread of the virus from South Sulawesi, West Java, and East Java.

Table 2 Cluster 1 from phylogenetics tree spreading of H9N2 in Indonesia

Sequence	DNA acc.	Length	Coll. date	Province
S-1	MF 164879	1605	2016	South Sulawesi
S-49	MK 559842	1605	2017	West Java
S-75	MK 559868	1605	2018	East Java
S-38	MK 559831	1605	2017	West Java

S-38 was a sequence in 2017, but the sequence is similar to MK 559868, which appeared in 2018.

3.3.2. Cluster 2

For Table 3, if we assume B, then B can be represented as $B = \{(((S_{58}, S_{59}), S_{70}), S_{71})\}$. In cluster 2, the distribution is from Lampung, Banten, and East Java.

Table 3 Cluster 2 from phylogenetics tree spreading of H9N2 in Indonesia

Sequence	DNA acc.	Length	Coll. date	Province
S-58	MK 559851	1605	2018	Lampung
S-59	MK 559852	1605	2018	Banten
S-70	MK 559863	1605	2018	Lampung
S-71	MK 559864	1605	2018	East Java

3.3.3. Cluster 3

For Table 4, if we assume C, that can be represented as $C = \{((S_{45}, S_{55}), (S_{44}, S_{48})), \quad S31.)\}$

Table 4 Cluster 3 from phylogenetics tree spreading of H9N2 in Indonesia

Sequence	DNA acc.	Length	Coll. date	Province
S-45	MK 559838	1605	2017	Central Java
S-55	MK 559848	1605	2017	West Java
S-44	MK 559837	1605	2017	Central Java
S-48	MK 559841	1555	2017	East Java
S-31	MK 559824	1605	2017	East Java

3.3.4. Cluster 4

In Table 5, if we assume cluster 4 as D,then D = $\{((((S_{62}, S_{63}), S_{24}), (S_{13}, S_{28})), (S_{26}, S_{27}))\}$.

Table 5 Cluster 4 from phylogenetics tree spreading of H9N2 in Indonesia

Sequence	DNA acc.	Length	Coll. date	Province
S-62	MK 559855	1605	2018	West Java
S-63	MK 559856	1605	2018	West Java
S-24	MK 559817	1605	2017	Central Java
S-13	MK 559806	1605	2017	West Java
S-28	MK 559821	1605	2017	Central Java
S-26	MK 559819	1605	2017	West Java
S-27	MK 559820	1605	2017	West Java

3.3.5. Cluster 5

Table 6 assumes the branching proximity representation as E so that $E = \{(((S_{25}, S_{40}), S_{51}), (S_{7}, S_{23.}))\}$.

Table 6 Cluster 5 from phylogenetics tree spreading of H9N2 in Indonesia

Sequence	DNA acc.	Length	Coll. date	Province
S-25	MK 559818	1605	2017	Central Java
S-40	MK 559833	1605	2017	East Java
S-51	MK 559844	1605	2017	West Java
S-7	MK 559800	1605	2017	North Sumatera
S-23	MK 559816	1605	2017	West Java

3.3.6. Cluster 6

For Table 7, if we assume F, that can be represented as $F = \{(((((S_{17}, S_{19}), S_{72}), S_{54}), S_{32}))\}$.

Table 7 Cluster 6 from phylogenetics tree spreading of H9N2 in Indonesia

Sequence	DNA acc.	Length	Coll. date	Province
S-17	MK 559810	1605	2017	East Java
S-19	MK 559812	1605	2017	East Java
S-72	MK 559865	1605	2018	West Java
S-54	MK 559847	1605	2017	East Java
S-32	MK 559825	1581	2017	West Java

3.3.7. Cluster 7

In Table 8, we assume the branching proximity representation as G, so that $G = \{(((((((S_{52}, S_{53}), S_{66}), (S_{47}, S_{65})), (S_{42}, S_{43})), S_{12}), S_{57})\}$

Table 8 Cluster 7 from phylogenetics tree spreading of H9N2 in Indonesia

Sequence	DNA acc.	Length	Coll. date	Province
S-52	MK 559845	1605	2017	West Java
S-53	MK 559846	1605	2017	East Java
S-66	MK 559859	1605	2018	North Sulawesi
S-47	MK 559840	1605	2017	East Java
S-65	MK 559858	1605	2018	West Java
S-42	MK 559835	1605	2017	East Java
S-43	MK 559836	1605	2017	North Sumatera
S-12	MK 559805	1605	2017	Central Java
S-57	MK 559850	1605	2018	Lampung

3.3.8. Cluster 8

In Table 9, for cluster 8, it is assumed to be H$\{((((\quad : H = S8, S_{69}), (S_{68}), S_{20}), \quad S74.))\}$.

Table 9 Cluster 8 from phylogenetics tree spreading of H9N2 in Indonesia

Sequence	DNA acc.	Length	Coll. date	Province
S-8	MK 559801	1605	2017	Banten
S-69	MK 559862	1605	2018	Lampung
S-68	MK 559861	1605	2018	West Sumatera
S-20	MK 559813	1605	2017	East Java
S-74	MK 559867	1605	2018	West Java

3.3.9. Cluster 9

Meanwhile, the representation for cluster 9 in Table 10 if it is assumed to be I is: $I = \{((((S_{33}, S_{37}), (S_{35}, S_{36}), S_{6}), S_{50}))\}$.

Table 10 Cluster 9 from phylogenetics tree spreading of H9N2 in Indonesia

Sequence	DNA acc.	Length	Coll. date	Province
S-33	MK 559826	1605	2017	East Java
S-37	MK 559830	1605	2017	South Sulawesi
S-35	MK 559828	1598	2017	East Java
S-36	MK 559829	1583	2017	East Java
S-6	MK 559799	1605	2017	Central Java
S-50	MK 559843	1598	2017	East Java

3.3.10. Cluster 10

For Table 11, if we assume J, then J can be represented as $J = \{(S_{46}, S_{61})\}$.

Table 11 Cluster 10 from phylogenetics tree spreading of H9N2 in Indonesia

Sequence	DNA acc.	Length	Coll. date	Province
S-46	MK 559839	1605	2017	West Java
S-61	MK 559854	1603	2018	West Java

3.3.11. Cluster 11

In cluster 11, if cluster 11 in Table 12 is assumed to be K, it can be represented as $K\left\{\left(\left((= S_9, S_{73}), S_{11}\right), (S_{30}, S_{60})\right)\right\}$.

Table 12 Cluster 11 from phylogenetics tree spreading of H9N2 in Indonesia

Sequence	DNA acc.	Length	Coll. date	Province
S-9	MK 559802	1605	2017	East Java
S-73	MK 559866	1605	2018	East Java
S-11	MK 559804	1605	2017	West Java
S-30	MK 559823	1605	2017	West Java
S-60	MK 559853	1605	2018	East Java

3.3.12. Cluster 12

Table 13 assumes the branching proximity representation as L to be shown as $L\left\{\left(\left((= S_{16}, S_{18}), S_{10}\right), S_{29}\right)\right\}$.

Table 13 Cluster 12 from phylogenetics tree spreading of H9N2 in Indonesia

Sequence	DNA acc.	Length	Coll. date	Province
S-16	MK 559809	1605	2017	East Java
S-18	MK 559811	1605	2017	East Java
S-10	MK 559803	1605	2017	Banten
S-29	MK 559822	1605	2017	East Java

3.3.13. Cluster 13

While representing the proximity of the branching in Table 14, if we assume cluster 13 as M, then $M\left\{\left(\left((\quad)\quad\right)\quad_{=S2,S4,S3,S34.}\right)\right\}$.

Table 14 Cluster 13 from phylogenetics tree spreading of H9N2 in Indonesia

Sequence	DNA acc.	Length	Coll. date	Province
S-2	MF 164880	1605	2016	South Sulawesi
S-4	MK 559797	1605	2017	West Java
S-3	MF 164881	1605	2016	Central Sulawesi
S-34	MK 559827	1598	2017	East Java

3.3.14. Cluster 14

As with the previous cluster, we represent cluster 14 in Table 15 as the N set. So that $N = \{(S_{56}, S_{67})\}$.

Table 15 Cluster 14 from phylogenetics tree spreading of H9N2 in Indonesia

Sequence	DNA acc.	Length	Coll. date	Province
S-56	MK 559849	1605	2018	East Java
S-67	MK 559860	1605	2018	West Sumatera

3.3.15. Cluster 15

Suppose cluster 15 in Table 16 is O, then we represent $O\{(= S_{15}, S_{22}), S_5\}$.

Table 16 Cluster 15 from phylogenetics tree spreading of H9N2 in Indonesia

Sequence	DNA acc.	Length	Coll. date	Province
S-15	MK 559808	1605	2017	East Java
S-22	MK 559815	1605	2017	Central Java
S-5	MK 559798	1598	2017	South Sulawesi

3.3.16. Cluster 16

Suppose cluster 16 is P. We represent P in a set based on the same branching. So that $P = \left\{\left((S_{39}, S_{41}), (S_{14}, S_{64})\right), S_{21}\right)\right\}$.

Table 17 Cluster 16 from phylogenetics tree spreading of H9N2 in Indonesia

Sequence	DNA acc.	Length	Coll. date	Province
S-39	MK 559832	1603	2017	Banten
S-41	MK 559834	1605	2017	Central Java
S-14	MK 559807	1605	2017	West Java
S-64	MK 559857	1601	2018	West Java
S-21	MK 559814	1605	2017	West Java

Cluster 1 shows the cluster of H9 virus spread in the early part of the spread. Since the initial rule of alignment is S-1, all sequences are aligned one by one with S-1. The result indicates S-1 is closer to S-49. If we look back [1], the first case of avian influenza subtype H9N2 on laying hens in Sidrap Regency, South Sulawesi, Indonesia, occurred in December 2016. When checked, S-49 has a close relationship with the first case in the previous year, namely 2016. In detail, the explanation is in the following section.

3.4. Mutation from the Initial Sequence to the Next Sequence

This section displays the mutation of the sequence originating from South Sulawesi, namely (M92_02/2016) access code MF164879 with South Sulawesi MK559798/2017. We used the Needleman-Wunsch algorithm simulated in MATLAB®2020 to conduct alignment. Fig. 4 displays the program.

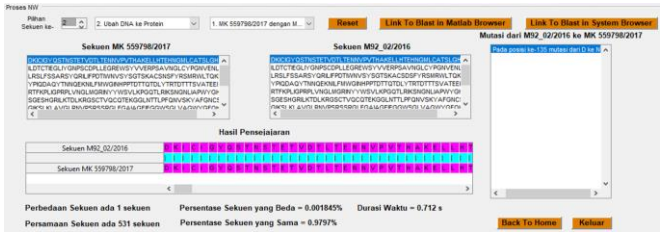


Fig. 4 Mutation from sequence MF164879 to MK559798

The sequence data used is DNA data, then we converted them into protein data with a 532 aa length in the simulation. The results showed that the difference in sequences was only 1; the same sequence was 531. The percentage of different sequences was 0.1845%, while the same sequence was 97.97%. Because the sequence difference is only 1, mutations also occur at one position, namely from protein D to protein N at the 135th position.

Table 18 Mutation of sequence from Sulawesi 2016 with sequence from Sulawesi 2017

Sequence 1	Sequence 2	Protein		Position	Mutation
		Diff.	Sim.		
MF164879 South Sulawesi 2016 (S-1)	MK559798 South Sulawesi 2017 (S-5)	1	531	135	D to N
MF164880 South Sulawesi 2016 (S-2)	MK559798 South Sulawesi 2017 (S-5)	2	530	135 172	D to N V to I

In summary, Table 18 shows the mutations from sequences originating from Sulawesi in 2016 (S-1 and S-2) with sequences from South Sulawesi in 2017 (S-5).

Table 19 displays mutations that occur from one sequence to another in cluster 1. The objective is to find out how many mutations happened in that cluster.

Table 19 Mutation of sequence from Cluster 1

Seq. 1	Seq. 2	Protein		Position (Mutation)	The same mutation	
		Diff.	Sim.			
S-1	S-49	6	529	34 (HQ), 79 (VI), 148 (ND), 183 (DN), 204 (KR), 384 (DE)	79 (VI), 384 (DE)	-
S-1	S-75	19	516	74 (RG), 79 (VI), 112 (QR), 169 (MR), 173 (NH), 177 (TK), 178 (DK), 179 (TN), 180 (TV), 181 (QL), 182 (TF), 195 (AT), 198 (ED), 212 (LR), 217 (MT), 363 (KR), 384 (DE), 467 (QR), 528 (FL)	79 (VI), 384 (DE)	112 (QR), 173 (NH), 180 (TV), 182 (TF)
S-1	S-38	10	525	79 (VI), 112 (QR), 173 (NH), 180 (TE), 182 (TI), 183 (DN), 186 (TK), 187 (RK), 204 (KN), 384 (DE)	79 (VI), 384 (DE)	112 (QR), 173 (NH), 180 (TV), 182 (TF)

In Cluster 1, the same mutation from sequence-1 and sequence-49, sequence-1 and sequence-75, sequence-1 and sequence-38, occurs in position 79 (from V to I) and384 (from D to E). Furthermore, the same mutation from sequence-1 and sequence-75, sequence-1 and sequence-38, occurs in positions 112 (from Q to R),173 (from N to H), 180 (from T to V), and 182 (from T to F).

Researchers have conducted several previous studies related to identifying the H9N2 subtype avian influenza virus in chickens in Indonesia [11]. It is also said the first case of H9N2 in Sidrap, South Sulawesi, Indonesia [9]. Moreover, some studies have performed global phylogenetic analysis to determine host species and geographic features of the evolution of avian influenza H9N2 hemagglutinin [18]. However, so far, all of the similar studies have focused on objects. They employ existing program packages such as MEGA, CLUSTAL, and PHYLIP to conduct analysis. This paper proposes to obtain research results using existing data in NCBI but working with other methods. The research focuses on methods, both on phylogenetic trees' formation with the evolutionary model used and mutations between sequences.

4. Conclusion

There were 16 clusters of H9N2 virus spread in Indonesia, identified from phylogenetic trees with Felsenstein models. We obtain mutation results that showed the evolution from the 2016 to 2018 sequence. We got the results that in one cluster, several mutations were the same in a specific location, as shown in the results of the Cluster 1 mutation.

Acknowledgments

The Basic Research Grant supports this study, the Ministry of Research and Technology/National Research and Innovation Agency of the Republic of Indonesia (Grant No. 3/E1/KP.PTNBH/2020).

References

[1] MUFLIHANAH, ANDESFHA E., WIBAWA H., ZENAL, F. C., HENDRAWATI, F., SISWANI, WAHYUNI, KARTINI D., RAHAYUNINGTYAS I., HADI S., MUKARTINI S., POERMADJAJA B., and TJATUR RASA F. S. Kasus pertama low pathogenic avian influenza subtipe H9N2 pada peternakan ayam petelur di Kabupaten Sidrap, Sulawesi Selatan Indonesia. *Diagnosa Veteriner*, 2017, 16(1): 1-13.

[2] MURAKAMI S., IWASA A., IWATSUKI-HORIMOTO K., ITO M., KISO M., KIDA H., TAKADA A., NIDOM C. A., MAI LE Q., YAMADA S., IMAI H., SAKAI-TAGAWA Y., KAWAOKA Y., and HORIMOTO T. Cross-clade

protective immunity of H5N1 influenza vaccines in a mouse model. *Vaccine*, 2008, 26(50): 6398-6404. https://doi.org/10.1016/j.vaccine.2008.08.053
[3] MONTEERARAT Y., SAKABE S., NGAMURULERT S., SRICHATRAPHIMUK S., JIAMTOM W., CHAICHUEN K., THITITHANYANONT A., PERMPIKUL P., SONGSERM T., PUTHAVATHANA P., NIDOM C. A., MAI LE Q., IWATSUKI-HORIMOTO K., KAWAOKA Y., and AUEWARAKUL P. Induction of TNF-alpha in human macrophages by avian and human influenza viruses. *Archives of Virology*, 2010, 155(8): 1273-1279. https://doi.org/10.1007/s00705-010-0716-y
[4] SAKABE S., TAKANO R., NAGAMURA-INOUE T., YAMASHITA N., NIDOM C. A., QUYNH LE M. T., IWATSUKI-HORIMOTO K., and KAWAOKA Y. Differences in cytokine production in human macrophages and in virulence in mice are attributable to the acidic polymerase protein of highly pathogenic influenza A virus subtype H5N1. *Journal of Infectious Diseases*, 2013, 207(2): 262-271. https://doi.org/10.1093/infdis/jis523
[5] DAVIDSON I., SHKODA I., GOLENDER N., PERK S., LAPIN K., KHINICH Y., and PANSHIN A. Genetic characterization of HA gene of low pathogenic H9N2 influenza viruses isolated in Israel during 2006-2012 periods. *Virus Genes*, 2013, 46(2): 255-263. https://doi.org/10.1007/s11262-012-0852-4
[6] YAMAJI R., YAMADA S., LE M. Q., LI C., CHEN H., QURNIANINGSIH E., NIDOM C. A., ITO M., SAKAI-TAGAWA Y., and KAWAOKA Y. Identification of PB2 mutations responsible for the efficient replication of H5N1 influenza viruses in human lung epithelial cells. *Journal of Virology*, 2015, 89(7): 3947-3956. https://doi.org/10.1128/JVI.03328-14
[7] SAKAI-TAGAWA Y., YAMAYOSHI S., KAWAKAMI C., LE M. Q., UCHIDA Y., SAITO T., NIDOM C. A., HUMAIRA I., TOOHEY-KURTH K., ARAFA A.-S., LIU M.-T., SHU Y., and KAWAOKA Y. Reactivity and sensitivity of commercially available influenza rapid diagnostic tests in Japan. *Scientific Reports*, 2017, 7(1): 14483. https://doi.org/10.1038/s41598-017-14536-0
[8] BONFANTE F., MAZZETTO E., ZANARDELLO C., FORTIN A., GOBBO F., MANIERO S., BIGOLARO M., DAVIDSON I., HADDAS R., CATTOLI G., and TERREGINO C. A G1-lineage H9N2 virus with oviduct tropism causes chronic pathological changes in the infundibulum and a long-lasting drop in egg production. *Veterinary Research*, 2018, 49(1): 83. https://doi.org/10.1186/s13567-018-0575-1
[9] PEACOCK T. H. P., JAMES J., SEALY J. E., and IQBAL M. A global perspective on H9N2 avian influenza virus. *Viruses*, 2019, 11(7): 620. https://doi.org/10.3390/v11070620
[10] NOVIANTI A. N., RAHARDJO K., PRASETYA R. R., NASTRI A. M., DEWANTARI J. R., RAHARDJO A. P., ESTOEPANGESTIE A. T. S., SHIMIZU Y. K., POETRANTO E. D., SOEGIARTO G., MORI Y., and SHIMIZU K. Whole-genome sequence of an avian influenza A/H9N2 virus isolated from an apparently healthy chicken at a live-poultry market in Indonesia. *Microbiology Resource Announcements*, 2019, 8(17): e01671-18. https://doi.org/10.1128/MRA.01671-18
[11] JONAS M., SAHESTI A., MURWIJATI T., LESTARININGSIH C. L., IRINE I., AYESDA C. S., PRIHARTINI W., and MAHARDIKA G. N. Identification of avian influenza virus subtype H9N2 in chicken farms in Indonesia. *Preventive Veterinary Medicine*, 2018, 159: 99-105. https://doi.org/10.1016/j.prevetmed.2018.09.003
[12] LIKIĆ V. A., MCCONVILLE M. J., LITHGOW T., and BACIC A. Systems biology: The next frontier for bioinformatics. *Advances in Bioinformatics*, 2010, 2010: 268925. https://doi.org/10.1155/2010/268925
[13] TAKAHASHI T., NIDOM C. A., QUYNH LE M. T., SUZUKI T., and KAWAOKA Y. Amino acid determinants are conferring stable sialidase activity at low pH for H5N1 influenza A virus neuraminidase. *FEBS Open Bio*, 2012, 2(1): 261-266. https://doi.org/10.1016/j.fob.2012.08.007
[14] AMIROCH S., & ROHMATULLAH A. Determining geographical spread pattern of MERS-CoV by distance method using Kimura model. *AIP Conference Proceedings*, 2017, 1825: 020001. https://doi.org/10.1063/1.4978970
[15] PRADANA M. S., & AMIROCH S. Protein sequence analysis of the Zika virus and the dengue virus using Smith-Waterman algorithm. *AIP Conference Proceedings*, 2019, 2084: 020011. https://doi.org/10.1063/1.5094275
[16] IRAWAN I., & AMIROCH S. Construction of phylogenetic tree using neighbor-joining algorithms to identify the host and the spreading of SARS epidemic. *Journal of Theoretical and Applied Information Technology*, 2015, 71(3): 424-429. http://www.jatit.org/volumes/Vol71No3/13Vol71No3.pdf
[17] FELSENSTEIN J. Evolutionary trees from DNA sequences: A maximum likelihood approach. *Journal of Molecular Evolution*, 1981, 17(6): 368-376. https://doi.org/10.1007/BF01734359
[18] DALBY A. R., & IQBAL M. A global phylogenetic analysis in order to determine the host species and geography dependent features present in the evolution of avian H9N2 influenza hemagglutinin. *PeerJ*, 2014, 2: e655. https://doi.org/10.7717/peerj.655
[19] BANKS J., SPEIDEL E. C., HARRIS P. A., and ALEXANDER D. J. Phylogenetic analysis of influenza A viruses of H9 haemagglutinin subtype. *Avian Pathology*, 2000, 29(4): 353-359. https://doi.org/10.1080/03079450050118485
[20] MONDAL S., KUMAR V., and SINGH S. P. Phylogenetic distribution and structural analyses of cyanobacterial glutaredoxins (Grxs). *Computational Biology and Chemistry*, 2020, 84: 107141. https://doi.org/10.1016/j.compbiolchem.2019.107141
[21] NATIONAL CENTER FOR BIOTECHNOLOGY INFORMATION, n.d. https://www.ncbi.nlm.gov/

参考文:

[1] MUFLIHANAH，ANDESFHA E.，WIBAWA H.，ZENAL，F. C.，HENDRAWATI，F.，SISWANI，WAHYUNI，KARTINI D.，RAHAYUNINGTYAS I.，HADI S.，MUKARTINI S.，POERMADJAJA B. 和 TJATUR RASA F. S. 第一种情况低致病性禽流感亚型西达拉普摄政层鸡场的 H9N2，印度尼西亚苏拉威西岛。诊断兽医，2017，16（1）：1-13。
[2] MURAKAMI S.，IWASA A.，IWATSUKI-HORIMOTO K.，ITO M.，KISO M.，KIDA H.，TAKADA A.，NIDOM C. A.，MAI LE Q.，YAMADA S.

，IMAI H.，SAKAI -TAGAWA Y.，KAWAOKA Y. 和 HORIMOTO T. 哈兴 1 流感疫苗在小鼠模型中的交叉保护性免疫。疫苗，2008，26（50）：6398-6404。https://doi.org/10.1016/j.vaccine.2008.08.053
[3] MONTEERARAT Y.，SAKABE S.，NGAMURULERT S.，SRICHATRAPHIMUK S.，JIAMTOM W.，CHAICHUNN K.，THITITHANYANONNT A.，PERMPIKUL P.，SONGSERM T.，PUTHAVATHANA P.，NIDOM C. A.，MAI LE Q.。，IWATSUKI-HORIMOTO K.，KAWAOKA Y. 和 AUEWARAKUL P. 禽和人流感病毒在人巨噬细胞中诱导肿瘤坏死因子-α。病毒学档案，2010，155（8）：1273-1279。https://doi.org/10.1007/s00705-010-0716-y
[4] SAKABE S.，TAKANO R.，NAGAMURA-INOUE T.，YAMASHITA N.，NIDOM C. A.，QUYNH LE M. T.，IWATSUKI-HORIMOTO K. 和 KAWAOKA Y. 人类巨噬细胞中细胞因子产生的差异和小鼠的毒性归因于高致病性甲型流感病毒哈兴 1 亚型的酸性聚合酶蛋白。传染病杂志，2013，207（2）：262-271。https://doi.org/10.1093/infdis/jis523
[5] DAVIDSON I.，SHKODA I.，GOLENDER N.，PERK S.，LAPIN K.，KHINICH Y. 和 PANSHIN A. 2006-2012 年间在以色列分离出的低致病性我们的权利流感病毒哈基因的遗传特征。病毒基因，2013，46（2）：255-263。https://doi.org/10.1007/s11262-012-0852-4
[6] YAMAJI R.，YAMADA S.，LE M. Q.，LI C.，CHEN H.，QUENIANINGSIH E.，NIDOM C. A.，ITO M.，SAKAI-TAGAWA Y. 和 KAWAOKA Y。哈兴 1 流感病毒在人肺上皮细胞中的有效复制。病毒学杂志，2015，89（7）：3947-3956。https://doi.org/10.1128/JVI.03328-14
[7] 坂井孝太郎，山口胜夫，川久美夫，LE M. Q.，内田勇夫，斋藤 T.，NIDOM C. A.，HUMAIRA I.，TOOHEY-KURTH K.，ARAFA A.-S.，LIU M .-T。，SHU Y. 和 KAWAOKA Y. 日本市售流感快速诊断测试的反应性和敏感性。科学报告，2017，7（1）：14483. https://doi.org/10.1038/s41598-017-14536-0
[8] BONFANTE F.，MAZZETTO E.，ZANARDELLO C.，FORTIN A.，GOBBO F.，MANIERO S.，BIGOLARO M.，DAVIDSON I.，HADDAS R.，CATTOLI G. 和 TERREGINO C. 一种 G1-带有输卵管嗜性的我们的权利谱系病毒会引起漏斗状的慢性病理变化，并使产蛋量长期下降。兽医研究，2018，49（1）：83. https://doi.org/10.1186/s13567-018-0575-1
[9] PEACOCK T. H. P.，JAMES J.，SEALY J. E. 和 IQBAL M。关于我们的权利禽流感病毒的全球性观点。病毒，2019，11（7）：620. https://doi.org/10.3390/v11070620
[10] NOVIANTI A. N.，RAHARDJO K.，PRASETYA R. R.，NASTRI A. M.，DEWANTARI J. R.，RAHARDJO A. P.，ESTOEPANGESTIE A. T. S.，SHIMIZU Y. K.，POETRANTO E. D.，SOEGIARTO G.，MORI Y. 和 SHIMIZU K. 禽类的全基因组序列。在印度尼西亚的活禽市场中从看似健康的鸡中分离出一种 /我们的权利流感病毒。微生物资源公告，2019，8（17）：e01671-18。https://doi.org/10.1128/MRA.01671-18
[11] JONAS M.，SAHESTI A.，MURWIJATI T.，LESTARININGSIH C. L.，IRINE I.，AYESDA C. S.，PRIHARTINI W. 和 MAHARDIKA G. N. 在印度尼西亚养鸡场中鉴定出禽流感病毒我们的权利亚型。预防兽医，2018，159：99-105。https://doi.org/10.1016/j.prevetmed.2018.09.003
[12] LIKIĆV. A.，MCCONVILLE M. J.，LITHGOW T. 和 BACIC A. 系统生物学：生物信息学的下一个前沿领域。生物信息学进展，2010，2010：268925. https://doi.org/10.1155/2010/268925
[13] TAKAHASHI T.，NIDOM C. A.，QUIYNH LE M. T.，SUZUKI T. 和 KAWAOKA Y。氨基酸决定簇在低 pH 值下对哈兴 1 甲型流感病毒神经氨酸酶具有稳定的唾液酸酶活性。FEBS 开放式生物学，2012，2（1）：261-266。https://doi.org/10.1016/j.fob.2012.08.007
[14] AMIROCH S. 和 ROHMATULLAH A. 使用木村模型通过距离法确定冠状病毒的地理扩散模式。AIP 会议论文集，2017，1825：020001. https://doi.org/10.1063/1.4978970
[15] PRADANA M. S. 和 AMIROCH S. 使用史密斯·沃特曼算法对寨卡病毒和登革热病毒的蛋白质序列进行分析。AIP 会议论文集，2019，2084：020011. https://doi.org/10.1063/1.5094275
[16] IRAWAN I. 和 AMIROCH S. 使用邻居连接算法来识别宿主和非典流行病的传播，以构建系统树。理论与应用信息技术学报，2015，71（3）：424-429。http://www.jatit.org/volumes/Vol71No3/13Vol71No3.pdf
[17] FELSENSTEIN J. 脱氧核糖核酸序列进化树：最大似然法。分子进化杂志，1981，17（6）：368-376。https://doi.org/10.1007/BF01734359
[18] DALBY A. R. 和 IQBAL M. 为了确定宿主我们的权利流感血凝素进化过程中存在的宿主物种和地理相关特征，进行了全面的系统发育分析。同行，2014，2：e655。https://doi.org/10.7717/peerj.655
[19] BANKS J.，SPEIDEL E. C.，HARRIS P. A. 和 ALEXANDER D. J. H9 血凝素亚型的 A 型流感病毒的系统发生分析。禽病理，2000，29（4）：353-359。https://doi.org/10.1080/03079450050118485
[20] MONDAL S.，KUMAR V. 和 SINGH S. P. 蓝藻戊二糖毒素（Grxs）的系统发育分布和结构分析。计算生物学与化学，2020，84：107141. https://doi.org/10.1016/j.compbiolchem.2019.107141
[21] 国家生物技术信息中心 https://www.ncbi.nlm.gov/