

## **SURAT KETERANGAN**

Nomor: 935/UNUSA-LPPM/Adm-I/VI/2022

Lembaga Penelitian dan Pengabdian Kepada Masyarakat (LPPM) Universitas Nahdlatul Ulama Surabaya menerangkan telah selesai melakukan pemeriksaan duplikasi dengan membandingkan artikel-artikel lain menggunakan perangkat lunak **Turnitin** pada tanggal 01 Juni 2022.

Judul : Bat Algorithm application for estimating Super Pairwise Alignment parameters on similarity analysis between virus protein sequences

Penulis : Dinita Rahmalia, Teguh Herlambang

No. Pemeriksaan : 2022.06.10.297

Dengan Hasil sebagai Berikut:

**Tingkat Kesamaan diseluruh artikel (*Similarity Index*) yaitu 17%**

Demikian surat keterangan ini dibuat untuk digunakan sebagaimana mestinya

Surabaya, 10 Juni 2022

Ketua LPPM,



Achmad Syafiuddin, Ph.D.

NPP. 20071300

**LPPM Universitas Nahdlatul Ulama Surabaya**

Website : lppm.unusa.ac.id

Email : lppm@unusa.ac.id

Hotline : 0838.5706.3867

# Bat Algorithm application for estimating Super Pairwise Alignment parameters on similarity analysis between virus protein sequences

*by Teguh Herlambang 44*

---

**Submission date:** 01-Jun-2022 12:28PM (UTC+0700)

**Submission ID:** 1848289102

**File name:** 44.\_Paper\_Jurnal\_Bat\_Algorithm\_application.pdf (940.86K)

**Word count:** 4833

**Character count:** 25415

1  
**Bat Algorithm application for estimating Super Pairwise Alignment parameters on similarity analysis between virus protein sequences**

Dinita Rahmalia<sup>1</sup>, Teguh Herlambang<sup>2</sup>

<sup>1</sup>Universitas Islam Darul Ulum, Jalan Airlangga 3 Sukodadi, Lamongan, Indonesia

<sup>2</sup>Universitas Nahdlatul Ulama Surabaya, Jalan Jemursari 59, Surabaya, Indonesia

**ARTICLE INFO**

**Article history:**

Received 14 October 2019,  
Revised 25 July 2020,  
Accepted 11 November 2020.

**Keywords:**

Meta-heuristic Method,  
Super Pairwise Alignment,  
Bioinformatics,  
Bat Algorithm,  
Similarity Analysis.

**ABSTRACT**

Either virus or bacteria could cause many diseases, and they can mutate to form a new disease. Sequence alignment is a method used in the researches of various diseases. In this research, we took a case study of the dengue virus and Zika virus. The alignment process of two virus sequences would be used to determine the similarity between the mutated virus and the original virus. Super Pairwise Alignment (SPA) is the method applied in bioinformatics for aligning two virus sequences. Because the similarity score was affected on SPA-parameters, in these research we would apply meta-heuristic method, such as Bat Algorithm (BA) algorithm for optimizing SPA-parameters which can maximize similarity score. BA with the optimization method which was resemble by the action of bats in using sonar called echolocation to detect prey, avoid obstacles. From the BA simulations, we obtain optimum SPA-parameters in approaching which can result maximum similarity score between two aligned each of dengue virus and zika virus protein sequences.

3  
This work is licensed under a Creative Commons Attribution-Share Alike 4.0



**Dinita Rahmalia,**

Universitas Islam Darul Ulum, Jalan Airlangga 3 Sukodadi, Lamongan, Indonesia

Email: [dinitarahmalia@gmail.com](mailto:dinitarahmalia@gmail.com)

**1. INTRODUCTION**

Either virus or bacteria can cause many diseases and they can mutate so that they could form the new disease. Mutations can create the growth or the death of cells, so that new disease is formed. Sequence alignment is the method so that it can be used to research various of disease whether in determining the variance, origin or epidemics development.

In this research, case study of dengue virus and zika virus will be taken as example. Zika virus has been found in 1947 in monkey and then it is identified in human in 1952 in Africa. *Aedes* mosquito can cause zika disease through bite with the symptoms are resemble to dengue fever [1]. Based on previous research, zika virus is mutation from dengue virus. In Indonesia, there are four types of dengue virus namely dengue virus-1, dengue virus-2, dengue virus-3 and dengue virus-4 [2]. In this research, each dengue virus protein sequence will be compared by zika virus protein sequence to be analyzed the similarity. For determining the similarity between mutated virus and the original virus, the alignment process of two virus sequences will be used. Super Pairwise Alignment (SPA) was the method in bioinformatics applied for aligning two virus sequences [3].

SPA-parameters selection is often applied by trial and error [4]. Because the similarity score is affected on SPA-parameters, the problems of this research are how we can estimate and optimize SPA-parameters maximizing similarity score between dengue virus and zika virus, so that the purpose of this research is applying meta-heuristic method, such as Bat Algorithm (BA) algorithm to estimate and optimize SPA-parameters which can maximize similarity score. In previous research, parameter estimations using heuristic method have been applied in forecasting [5], control problem applied in epidemics [6], and control problem applied in transportation [7]. In particular researches, SPA-parameters have also been optimized by Particle Swarm Optimization in dengue virus [8] and Artificial Bee Colony in mutation of zika virus [9]. In this research, we will optimize SPA parameters between two different disease, i.e. dengue virus and zika virus protein sequence.

2  
 Bat Algorithm (BA) was discovered by Xin She Yang in 2011. The bats use sonar called echolocation for locating in the dark, detect prey, and avoid obstacles. In the BA, there are local solution among the selected best solution based on loudness and pulse rates [10][11]. There are many researches about BA applications such as optimization problem by modified BA [12], transport network design problem [13], Travel Salesman Problem (TSP) [14], Vehicle Routing Problem (VRP) with time windows [15], flowshop scheduling problem [16], jobshop scheduling problem [17], goal programming on multiobjective function [18], image classification [19], designing fuzzy controller [20], and designing PID controller [21].

BA simulations have been applied between each dengue virus and zika virus. We obtain optimum SPA parameters in approaching which can result maximum similarity score between two aligned each of dengue virus and zika virus protein sequences. Moreover, we can obtain the position and length of sequence unit which is occurred insertions and deletions.

## 2. RESEARCH METHOD

### 2.1. Biological Sequences

There are terms in the study of biological sequences like DNA, RNA, and protein sequences. DNA (deoxyribonucleic acid) molecule is two strands of nucleotides twisted. Protein-coding genes are for producing proteins. They are linear polymers of twenty amino acids connected by peptide bond. RNA (ribonucleic acid) is single-stranded and related to DNA [22].

In DNA, RNA or protein sequence, their structure consists of the nucleotides component. Generally the following description of a biological sequence is as follows

$$A = (a_1, a_2, \dots, a_{n_a}), \bar{B} = (b_1, b_2, \dots, b_{n_b}), \bar{C} = (c_1, c_2, \dots, c_{n_c})$$

With  $\hat{A}, \hat{B}$  are the sequences, and  $a_i, b_i$  are the units of the sequence, at position  $i$

$\hat{A}, \hat{B}$  are DNA when the unit of the sequence are  $\{A, C, G, T\}$

$\hat{A}, \hat{B}$  are RNA when the unit of the sequence are  $\{A, C, G, U\}$

$\hat{A}, \hat{B}$  are protein sequences when the unit of the sequences are  $\{A, C, D, E, F, G, H, I, K, L, M, N, P, Q, R, S, T, V, W, Y\}$

### 2.2. Sequence Alignment

If the sequence has relatedness to the its pair, then two sequences are homologous. The step for inferring homology is determining the sequence similarity. To determine the similarity, we have to align them properly. For example, the two nucleotides sequences, X and Y

X : TACCAGT

Y : CCCGTAA

If there are gaps, there are many possible alignment either case 1 or case 2 so that we must choose better alignment.

Case 1 :

X : TACCAGT --

Y : C -- CC -- GTAA

Case 2 :

X : TACCAGT --

Y : -- CCCGTAA

### 2.3. The Type of Mutation

In molecular biology, the mutation of sequence A will change sequence A into sequence B. There are four types of the mutations [3]:

Type I : The mutation by segment changing to another.

Type II : The mutation by segment permuting its position.

Type III : The mutation by inserting new segment into sequence.

Type IV : The mutation by deleting the segment from sequence.

In this research, we restrict the problem only using type III and type IV mutations.

### 2.4. Comparison Two Sequences

The sequence alignment problem is finding the structure so that the sequences have difference or similarity. It is noted in penalty matrix W whose the elements of penalty matrix can be seen in equation (1).

$$w(a_i, b_i) = \begin{cases} 0, & \text{if } a_i = b_i \\ 1, & \text{otherwise} \end{cases}, i = 1, 2, \dots, n \quad (1)$$

Let two sequences  $(A = (a_1, a_2, \dots, a_n))$  and  $(B = (b_1, b_2, \dots, b_n))$ . Sliding Window Function of sequence A and sequence B can be seen in equation (2) with  $i$  the sequence unit initial number of sequence A and  $j$  is the sequence unit initial number of sequence B, and  $n$  is the length of sequence.

$$w(A, B, i, j, n) = \frac{1}{n} \sum_{k=1}^n w(a_{i+k}, b_{j+k}) \quad (2)$$

### 2.5. Shifting Mutation Modulus Structure

If sequence A mutate into sequence B by shifting mutations, then it can be constructed as in equation (3).

$$A = (a_1, \dots, a_{n_a}) \xrightarrow{III} C = (c_1, \dots, c_{n_c}) \xrightarrow{IV} B = (b_1, \dots, b_{n_b}) \quad (3)$$

By shifting mutation  $T$ , and  $A \rightarrow B$ , the mutation operations applied is as follows:

$$T = \{(i_k, l_k), k = 1, 2, \dots, k_a\} \quad (4)$$

The  $k_a$  in  $T$  is the shifting mutations total of sequence A. The pair  $(i_k, l_k)$  consist of  $i_k$  is the shifting mutation position, and  $l_k$  is the the type and length of the  $k$ -th shifting mutation. If  $l_k > 0$ , it is the type III mutation and if  $l_k < 0$ , it is the type IV mutation.  $|l_k|$  is the length of the deleted or inserted segment [3].

### 2.6. Super Pairwise Alignment (SPA) Algorithm

SPA algorithm is used for computing similarity between two sequences. Let  $(A, B)$  are two sequences with sequence A mutates into sequence B. SPA works by partiting the sequence and analyzing the similarity on the partited sequence respectively. The parameters of SPA are  $(n, \theta, \phi)$  with interval are  $n > 0$ ;  $0.3 \leq \theta \leq 0.4$ ;  $\theta < \phi \leq 1$  [3]. The parameters of SPA consist of the length of partited sequence unit  $n$ , the threshold where two partited sequences have almost been similar  $\theta$ , the threshold where two partited sequences have not been similar yet  $\phi$ , so that they may be mutated and require shifting process either insertion or deletion.

#### 1. Estimate the first mutation position $\hat{i}_1$ in $T$

Initialization  $i = j = 0$  and compute  $(w(A, B, i, j, n))$  based on equation (2).

a. If  $w(A, B, i, j, n) \geq \phi$

$$\hat{i}_1 = 0$$

and

b. If  $w(A, B, i, j, n) < \phi$

Let  $i = j = (k_1(n - \tau))$  with  $k_1$  is the integer.

Compute  $(w(A, B, i, j, n))$ .

If  $w(A, B, i, j, n) < \phi$ , do the steps 1(b) by updating  $k \leftarrow k + 1$  until  $w(A, B, i, j, n) \geq \phi$ ,

and store the value of  $i$  and  $j$  as  $\hat{i}_1$

End

#### 2. Based on the estimation $\hat{i}_1$ , estimate $\hat{l}_1$ using this procedure

$$w(A, B, \hat{i}_1 + l, \hat{i}_1 + l, n), w(A, B, \hat{i}_1 - l, \hat{i}_1 - l, n), l = 1, 2, 3, \dots, l_{\max}$$

a. If  $w(A, B, \hat{i}_1 + l, \hat{i}_1 + l, n) \leq \theta$ , then  $\hat{l}_1 = -l$ . This means occuring deletion  $l$  segments from sequence A to sequence B.

b. If  $w(A, B, \hat{l}_i, \hat{l}_i + l, n) \leq \theta$ , then  $\hat{l}_i = l$ . This means occurring insertion  $l$  segments from sequence A to sequence B.

Then, local mutation mode  $T_1 = \{(i_1, l_1)\}$ , and locally uniform alignment  $(C_1, D_1)$  are formed.

3. Repeat its process until  $k_0$

4. Compute objective :

Similarity score between whole aligned sequence  $C_{k_0}$  and sequence  $D_{k_0}$  in equation (5)

$$s(C_{k_0}, D_{k_0}) = \frac{1}{N} \sum_{j=1}^N s(c_j, d_j) \quad (5)$$

whose the element

$$s(c_j, d_j) = \begin{cases} 1, & \text{if } c_j = d_j \\ 0, & \text{otherwise} \end{cases}$$

## 2.7. Bat Algorithm

Bat Algorithm (BA) was discovered by Xin She Yang in 2010. The bats use sonar named echolocation for locating in the dark, detect prey, and avoid obstacles. The action of bats are as follows [10],[11]:

1. Echolocation is applied by bats for sensing the food, distances, barriers, and prey.
2. Bat flies randomly in position  $x_i$  with velocity  $v_i$  by various wavelength, fixed frequency  $f_{min}$ , and loudness  $A^0$  for searching the prey. They adapt the pulse rate.
3. The loudness changes from the large positive  $A^0$  to the small  $A_{min}$

Based on action of bats, the BA applied to optimization model of SPA can be constructed as follows :

Suppose objective function  $f : R^D \rightarrow R$  with  $D$  is the search space dimension.

In initialization, build the initial solutions randomly  $x_{ij}, i=1..maxpop, j=1..D$  and compute the fitness

$f(x_{ij}), i=1..maxpop, j=1..D$ . In BA applied to optimization model of SPA, let  $X = (n, \theta, \phi)$  as decision variables with objective in equation (5) for maximum similarity score.

1. Initialize the population of bat position  $x_i, i=1, 2, \dots, maxpop$ , velocity  $v_i, i=1, 2, \dots, maxpop$ , and pulse frequency  $f_i, i=1, 2, \dots, maxpop$

Initialize loudness  $A_i$  and pulse rates  $r_i$

3. Do procedure as follows :

for  $t = 1 : t_{max}$

for  $i = 1 : maxpop$

$$f_i = f_{min} + (f_{max} - f_{min})\beta, \beta \sim U(0,1) \quad (6)$$

$$v_i^t = v_i^{t-1} + (x_i^t - x^*)f_i \quad (7)$$

$$x_i^t = x_i^{t-1} + v_i^t \quad (8)$$

end

if ( $rand > r_i$ )

- Generate local solution among the selected best solution

$$x_{new} = x_{old} + \epsilon A^t, \text{ with } \epsilon \sim U(-1,1) \quad (9)$$

end

if ( $rand > A_i$  &  $f(x_i) > f(x^*)$ )

- Keep new solution  $x_i$
- Reduce  $A_i$  and increase  $r_i$  and so that  $A_i^t \rightarrow 0, r_i^t \rightarrow r_i^0$

$$r_i^{t+1} = r_i^0 (1 - \exp(-\gamma t)), \text{ with } \gamma > 0 \quad (10)$$

$$A_j^{t+1} = \alpha A_j^t, \text{ with } 0 < \alpha < 1 \tag{11}$$

```

end
Find current best  $x^*$ 
end
    
```

### 3. RESULTS AND DISCUSSION

Based on previous research [1], zika virus is mutation from dengue virus. In Indonesia, there are four types of dengue virus namely dengue virus 1, dengue virus 2, dengue virus 3 and dengue virus 4. In this research, each dengue virus protein sequence will be compared by zika virus protein sequence to be analyzed the similarity. The source of data are from National Center for Biotechnology Information (NCBI) taken on May 2, 2019. The characteristics of dengue virus and zika virus protein data can be seen in Table 1, where each virus type has unique access code with length is the number of amino acids of virus protein sequences.

**Table 1.** Dengue virus and zika virus protein data.

Access Code	Virus Type	Length (bp)
AHG06327	Dengue virus-1	3392
AHG06364	Dengue virus-2	3391
AHG06376	Dengue virus-3	3390
AHG06382	Dengue virus-4	3387
AMK49492	Zika virus	3429

The samples of dengue virus protein sequences which have been obtained from NCBI can be seen in Figure 1(a-d) while zika virus protein sequences can be seen in Figure 2. In protein sequences, there are protein-coding genes. They are linear polymers of twenty amino acids connected by peptide bond [22], with amino acid codes can be seen in Table 2.

**Table 2.** Amino acid of protein sequences.

Code	Amino Acid
'A'	"Alanine"
'C'	"Cysteine"
'D'	"Aspartic acid"
'E'	"Glutamic acid"
'F'	"Phenylalanine"
'G'	"Glycine"
'H'	"Histidine"
'I'	"Isoleucine"
'K'	"Lysine"
'L'	"Leucine"
'M'	"Methionine"
'N'	"Asparagine"
'P'	"Proline"
'Q'	"Glutamine"
'R'	"Arginine"
'S'	"Serine"
'T'	"Threonine"
'V'	"Valine"
'W'	"Tryptophan"
'Y'	"Tyrosine"



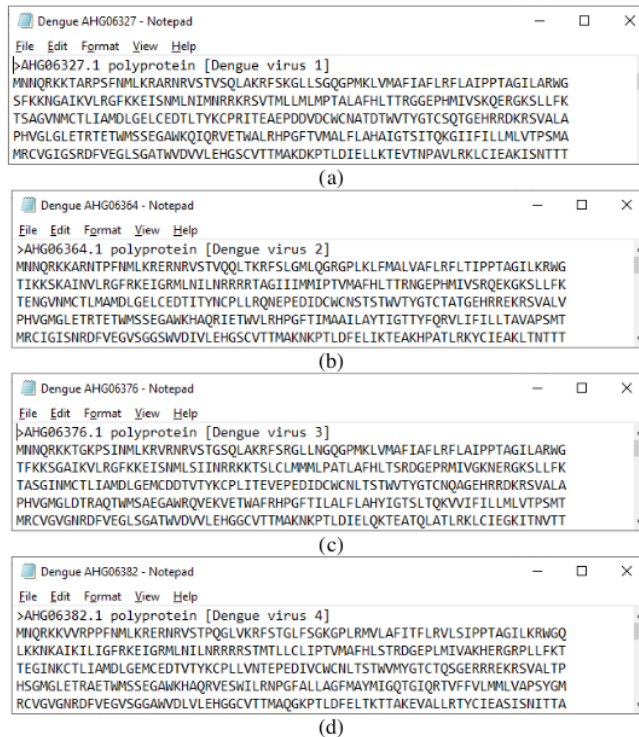


Fig. 1. The sample of dengue virus protein sequence (a) Dengue virus-1 (b) Dengue virus-2 (c) Dengue virus-3 (d) Dengue virus-4



Fig. 2. The sample of zika virus protein sequence

Bat Algorithm (BA) simulation is for finding SPA-parameters ( $n, \theta, \phi$ ) which result maximum similarity score (excluding gaps) between two aligned dengue virus and zika virus protein sequences as objective. BA parameters used in each simulation are :

- The number of population : 10
- The maximum iteration : 50
- The fixed Frequency :  $f_{\min} = 0$   $f_{\max} = 1$
- The loudness :  $A_0 = 10$   $A_{\min} = 0$
- The rate of pulse emission :  $r_0 \in (0,1)$

Simulations are run by Matlab in Asus Laptop, CPU Intel 2 Core N3350, 2.4 GHz, Memory (RAM) 4 GB, HDD 500 GB, OS Windows 10.

### 3.1. Similarity Analysis Between Dengue Virus 1 and Zika Virus

Figure 3(a) shows BA process of similarity score maximization of aligned dengue virus 1 and zika virus protein sequence. In the initial, bats fly randomly with the position as SPA-parameters with varying wavelength, a fixed frequency, and loudness for searching prey. In the optimization, the loudness decreases



and the pulse rate increases so that bat can find SPA-parameters with maximum score. The optimum SPA-parameters are  $((n, \theta, \phi) = (13; 0.3973; 0.5126))$  with maximum similarity score based on equation (5) is 0.5121 or 51.21%. Because the similarity score is more than 25 %, we can declare that sequences are homologous (have similarity each other) [3].

Figure 3(b) shows the positions  $i_k$  and lengths  $l_k$  of optimum SPA-parameters. Optimum SPA-parameters result iteration  $k = 123$  when partition processes of 5 hole sequences finish. From the graphs, positions  $i_k$  and lengths  $l_k$  can be seen. If  $l_k > 0$ , insertion occurs (type III mutation), if  $l_k < 0$ , deletion occurs (type IV mutation), and if  $l_k = 0$ , mutation shifting is nothing.

As example, in iteration  $k=14$  in Figure 3(b), there are insertion  $l_k$  by 8 segments into zika virus protein sequence from dengue virus 1 in position 170. In iteration  $k=86$ , there are deletion  $l_k$  by 2 segments into zika virus protein sequence from dengue virus 1 in position 1936. When in Figure 3(b) does not show insertion or deletion, then there is no mutation shifting in the corresponding position  $i_k$ .

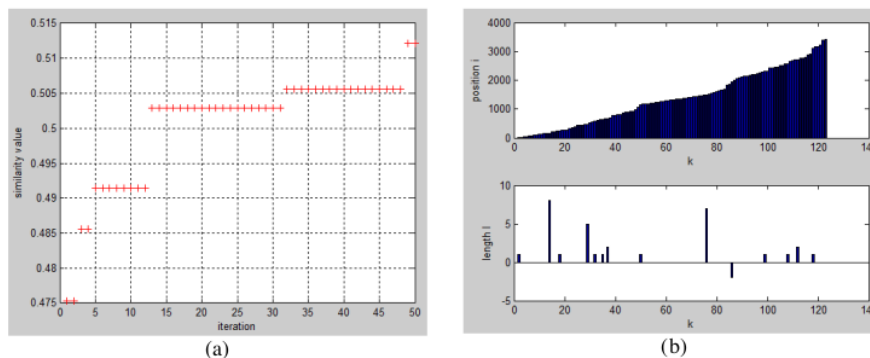


Fig. 3. Optimization process of aligned dengue virus 1 and zika virus protein sequences (a) BA optimization process (b) Lengths and positions of optimum SPA-parameters

### 3.2. Similarity Analysis Between Dengue Virus 2 and Zika Virus

Figure 4(a) shows BA process of similarity score maximization of aligned dengue virus 2 and zika virus protein sequence. In the initial, bats fly randomly with the position as SPA-parameters with varying wavelength, a fixed frequency, and loudness for searching prey. In the optimization, the loudness decreases and the pulse rate increases so that bat can find SPA-parameters with maximum score. The optimum SPA-parameters are  $((n, \theta, \phi) = (9; 0.3674; 0.5348))$  with maximum similarity score based on equation (5) is 0.5285 or 52.85%. Because the similarity score is more than 25 %, we can declare that sequences are homologous (have similarity each other) [3].

Figure 4(b) shows the positions  $i_k$  and lengths  $l_k$  of optimum SPA-parameters. Optimum SPA-parameters result iteration  $k = 179$  when partition processes of 5 hole sequences finish. From the graphs, positions  $i_k$  and lengths  $l_k$  can be seen. If  $l_k > 0$ , insertion occurs (type III mutation), if  $l_k < 0$ , deletion occurs (type IV mutation), and if  $l_k = 0$ , mutation shifting is nothing.

As example, in iteration  $k=20$  in Figure 4(b), there are insertion  $l_k$  by 9 segments into zika virus protein sequence from dengue virus 2 in position 177. In iteration  $k=124$ , there are deletion  $l_k$  by 2 segments into zika virus protein sequence from dengue virus 2 in position 1686. When in Figure 4(b) does not show insertion or deletion, then there is no mutation shifting in the corresponding position  $i_k$ .

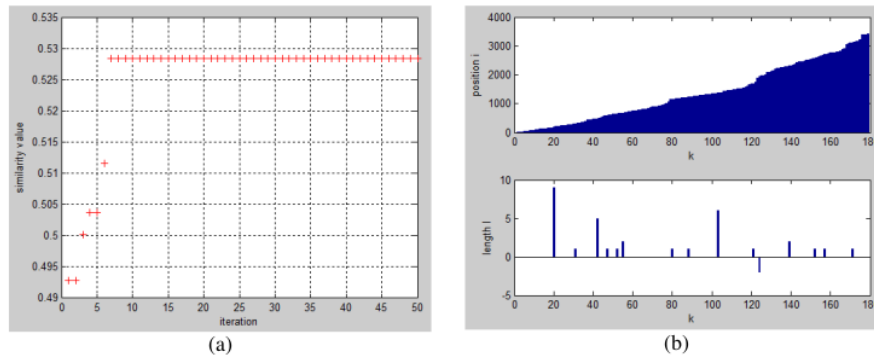


Fig. 4. Optimization process of aligned dengue virus 2 and zika virus protein sequences (a) BA optimization process (b) Lengths and positions of optimum SPA-parameters

### 3.3. Similarity Analysis Between Dengue Virus 3 and Zika Virus

Figure 5(a) shows BA process of similarity score maximization of aligned dengue virus 3 and zika virus protein sequence. In the initial, bats fly randomly with the position as SPA-parameters with varying wavelength, a fixed frequency, and loudness for searching prey. In the optimization, the loudness decreases and the pulse rate increases so that bat can find SPA-parameters with maximum score. The optimum SPA-parameters are  $((n, \theta, \phi) = (9; 0.3366; 0.5604))$  with maximum similarity score based on equation (5) is 0.5319 or 53.19%. Because the similarity score is more than 25 %, we can declare that sequences are homologous (have similarity each other) [3].

Figure 5(b) shows the positions  $i_k$  and lengths  $l_k$  of optimum SPA-parameters. Optimum SPA-parameters result iteration  $k = 139$  when partition processes of 5 hole sequences finish. From the graphs, positions  $i_k$  and lengths  $l_k$  can be seen. If  $l_k > 0$ , insertion occurs (type III mutation), if  $l_k < 0$ , deletion occurs (type IV mutation), and if  $l_k = 0$ , mutation shifting is nothing.

As example, in iteration  $k=14$  in Figure 5(b), there are insertion  $l_k$  by 8 segments into zika virus protein sequence from dengue virus 3 in position 155. In iteration  $k=99$  in, there are deletion  $l_k$  by 2 segments into zika virus protein sequence from dengue virus 3 in position 1939. When in Figure 5(b) does not show insertion or deletion, then there is no mutation shifting in the corresponding position  $i_k$ .

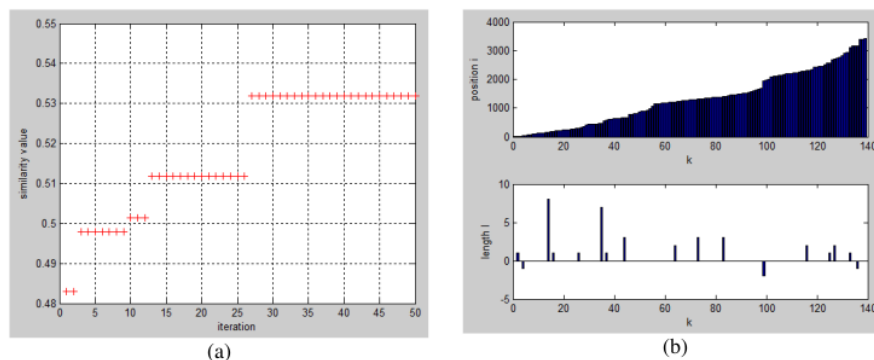


Fig. 5. Optimization process of aligned dengue virus 3 and zika virus protein sequences (a) BA optimization process (b) Lengths and positions of optimum SPA-parameters

### 3.4. Similarity Analysis Between Dengue Virus 4 and Zika Virus

Figure 6(a) shows BA process of similarity score maximization of aligned dengue virus 4 and zika virus protein sequence. In the initial, bats fly randomly with the position as SPA-parameters with varying

wavelength, a fixed frequency, and loudness for searching prey. In the optimization, the loudness decreases and the pulse rate increases so that bat can find SPA-parameters with maximum score. The optimum SPA-parameters are  $((n, \theta, \phi) = (8; 0.3889; 0.6267))$  with maximum similarity score based on equation (5) is 0.5338 or 53.38%. Because the similarity score is more than 25 %, we can declare that sequences are homologous (have similarity each other) [3].

Figure 6(b) shows the positions  $i_k$  and lengths  $l_k$  of optimum SPA-parameters. Optimum SPA-parameters result iteration  $k = 116$  when partition processes of 5 hole sequences finish. From the graphs, positions  $i_k$  and lengths  $l_k$  can be seen. If  $l_k > 0$ , insertion occurs (type III mutation), if  $l_k < 0$ , deletion occurs (type IV mutation), and if  $l_k = 0$ , mutation shifting is nothing.

As example, in iteration  $k=14$  in Figure 6(b), there are insertion  $l_k$  by 8 segments into zika virus protein sequence from dengue virus 4 in position 152. In iteration  $k=84$ , there are deletion  $l_k$  by 2 segments into zika virus protein sequence from dengue virus 4 in position 1942 (see Figure 6(b)). When in Figure 6(b) does not show insertion or deletion, then there is no mutation shifting in the corresponding position  $i_k$ .

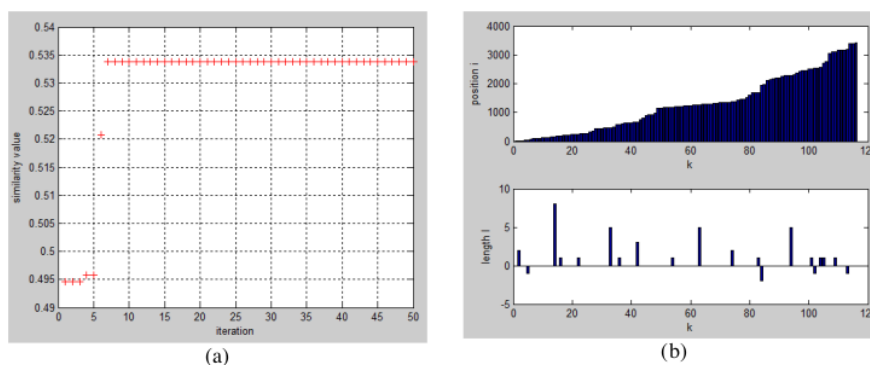


Fig. 6. Optimization process of aligned dengue virus 4 and zika virus protein sequences (a) BA optimization process (b) Lengths and positions of optimum SPA-parameters

#### 4. CONCLUSION

Dengue virus protein sequence is the mutation from zika virus protein sequence, so that they are homologous sequence and they have similarity. For determining the similarity between mutated virus and the original virus, the alignment process of two virus sequences will be used. Super Pairwise Alignment (SPA) is the method in bioinformatics applied for aligning two virus sequences. In SPA, the similarity score is affected on SPA parameters. The SPA-parameters can be optimized by meta-heuristic method, such as Bat Algorithm (BA) from the action of bats. In the case study of dengue virus and zika virus protein, BA can find optimum SPA-parameters in approaching which can result maximum similarity score. Moreover, we can obtain the position and length of sequence unit which is occurred insertions and deletions. The development of this research is optimization the similarity of more than two sequences (multiple sequences).

#### Acknowledgments

Appreciation to the Kemenristekdikti, for the fund for this research conducted in 2019.

#### REFERENCES

- [1] M.S. Pradana, S. Amiroch, "Protein Sequence Analysis of the Zika Virus and Dengue Virus Using Smith Waterman Algorithm," *AIP Conference Proceeding*, Vol 2084, 2019. DOI: <https://doi.org/10.1063/1.5094275>
- [2] K.C. Mulyatno, A. Yamanaka, S. Yotopranoto, E. Konishi, "Vertical Transmission of Dengue Virus in Aedes aegypti Collected in Surabaya, Indonesia, during 2008-2011," *Jpn. J. Infect. Dis.*, vol. 65, pp. 274-276, 2012. DOI: <https://doi.org/10.7883/yoken.65.274>
- [3] S.N. Shen, J.A. Tuszynski, *Theory and Mathematical Method for Bioinformatics*. Berlin: Springer, 2008. DOI: <https://doi.org/10.1007/978-3-540-74891-5>

- [4] M.L. Shahab, M.I. Irawan, "Sequence Alignment Using Nature-Inspired Metaheuristic Algorithm," *International Journal of Computing Science and Applied Mathematics* vol. 3, pp. 27-31, 2017. DOI: <https://doi.org/10.12962/j24775401.v3i1.2118>
- [5] D. Rahmalia, "Estimation of Exponential Smoothing Parameter on Pesticide Characteristic Forecast Using Ant Colony Optimization (ACO)," *Eksakta: Jurnal Ilmu-ilmu MIPA*, vol. 18, pp. 56-63, 2018. DOI: <https://doi.org/10.20885/eksakta.vol18.iss1.art6>
- [6] D. Rahmalia, T. Herlambang, "Weight Optimization of Optimal Control Influenza Model Using Artificial Bee Colony," *International Journal of Computing Science and Applied Mathematics*, vol. 4, pp. 27-31, 2018. DOI: <https://doi.org/10.12962/j24775401.v4i1.2997>
- [7] T. Herlambang, D. Rahmalia, T. Yulianto, "Particle Swarm Optimization (PSO) and Ant Colony Optimization (ACO) for Optimizing PID Parameters on Autonomous Underwater Vehicle (AUV) Control System," *Journal of Physics : Conference Series*, Vol 1211, 2019. DOI: <https://doi.org/10.1088/1742-6596/1211/1/012039>
- [8] D. Rahmalia, A. Rohmatullah, M.S. Pradana, "Estimasi Parameter Super Pairwise Alignment pada Kombinasi Virus Dengue Menggunakan Particle Swarm Optimization," *TechnoCom*, vol. 18, no. 3, pp. 264-274, 2019. DOI: <https://doi.org/10.33633/tc.v18i3.2528>
- [9] D. Rahmalia, A. Rohmatullah, M.S. Pradana, "Estimation of Super Pairwise Alignment (SPA) Parameters on Zika Virus Mutation Using Artificial Bee Colony," *Journal of Physics : Conference Series*, Vol 1417, 2019. DOI: <https://doi.org/10.1088/1742-6596/1417/1/012029>
- [10] X.S. Yang, Z. Chui, R. Xiao, *Swarm Intelligent and Bio-Inspired Computation*. London: Elsevier, 2013. DOI: <https://doi.org/10.1016/B978-0-12-405163-8.00001-6>
- [11] X.S. Yang, *A New Metaheuristic Bat-Inspired Algorithm in Nature Inspired Cooperative Strategies for Optimization (NISCO 2010)*. Berlin: Springer, 2010. DOI: [https://doi.org/10.1007/978-3-642-12538-6\\_6](https://doi.org/10.1007/978-3-642-12538-6_6)
- [12] S. Yilmaz, E.U. Kucuksille, "A New Modification Approach on Bat Algorithm for Solving Optimization Problem," *Applied Soft Computing*, vol. 28, pp. 259-275, 2015. DOI: <https://doi.org/10.1016/j.asoc.2014.11.029>
- [13] S. Srivastava, S.K. Sahana, "Application of Bat Algorithm for Transport Network Design Problem," *Applied Computational Intelligence and Soft Computing*, vol. 2019, pp. 1-12, 2019. DOI: <https://doi.org/10.1155/2019/9864090>
- [14] E. Osaba, X.S. Yang, F. Diaz, P.L. Garcia, R. Carballedo, "An Improved Discrete Bat Algorithm for Symmetric and Asymmetric Travel Salesman Problem," *Engineering Application of Artificial Intelligence*, vol. 48, no. 1, pp. 59-71, 2016. DOI: <https://doi.org/10.1016/j.engappai.2015.10.006>
- [15] E. Osaba, R. Carballedo, X.S. Yang, "On Efficiently Solving the Vehicle Routing Problem with Time Windows Using Bat Algorithm with Random Reinsertion Operators," *Natural-Inspired Algorithm and Applied Optimization*. Springer, pp. 69-89, 2017. DOI: [https://doi.org/10.1007/978-3-319-67669-2\\_4](https://doi.org/10.1007/978-3-319-67669-2_4)
- [16] N. Kaur, S. Singh, "A Budget-Constrained Time and Reability Optimization Bat Algorithm for Scheduling Workflow Application in Clouds," *Procedia Computer Science*, vol. 98, pp. 199-204, 2016. DOI: <https://doi.org/10.1016/j.procs.2016.09.032>
- [17] X. Chen, B. Zhang, D. Gao, "An Improved Bat Algorithm for Job Shop Scheduling Problem," *IEEE International Conference on Mechantronics and Automation (ICMA)*, 2019. DOI: <https://doi.org/10.1109/ICMA.2019.8816578>
- [18] D. Rahmalia, N.E. Chandra, S.A. Rohmaniah, L. Muzdalifah, "Goal Programming on Optimal Pairings Selection from Flight Schedule Using Bat Algorithm," *Journal of Physics : Conference Series*, Vol 1940, 2020. DOI: <https://doi.org/10.1088/1742-6596/1490/1/012036>
- [19] Z. Ye, X. Hou, X. Zhang, J. Yang, "Application of Bat Algorithm for Texture Image Classification," *International Journal Intelligent Systems and Application*, vol. 5, pp. 42-50, 2018. DOI: <https://doi.org/10.5815/ijisa.2018.05.05>
- [20] N. Talbi, "Design of Fuzzy Controller Rule Base Using Bat Algorithm," *Energy Procedia*, vol. 162, pp. 241-250, 2019. DOI: <https://doi.org/10.1016/j.egypro.2019.04.026>
- [21] D. Pebrianti, L. Bayuaji, Y. Arumgam, "PID Controller Design for Mobile Robot Using Bat Algorithm with Mutation (BAM)," *6<sup>th</sup> International Conference on Electrical Engineering, Computer Science, and Informatics (EECSI)*, 2019. DOI: <https://doi.org/10.23919/EECSI48112.2019.8976932>
- [22] A. Isaev, *Introduction to Mathematical Methods in Bioinformatics*. Berlin: Springer, 2004.

## BIOGRAPHY OF AUTHORS

**Dinita Rahmalia**, Lecturer of Mathematics Department. Universitas Islam Darul Ulum Lamongan. The research fields are optimization and operations research, control system, and data mining  
Email : [dinitarahmalia@gmail.com](mailto:dinitarahmalia@gmail.com)

**Teguh Herlambang**, Lecturer of Information System Department. Universitas Nahdlatul Ulama Surabaya. The research fields are applied mathematics, modelling, navigation and guidance  
Email : [teguh@unusa.ac.id](mailto:teguh@unusa.ac.id)

# Bat Algorithm application for estimating Super Pairwise Alignment parameters on similarity analysis between virus protein sequences

## ORIGINALITY REPORT

17%

SIMILARITY INDEX

9%

INTERNET SOURCES

15%

PUBLICATIONS

4%

STUDENT PAPERS

## PRIMARY SOURCES

1	<a href="http://www.sciencegate.app">www.sciencegate.app</a> Internet Source	3%
2	D Rahmalia, N E Chandra, S A Rohmaniah, L Muzdalifah, E F Kurniawati. "Goal programming on optimal pairings selection from flight schedule using Bat Algorithm", <i>Journal of Physics: Conference Series</i> , 2020 Publication	3%
3	<a href="http://pdfs.semanticscholar.org">pdfs.semanticscholar.org</a> Internet Source	3%
4	"Super Pairwise Alignment", <i>Biological and Medical Physics Biomedical Engineering</i> , 2008 Publication	3%
5	"Chapter 4 Modulus Structure Theory", Springer Science and Business Media LLC, 2008 Publication	2%
6	"Introduction", <i>Biological and Medical Physics Biomedical Engineering</i> , 2008	2%

7

Submitted to University of Malaya  
Student Paper

2%

---

Exclude quotes On

Exclude matches < 2%

Exclude bibliography On