International Journal of Trends in Mathematics Education Research
Vol. 2, No. 1, February 2019, pp. 9-12
Available online at http://ijtmer.com

E-ISSN : 2621-8488

RESEARCH ARTICLE

Implementation of Basic Local Alignment Search for Detection of H1N1 Sequence Alignment

Retno Tri Vulandari[1], Sri Siswanti[2], Andriani Kusumaningrum K[3], Kumaratih Sandradewi[4]

[2]Informatics Engineering, STMIK Sinar Nusantara, Surakarta, Indonesia
[3, 4]Computer Accounting, STMIK Sinar Nusantara, Surakarta, Indonesia
**Corresponding author: retno.tv@gmail.com*

How to Cite: Vulandari, R.T., Siswanti, S., Kusumaningrum. A.K, & Sandradewi. K. (2019). Implementation of Basic Local Alignment Search for Detection of H1N1 Sequence Alignment .*International Journal of Trends in Mathematics Education Research,* 2(1), 9-12. DOI: **http://dx.doi.org/10.33122/ijtmer.v2i1.20**

ARTICLE HISTORY

Received: 2 October 2018
Revised: 17 February 2019
Accepted: 26 February 2019

KEYWORDS

Swine influenza
H1N1 virus
Sequence Alignment

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.

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 (Narita, 2012,).
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.

2. RESEARCH METHOD

2.1 Swine influenza H1N1

Swine influenza is a respiratory disease caused by influenza viruses. This disease is actually often attack the pigs, but now it has undergone drastic changes and is able to infect humans. The symptoms that arise on any human being is similar to what happens in pigs.
Swine flu was first isolated from a pig in 1930 are infected in the United States. In its development, this disease can be passed to humans especially attacking those that close contact with pigs. Long time no sound again reportedly turned out to be a series of mutations of the virus so that it appears a new variant which first attacked humans in Mexico in early 2009. This new variant, known as the H1N1 virus which is an abbreviation of two major virus antigens i.e. hemagglutinin type 1 and neuraminidase type 1. Here is a picture of the H1N1 virus (Gavaskar, 2010, Altschul, 1990, Ge, 2010)

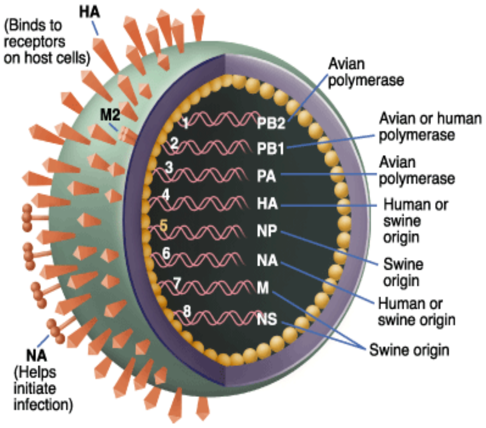


Fig 1. H1N1 Virus (Kumalasari, 2011)

2.2 NCBI and BLAST

NCBI (National Center for Biotechnology Information) is an institution that concentrated as a source of information on the development of molecular biology. NCBI database made accessible to the public, stimulating the research of biological terkomputasi data Analyzer software, developing a genome, and spread the information of which expected the biomedical leads to a better understanding of molecular processes affecting human and health. Here are the menus provided by the NCBI (Hariyadi, 2011).

1. Entrez
   Entrez is a search system information in the NCBI provides integrated access to do sequencing, mapping (mapping), taxonomy and structural data. Entrez also provides a graphical depiction for mapping sequence and chromosome. The characteristic and advantage Entrez is the ability to search related information sequence, structure and references. Available journal literature is accessible through PubMed. PubMed is a connecting tool web search which provides access to over 11 million citations in MEDLINE journals.
2. Nucleotide Database
   A database is a collection of nucleotide sequence from several sources, including GenBank, the Reference Sequence (RefSeq), Third Party Annotation (TPA) and the Protein Data Bank (PDB). GenBank is the NIH genetic sequence database of the (National Institutes of Health), a collection of DNA sequence which can be known by the public. Database of GenBank financed and distributed by NCBI. The data sequence is sent to GenBank by researchers from all over the world.
3. Blast
   Basic Local Alignment Search Tool (BLAST) is a program for searching sequence similarity (sequence similarity) and is a great tool in the identification of genes and genetic characters. BLAST sequence search can do through a comparison with DNA databases within a short time (less than 15 seconds).

There are 5 major programs in BLAST, namely (Krauthammer, 2000):

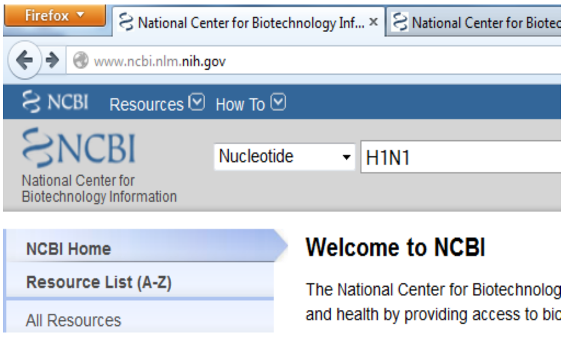
1. the nucleotide blast (blastn): compare a nucleotide sequence that we have with the nucleotide sequence database
2. protein blast (blastp): comparing a sequence of amino acids that we have with a protein sequence database
3. blastx: compare products 6-frame translation of the concept of a sequence of nucleotides (nucleotide translated) that we have with a protein sequence database
4. tblastn: compare a protein sequence that we have with the nucleotide sequence database dynamically ditranslasi on all 6 reading frames
5. tblastx: comparing a translation of 6 frames of nucleotides.

3. RESULT AND ANALYSIS

3.1 Data Process dan Data Capture

The steps use the BLAST indicated on the flow of the following methods,

1. Open NCBI's home page.



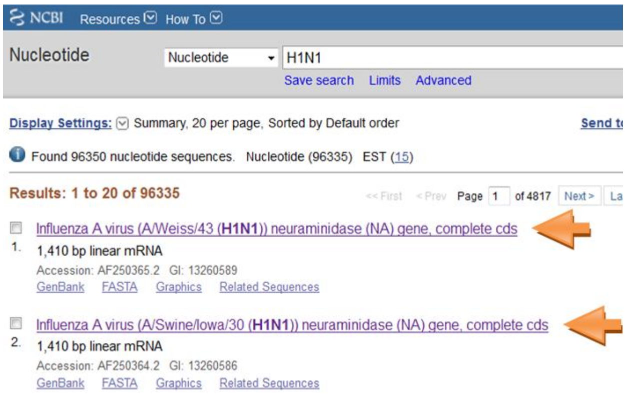
Welcome to NCBI

The National Center for Biotechnolog
and health by providing access to bic

Fig 2. Home Page NCBI

2. Locate the access code sequences which will be compared

Before the start of BLAST, look for code access sequences of nucleotides that are compared. In this paper are described the process of pensejajaran between the H1N1 virus sequences at Nucleotide database put the keywords in the search field of the H1N1.



Display Settings: Summary, 20 per page, Sorted by Default order

Found 96350 nucleotide sequences. Nucleotide (96335) EST (15)

Results: 1 to 20 of 96335

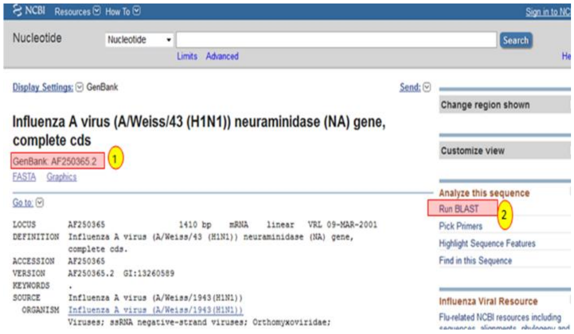
☐ 1. Influenza A virus (A/Weiss/43 (**H1N1**)) neuraminidase (NA) gene, complete cds
1,410 bp linear mRNA
Accession: AF250365.2 GI: 13260589
GenBank FASTA Graphics Related Sequences

☐ 2. Influenza A virus (A/Swine/Iowa/30 (**H1N1**)) neuraminidase (NA) gene, complete cds
1,410 bp linear mRNA
Accession: AF250364.2 GI: 13260586
GenBank FASTA Graphics Related Sequences

Fig 3. Nucleotide Sequence Search Results

Sequences which we will align on paper is the H1N1 virus sequences. After a search, found 96,350 nucleotide sequence and a sequence taken 2 are compared, one of which can be taken in sequence 1 and 2.

After the selected sequence 1, will be obtained the access codes the sequences in GenBank IE AF 250365.2 as in point (1) in figure (4). The code is used at Run BLAST.



Display Settings: GenBank

Influenza A virus (A/Weiss/43 (H1N1)) neuraminidase (NA) gene, complete cds

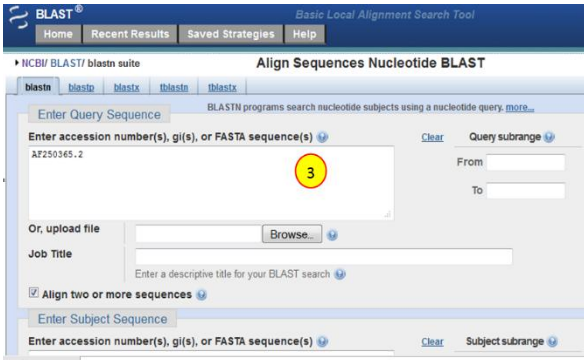
GenBank: AF250365.2

```
LOCUS       AF250365               1410 bp    mRNA    linear   VRL 09-MAR-2001
DEFINITION  Influenza A virus (A/Weiss/43 (H1N1)) neuraminidase (NA) gene,
            complete cds.
ACCESSION   AF250365
VERSION     AF250365.2  GI:13260589
KEYWORDS    .
SOURCE      Influenza A virus (A/Weiss/1943(H1N1))
  ORGANISM  Influenza A virus (A/Weiss/1943(H1N1))
            Viruses; ssRNA negative-strand viruses; Orthomyxoviridae;
```

Fig 4. Nucleotide Sequence 1

3. Process of Nucleotide Sequence Alignment BLAST

By entering your mobile number access, gi, or FASTA sequence 1 in column (3) in the figure (5) and the access number, gi, or FASTA sequence 2 in column (4) in Fig. (6) and press the BLAST on the picture (6) then it will be the second sequence alignment results obtained.



NCBI/ BLAST/ blastn suite

Align Sequences Nucleotide BLAST

BLASTN programs search nucleotide subjects using a nucleotide query.

Enter Query Sequence

Enter accession number(s), gi(s), or FASTA sequence(s): AF250365.2

Query subrange: From / To

Or, upload file

Job Title: Enter a descriptive title for your BLAST search

☑ Align two or more sequences

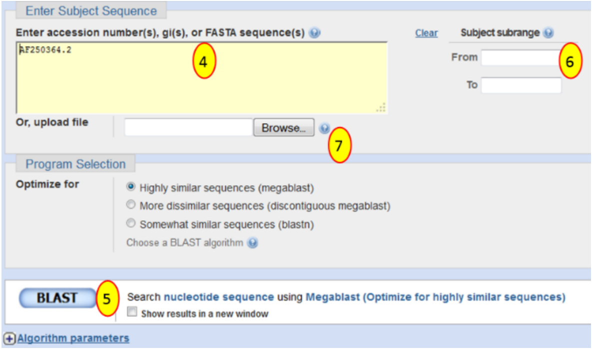
Enter Subject Sequence

Enter accession number(s), gi(s), or FASTA sequence(s)

Subject subrange

Fig 5. BLAST Running Process (1)

In column (6) the subject subrange, can be used if you don't want to use the entire sequence. For example only want to align the first subsekuens up to 100, then it can be filled from 1 to 100. In column (7) upload a file, it can be used if the DNA sequence fasta rather than from the NCBI GenBank.



Enter Subject Sequence

Enter accession number(s), gi(s), or FASTA sequence(s): AF250364.2

Subject subrange: From / To

Or, upload file

Program Selection

Optimize for:

● Highly similar sequences (megablast)

○ More dissimilar sequences (discontiguous megablast)

○ Somewhat similar sequences (blastn)

Choose a BLAST algorithm

Search nucleotide sequence using Megablast (Optimize for highly similar sequences)

☐ Show results in a new window

Fig 6. BLAST Running Process (2)

3.2 Interpretation Results with Blast

The alignment results of sequences 1 with sequences 2 is as follows:

NCBI/ BLAST/ blastn suite-2sequences/ Formatting Results - D1DSG4XD114

Blast 2 sequences

gb|AF250365.2| (1410 letters)

RID D1DSG4XD114 (Expires on 01-12 10:14 am)

Query ID gi|13260589|gb|AF250365.2|AF250365

Description Influenza A virus (A/Weiss/43 (H1N1)) neuraminidase (NA) gene, complete cds

Molecule type nucleic acid

Query Length 1410

Subject ID gi|13260586|gb|AF250364.2|AF250364

Description Influenza A virus (A/Swine/Iowa/30 (H1N1)) neuraminidase (NA) gene, complete cds

Molecule type nucleic acid

Subject Length 1410

Program BLASTN 2.2.29+

Fig 7. Description of Sequences 1 and Sequences 2

In figure (7) 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.

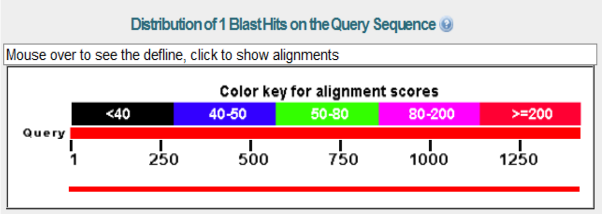


Fig 8. Graphic of Sequences 1 and Sequences 2

In figure (8), 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.

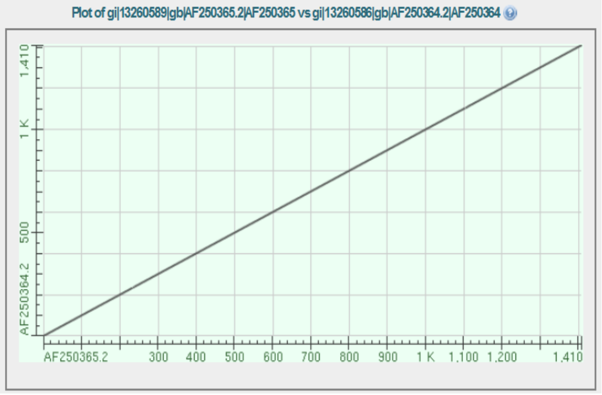


Fig 9. Dot Matrix of Sequences 1 and Sequences 2

In figure (9) 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.

	Description	Max score	Total score	Query cover	E value	Ident	Accession
☐	Influenza A virus (A/Swine/Iowa/30 (H1N1)) neuraminidase (NA) gene, complete cds	1834	1834	100%	0.0	90%	AF250364.2

Fig.10 Description of Sequence Alignment 1 and Sequence Alignment 2

In figure (10) 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.

4. CONCLUSION

Based on the previous description, can be summed up as follows BLAST is a program for searching sequence similarity and is a great tool in the identification of genes and genetic characters. BLAST can compare either the H1N1 virus sequences for pairwise alignment or for some multiple sequence alignment. Alignment results between H1N1 Weiss AF 250365.2 and H1N1 AF 250364.2 is the second level of 90% sequence similarity, it can be said both viruses in both the journal has a close kinship. From the results of the sequences alignment H1N1 virus Weiss AF 250365.2 with some other H1N1 virus sequences contained in the NCBI database, note that there is still one other H1N1 virus sequences are very similar or identical to the H1N1 virus sequences Weiss AF 250365.2 that has levels of 100% similarity.

ACKNOWLEDGEMENTS

The preferred spelling of the word “acknowledgment” in America is without an “e” after the “g”. Avoid the stilted expression, “One of us (R. B. G.) thanks . . .” Instead, try “R. B. G. thanks”. Put sponsor acknowledgments in the unnum-bered footnote on the first page.

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.