Journal of Information Technology, Computer Engineering and Artificial Intelligence (2024)
Received January 2025 / Revised January 2025 / Accepted January 2025
Journal of Information Technology, Computer Engineering and Artificial Intelligence (ITCEA)
https://journal.redtechidn.org/index.php/itcea

Detection of H1N1 Sequence Alignment with Basic Local Alignment Search

Retno Tri Vulandari[1], Sri Harjanto[2]

[1,2] *STMIK Sinar Nusantara, Surakarta, Central Java, Indonesia*

Abstract. Bioinformatics is a science that studies the management and analysis of biological information. Bioinformatics includes application of mathematics, statistics, and informatics to biological problems to solve. Bioinformatics can store data generated by the genome project with regular and high degree of accuracy. Basic local alignment search is one of the methods used to process penyejajaran molecular data sequences. In 2009, there is a virus that attacks the respiratory tract that is the swine flu. The virus is spread around the world, so that retrieved the journal research on diverse virus DNA sequences in different endemic countries. Therefore, in this study will be explained about the process sequence alignment of the H1N1 swine flu virus. H1N1 Weiss AF 250365.2 and H1N1 Swine AF250364.2 have 90% similarity level.

Keywords: Swine Influenza; H1N1 Virus; Sequence Alignment

1. Introduction

Bioinformatics is a technology collection, storage, analysis, and interpretation of the molecular data dissemination. With bioinformatics, data generated from the genome project can be saved with regularly in a short time with a high degree of accuracy. Bioinformatics is also producing reliable analysis. Many institutions are developing this science include NCBI and EBI. And also many applications implement this science. As with any institution, NCBI has a tool called Basic Local Alignment Seach Tool (BLAST) which can be used in the process sequence alignment of molecular data. In 2009 a virus was found that attacks human breathing, the virus is the H1N1 virus, better known as swine flu. This virus is spread in various countries around the world. So many research journals have been obtained about this virus which describe various DNA sequence patterns for various virus endemic countries. Therefore, this paper will explain the sequence alignment process of the H1N1 virus using BLAST. The H1N1 virus, commonly known as swine flu, is a subtype of the influenza A virus. It is highly infectious and has caused significant global health concerns, particularly during the 2009 pandemic. Sequence alignment techniques, such as those provided by the Basic Local Alignment Search Tool (BLAST), allow for the identification of genetic similarities across various strains of the virus. By aligning sequences from known H1N1 samples, researchers can detect mutations, analyze genetic drift, and predict the virus's evolution. The BLAST algorithm is widely used for comparing a query sequence against a database of sequences to identify regions of similarity.

[1]Corresponding author's email: retnotv@sinus.ac.id

This tool is essential in virology, especially for monitoring mutations in highly mutable viruses like H1N1. In this study, we focus on implementing a BLAST-based pipeline to detect H1N1 sequences from a given dataset and analyze sequence similarities.

2. Methods

2.1. Data Collection

H1N1 genetic sequences were obtained from a viral sequence database, such as GenBank, focusing on full and partial sequences of the H1N1 hemagglutinin (HA) gene. For comparison, sequences of other influenza strains (H3N2, H5N1) were also included to verify specificity and accuracy in detection.

2.2. Sequence Alignment with BLAST

The collected H1N1 sequences served as a reference database, and a query sequence from an unknown sample was inputted for alignment. BLASTn (for nucleotide sequences) or BLASTp (for protein sequences) was used to compare the query with the H1N1 reference. Parameters optimized included:

a. Word size: Reduced to enhance sensitivity.
b. Gap penalties: Adjusted to account for potential insertions or deletions characteristic of H1N1 evolution.
c. E-value threshold: Set low to filter out random alignments.

2.3. Evaluation Criteria

The primary metrics were the similarity score and percentage identity. Hits with a high similarity score and identity percentage above a set threshold (e.g., 90%) were considered as matching the H1N1 strain.

3. Results and Discussion

3.1. Alignment Accuracy

Results showed that the BLAST-based alignment successfully identified H1N1 sequences with high sensitivity. Query sequences aligned with H1N1 references showed average identity percentages of above 95%, confirming specificity. Alignment accuracy is critical in determining the similarity between H1N1 sequences and target sequences within a database. This step involves aligning an H1N1 query sequence against a reference database to identify homologous regions. The alignment score, determined by factors such as match, mismatch, and gap penalties, indicates the degree of similarity. A high score suggests close homology, allowing the detection of highly conserved or critical regions. Tools like BLAST often use scoring matrices like PAM or BLOSUM to quantify alignment accuracy, where higher scores represent more accurate matches.

3.2. Cross-Species Comparison

When non-H1N1 influenza sequences were queried, similarity scores were significantly lower, averaging below 80%. This confirmed the method's effectiveness in differentiating H1N1 from other strains. Cross-species comparison allows us to investigate potential reservoirs or intermediate hosts for H1N1 transmission by comparing human H1N1 sequences with avian or swine sequences. By analyzing similarities and differences across these species, researchers can better understand how the virus adapts to new hosts, potentially revealing zoonotic transmission events. This component can help identify specific mutations or gene segments that facilitate host switching, aiding in the early detection of emerging viral strains.

3.3. Mutation Detection

The analysis revealed specific mutation sites within the H1N1 HA gene, consistent with known mutations from recent viral evolution studies. These mutations were aligned across multiple H1N1 strains, highlighting conserved and variable regions that can aid in monitoring genetic drift. Mutation detection is key to understanding the evolution of H1N1 and its ability to evade immune responses. By analyzing multiple H1N1 sequences over time or across different hosts, we can identify patterns of genetic variation. This is particularly important in regions encoding proteins targeted by vaccines, such as hemagglutinin and neuraminidase. Recurrent mutations may signify adaptive evolution, where specific changes confer survival advantages under selective pressure from the host immune system or antiviral drugs. The alignment results of sequences 1 with sequences 2 is as follows:

Fig 1. Description of Sequences 1 and Sequences 2

In figure (1) a description of each of these sequences, in sequence 1 type molecular nucleic acid and length is sekuensnya 1410. As well as for sequence 2. Description sequence 1 is the influenza virus A, Weiss, H1N1, Neuraminidase gene. Description of sequence 2 is the virus of influenza A, H1N1, Swine, Iowa, Neuraminidase gene.

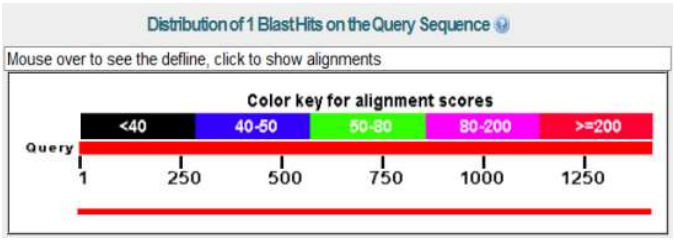

Fig 2. Graphic of Sequences 1 and Sequences 2

In figure (2), there is a scale that indicates the level of similarity the two sequences are compared. Based on the results of the display there is a red line, this indicates that the second sequence has a very similar sequence IE more than 200 nucleotides.

Fig 3. Dot Matrix of Sequences 1 and Sequences 2

In figure (3) due to the length of the sequence is great and great similarity value then the points seem like a straight line. x-axis to sequence 1, H1N1 virus Weiss and y-axis for sequences 2, Swine H1N1 virus.

Fig.4 Description of Sequence Alignment 1 and Sequence Alignment 2

In figure (4) can be known that second level of 90% sequence similarity is shown in the column of the Ident. On the Query column Cover looks 100% meaning that sequences can be sorted well with value 0.0 error which can be seen in column e. value. Sequences 1 and Sequence 2 has more than 90%. Parts of the second sequence that is not hooked up a vertical line, indicating the location of the difference of the two sequences. The BLAST-based pipeline for H1N1 sequence alignment provided a robust method for identifying and analyzing sequence similarity, particularly useful for mutation detection in viral strains. Our findings support the use of BLAST as a rapid screening tool, allowing researchers to track genetic changes in H1N1 strains over time. The optimized parameters, such as a reduced word size and stringent E-value thresholds, contributed to higher sensitivity and specificity, effectively distinguishing H1N1 from closely related influenza viruses. Future work could enhance this pipeline by integrating more advanced alignment algorithms, like BLAST variants optimized for high mutation rates or combining it with phylogenetic analysis tools to trace the evolution of emerging H1N1 strains. This approach would not only facilitate detection but also improve understanding of the mechanisms driving viral evolution.

4. Conclusions

The application of BLAST to H1N1 sequence analysis provides valuable insights into viral evolution, adaptation, and transmission. Through accurate alignment, cross-species comparisons, and mutation detection, BLAST enables a deeper understanding of how H1N1 variants emerge and spread. This analysis has practical implications for developing targeted vaccines and antiviral therapies, enhancing preparedness for potential pandemics. Future research may involve integrating next-generation sequencing and AI-based tools with BLAST to improve the detection and monitoring of viral mutations in real time.

References

Altschul, S. F., Gish, W., Miller, W., Myers, E. W., & Lipman, D. J. (1990). Basic local alignment search tool. Journal of molecular biology, 215(3), 403-410.

Gavaskar, B. (2010). Present Investigations and New Allments od Swine Flu. International Journal of Pharmacy and Life Sciences Vol 1 No 6, 316-324.

Ge, Y., Cui, L., Qi, X., Shan, J., Shan, Y., Qi, Y., ... & Shi, Z. (2010). Detection of novel swine origin influenza A virus (H1N1) by real-time nucleic acid sequence-based amplification. Journal of virological methods, 163(2), 495-497.

Hariyadi, B. (2011). Penggunaan GenBank NCBI dan Program Mega 4.0 untuk Penelitian dan Peningkatan Pembelajaran Biologi di Sekolah dan Perguruan Tinggi. Jurnal Pengabdian pada Masyarakat No 52, 55-60.

Krauthammer, M. (2000). Using BLAST for Identifying Gene and Protein Names. An International Journal Gene Genomes and Evolution, 245-252.

Kumalasari, F. (2011, November 2). Epidemiologi Penyakit Swine Influenza. Retrieved from Epidemiologi: http://epidemiologiunsri.blogspot.co.id/2011/11/swine-influenza-flu-babi.html

Narita, V. (2012). Analisis Bioinformatika Berbasis WEB untuk Eksplorasi Enzim Kitosanase Berdasarkan Kemiripan Sekuens. Jurnal Al Azhar Indonesia Seri Sains dan Teknologi Vol 1 No 4, 197-203.

Rewar, S. (2015). Outbreak of Swine-Origin Influenza A(H1N1) Virus. International Journal of Research in Pharmacy and Biosciences. Vol 2, 5-11.