

Analisis *Multiple Alignment* pada Penyebaran Covid-19 Varian E.G.5 dengan Metode *Maximum Likelihood*

Nelly Alviana

Program Studi Matematika, FMIPA, Universitas Islam Darul ‘Ulum Lamongan
email: nelly.2020@mhs.unisda.ac.id

Abstract. One method that can be used to construct a phylogenetic tree is the maximum likelihood method. This study was conducted to determine how to identify the spread of the covid virus variant E.G.5 using the maximum likelihood method. The results of the mutation area network analysis show that the spread of the Covid-19 E.G.5 epidemic originated from South Jakarta (EPI_ISL_17664319) and from the phylogenetic tree analysis with maximum likelihood it is known that there are 16 clades with 6 large clades in the covid-19 variant E.G.5 virus spread cluster.

Keyword: *maximum likelihood method, phylogenetics tree, Covid-19 var E.G.5*

Abstrak. Salah satu metode yang dapat digunakan untuk mengkonstruksi pohon filogenetik adalah *maximum likelihood method*. Penelitian ini dilakukan untuk mengetahui bagaimana mengidentifikasi suatu penyebaran virus covid varian E.G.5 menggunakan metode *maximum likelihood*. Hasil analisa jaringan wilayah mutasi menunjukkan bahwa penyebaran epidemi Covid-19 E.G.5 berasal dari Jakarta Selatan (EPI_ISL_17664319) dan dari analisis pohon filogenetik dengan *maximum likelihood* diketahui ada 16 klad dengan 6 klad besar pada kluster penyebaran virus covid-19 varian E.G.5.

Kata kunci : metode *maximum likelihood*, pohon filogenetik, Covid-19 varian E.G.5.

1 Pendahuluan

Filogenetik molekuler merupakan cabang ilmu yang sedang sangat berkembang, digunakan pada hampir seluruh cabang ilmu biologi. Filogenetik molekuler merupakan teknik yang mengkombinasikan metode molekuler dan statistik dalam menentukan hubungan kekerabatan secara evolusi antar organisme atau gen menggunakan struktur dan fungsi molekul beserta informasi perubahannya terhadap waktu. Kemajuan teknologi dan algoritma-algoritma statistik yang telah diciptakan membuat proses *sequencing* genom menjadi lebih cepat, murah, dan efektif. Dengan banyaknya data genom yang dipublikasikan, hal ini membuat filogenetik molekuler terus berkembang dan memberikan banyak aplikasi. Tujuan utama analisis filogenetik molekuler adalah menganalisis adanya proses evolusi dan

menyajikannya dalam bentuk pohon filogenetik yang secara grafis menunjukkan kekerabatan antar spesies atau gen terhadap waktu [1].

Pohon filogenetik dikenal pula dengan istilah filogeni. Filogeni merupakan diagram berupa pohon yang menunjukkan garis evolusi dari spesies, organisme, atau gen berbeda dari suatu nenek moyang bersama. Filogeni sangat bermanfaat dalam mengetahui diversitas biologis, menyusun klasifikasi, dan menjelaskan fenomena yang terjadi selama proses evolusi [2]. Hubungan kekerabatan antar spesies atau gen dapat dijelaskan dengan filogeni yang merupakan hasil analisis filogenetik molekuler. Oleh karena itu, filogenetik molekuler digunakan sebagai pendekatan untuk membandingkan susunan gen atau genom antar organisme, dalam percobaan ini adalah bakteri yang belum teridentifikasi. Informasi yang tersedia merupakan *sequences* nukleotida gen hasil *sequence* DNA dari suatu sampel bakteri. Salah satu metode yang digunakan dalam menentukan pohon filogenetik adalah metode *maximum likelihood* [3].

Covid-19 varian E.G.5 adalah salah satu contoh kasus yang dapat diidentifikasi dengan pohon filogenetik ini, dimana pola penyebaran virus corona ini diperkirakan menjadi pohon filogenetik yang menghubungkan semua virus covid varian terbaru satu sama lain. Virus ini bermula muncul di China dan beberapa penelitian menemukan adanya varian terbaru yang tersebar di beberapa negara termasuk Indonesia. Salah satu penelitian mengenai analisis filogenetik pada virus covid varian omicron dan delta yang menunjukkan hasil perbedaan genetik yang cukup signifikan [4]. Selain itu, ada penelitian yang menganalisis analisis mengenai mutasi varian dan rekonstruksi filogenetik menggunakan berbasis jarak dan algoritma neighbour joining yang disimulasikan di Phylip dan untuk multiple alignment menggunakan CLustalW [5], [6]. Dan ada penelitian lain yang dilakukan oleh [7] mengenai analisis virus SARS menggunakan metode *maximum likelihood*. Pada penelitian tersebut menunjukkan bahwa nilai *maximum likelihood* adalah 269135,56466 yang dapat digunakan untuk menentukan pohon dengan penyelarasan urutan dengan lebih baik.

Sehubungan dengan penelitian yang telah dilakukan diatas, penulis melakukan penelitian yang menjelaskan cara mengidentifikasi penyebaran covid-19 varian E.G.5 melalui pembentukan pohon filogenetik menggunakan metode *maximum likelihood* serta analisisnya.

2 Tinjauan Pustaka

2.1 Sequence DNA

Sequence DNA, deoxyribonucleic acid, (selanjutnya hanya akan disebut dengan *sequence*) adalah barisan dari empat macam *nucleotide*, yaitu adenine (A),

cytosine (C), guanine (G), dan thymine (T) [8]. Sebagai contoh, TGGTCTCGATAGTACTTGCA adalah *sequence* yang terdiri dari dua puluh *nucleotide*.

Dari *sequence* yang sudah ada, dapat terbentuk *sequence* baru apabila terjadi suatu mutasi pada *nucleotide-nucleotide* dari *sequence* tersebut. Secara umum, terdapat empat tipe mutasi yang bisa terjadi yaitu [3]:

1. Mutasi tipe 1, adalah mutasi yang disebabkan suatu *nucleotide* berubah menjadi *nucleotide* lainnya, misalnya A pada TGGTTCAGTA berubah menjadi G sehingga didapat TGGTTCGGTA.
2. Mutasi tipe 2, adalah mutasi yang disebabkan suatu *nucleotide* bertukar posisi dengan *nucleotide* didepan atau dibelakangnya, misalnya A bertukar posisi dengan C pada *sequence* TGGTTCAGTA sehingga didapat TGGTACGTA.
3. Mutasi tipe 3, adalah mutasi yang disebabkan adanya suatu *nucleotide* yang disisipkan pada *sequence*, misalnya G disisipkan pada TGGTTCAGTA sehingga didapat TGTGTCAGTA.
4. Mutasi tipe 4, adalah mutasi yang disebabkan adanya suatu *nucleotide* pada *sequence* yang dihapus, misalnya T pada TGGTTCAGTA dihapus sehingga didapat TGTCAGTA.

Selanjutnya, sembarang *sequence* dapat disimbolkan dengan s_i untuk suatu $E \in N$ dan *nucleotide* dari *sequence* adalah elemen dari $N_4 = \{A, C, G, T\}$.

2.2 Multiple Sequence Alignment

Secara umum, multiple *sequence* alignment adalah proses penjajaran n *sequence*, dimana $n \geq 3$. Perhatikan penjelasan secara matematis mengenai multiple *sequence* alignment berikut ini.

Misalkan terdapat n *sequence*, yaitu s_1, s_2, \dots, s_n , yang tersusun dari elemen-elemen dari N_4 . Kemudian disisipkan beberapa gap “-” ke dalam s_1, s_2, \dots, s_n sehingga didapatkan n *sequence*, yaitu s_1', s_2', \dots, s_n' , yang tersusun dari elemen-elemen dari $N_5 = \{A, C, G, T, -\}$. Dimana [3]:

1. *Sequence* s_1' disebut ekspansi dari s_1 apabila s_1^E diperoleh dengan menyisipkan beberapa gap “-” ke dalam s_1 .
2. *Sequence* s_1', s_2' disebut ekspansi dari s_1, s_2 apabila s_1^E, s_2^E berturut-turut adalah ekspansi dari s_1, s_2 .
3. Multiple *sequence* $S^E = \{s_1^E, s_2^E, \dots, s_n^E\}$ disebut ekspansi dari multiple *sequence* $S = \{s_1, s_2, \dots, s_n\}$ apabila setiap S^E adalah ekspansi dari s_i .
4. S disebut *sequence* asal dari S^E apabila multiple *sequence* S^E adalah ekspansi dari S . Kemudian dinotasikan *sequence* dalam S^E dengan

$$s_i' = s_{i,1}^E s_{i,2}^E \dots s_{i,l_i}^E \quad (2.1)$$

Dimana $S_{i,j} \in \mathbb{N}$.

- Multiple *sequence* S' disebut multiple *sequence* alignment dari S jika $l_1 = l_2 = \dots = l_n = L$, jika S^E adalah ekspansi dari S , dan jika terdapat salah satu dari $s_1^E, s_2^E, \dots, s_n^E$, yang bukan merupakan gap “-” untuk $1 \leq j \leq L$.

Multiple *Sequence* Alignment dilakukan dengan menggunakan program Clustal Omega [9].

2.3 Markov Chain Continuous Time

Markov Chain Continuous Time merupakan proses stokastik dari variable random $\{X(t), t \geq 0\}$ yang mempunyai sifat Markov. Dengan demikian, probabilitas transisi dari keadaan i ke keadaan j pada waktu $t \geq 0$,

$$P_{ij}(t) = P(X(t) = j | X(0) = i) \quad (2.2)$$

dengan $0 \leq P_{ij} \leq 1, \sum_j P_{ij}(t) = 1, \forall j$

Matriks probabilitas transisi milik DNA yang memiliki empat nukleotid, yaitu $A, C, G,$ dan T yang dinotasikan oleh:

$$P(t) = \begin{pmatrix} P_{AA}(t) & P_{AC}(t) & P_{AG}(t) & P_{AT}(t) \\ P_{CA}(t) & P_{CC}(t) & P_{CG}(t) & P_{CT}(t) \\ P_{GA}(t) & P_{GC}(t) & P_{GG}(t) & P_{GT}(t) \\ P_{TA}(t) & P_{TC}(t) & P_{TG}(t) & P_{TT}(t) \end{pmatrix} \quad (2.3)$$

Substitusi nukleotid dari *state* i ke *state* j merupakan rantai Markov dengan waktu kontinu. Dengan begitu, perubahan nukleotid dari *state* i ke *state* j selama $t + s$ waktu itu berarti perubahan nukleotid dari *state* i ke *state* k selama t waktu, kemudian dilanjutkan *state* k ke *state* j selama s waktu. Dengan demikian, berdasarkan persamaan Chapman Kolmogorov dapat dituliskan bahwa,

$$P_{ij}(t+s) = \sum_{k=1}^{\infty} P_{ik}(t) P_{kj}(s) \quad (2.4)$$

Karena substitusi nukleotid tersebut merupakan rantai Markov dengan waktu kontinu, dapat dikatakan sebagai rantai Markov dengan waktu kontinu regular [10].

2.4 Metode Maximum likelihood

Pada evolusi, mutasi merupakan perubahan peristiwa. Probabilitas menemukan mutasi sepanjang satu cabang pada pohon filogenetik dapat dihitung dengan menggunakan metode *maximum likelihood*. Metode tersebut digunakan untuk menentukan topologi pohon, panjang cabang, dan parameter dari model evolusi yang memaksimalkan probabilitas dari *sequence* [8]. Dengan kata lain, fungsi likelihood adalah probabilitas bersyarat dari data yang diberikan. Dengan demikian,

$$L(\tau\theta) = P(\text{Data} | \tau, \theta) \quad (2.5)$$

$$= P(\text{align sequence} | \text{pohon, model evolusi})$$

MLE dari τ dan θ yaitu τ^{\wedge} adalah θ , adalah membuat fungsi Likelihood sebisa mungkin menjadi,

$$(\tau^{\wedge}, \theta^{\wedge}) = \arg_{\tau, \theta} \max(\tau, \theta).$$

1. Metode *maximum likelihood* untuk dua *sequence*

$$L = \varphi_{x1} p_{x1y1}(t) \varphi_{x2} p_{x2y2}(t) \dots \varphi_{xn} p_{xny_n}(t) \quad (2.6)$$

Probabilitas transisi dari *sequence* satu ke *sequence* yang lain merupakan jumlah probabilitas dari semua kemungkinan jalur yang menghubungkan kedua *sequence* tersebut. Jalur khusus dari transisi satu *sequence* ke *sequence* yang lain dapat dinyatakan sebagai penyejajaran [11].

2. Kemungkinan Pohon Likelihood

Dalam rekonstruksi filogenetik perlu menemukan pohon yang dapat memaksimalkan kemungkinan global dari semua struktur pohon. Sejumlah pohon topologi biner tidak berakar meningkat dengan sejumlah taksa (n), yang dapat dihitung menurut:

$$t_n = \frac{2^n - 5!}{2^{n-3}(n-3)!} = \prod_{i=1}^n 2^n - 5 \quad (2.7)$$

Saat menghitung pohon *maximum likelihood*, parameter model dan panjang cabang harus dihitung untuk setiap pohon, kemudian pohon yang menghasilkan nilai likelihood tertinggi akan dipilih. Nilai MLE yang paling tinggi menandakan bahwa pohon tersebut dapat menjelaskan persejajaran *sequence* dengan lebih baik [12].

3. Metode Penelitian

3.1 Dataset

Data yang digunakan pada studi ini adalah 93 data DNA *sequence* dari virus covid-19 varian E.G.5 yang tersaji pada Tabel 1. Data tersebut diambil dari Genbank yaitu *database* gen terbesar di dunia milik pemerintah Amerika Serikat. Pengambilan dilakukan dengan mengakses Pusat Informasi Bioteknologi Nasional yaitu www.ncbi.nlm.nih.gov dan GISAID www.gisaid.org berupa fasta.

3.2 Multiple Alignment

Alignment bertujuan untuk menata semua *sequence* agar satu sama lain diletakkan sesuai dengan posisi homologi antar *sequence* sehingga dapat dibandingkan. File FASTA digunakan sebagai data input. *Alignment* total (*do*

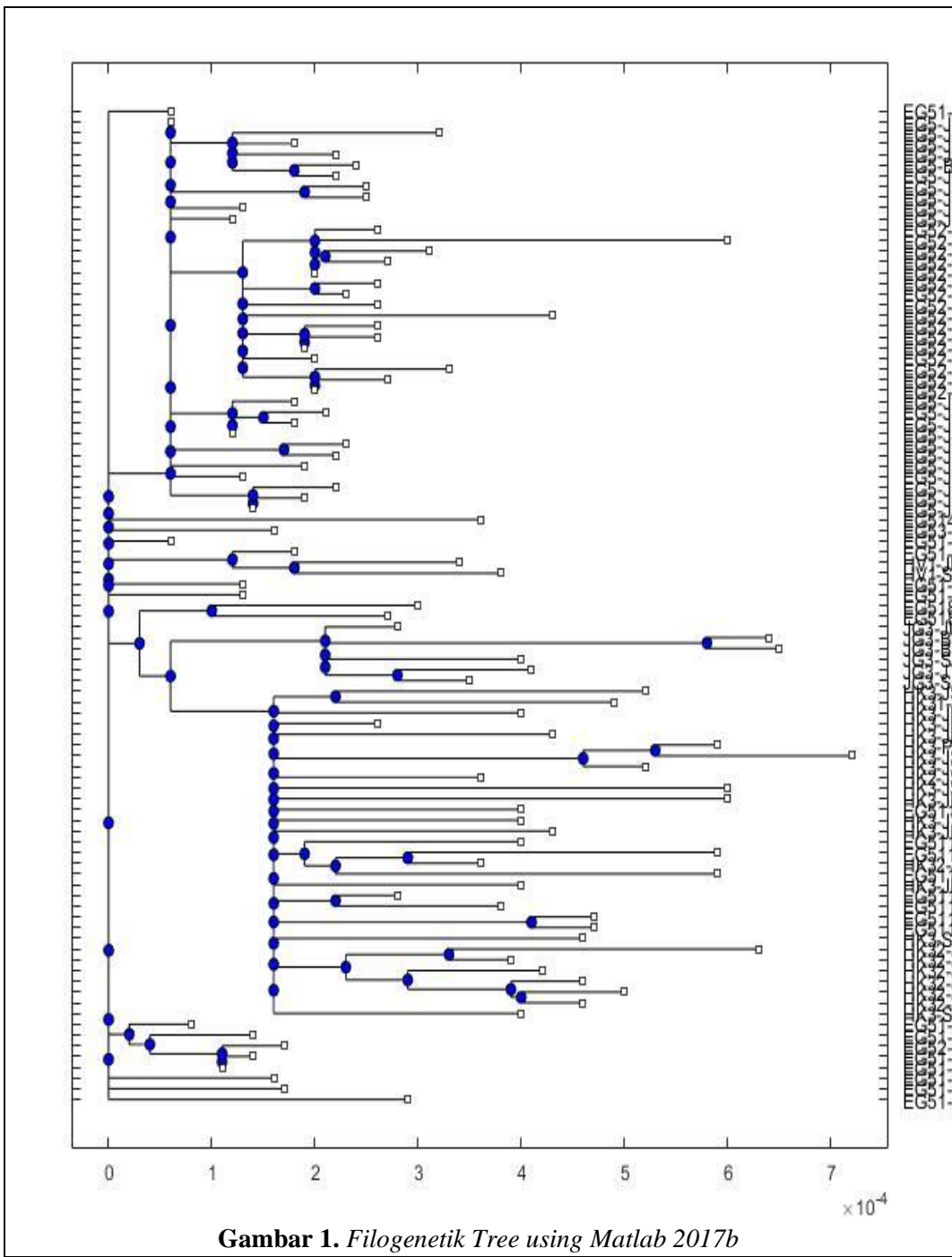
complete alignment) dilakukan agar diperoleh file yang kompatibel untuk program Phylip.

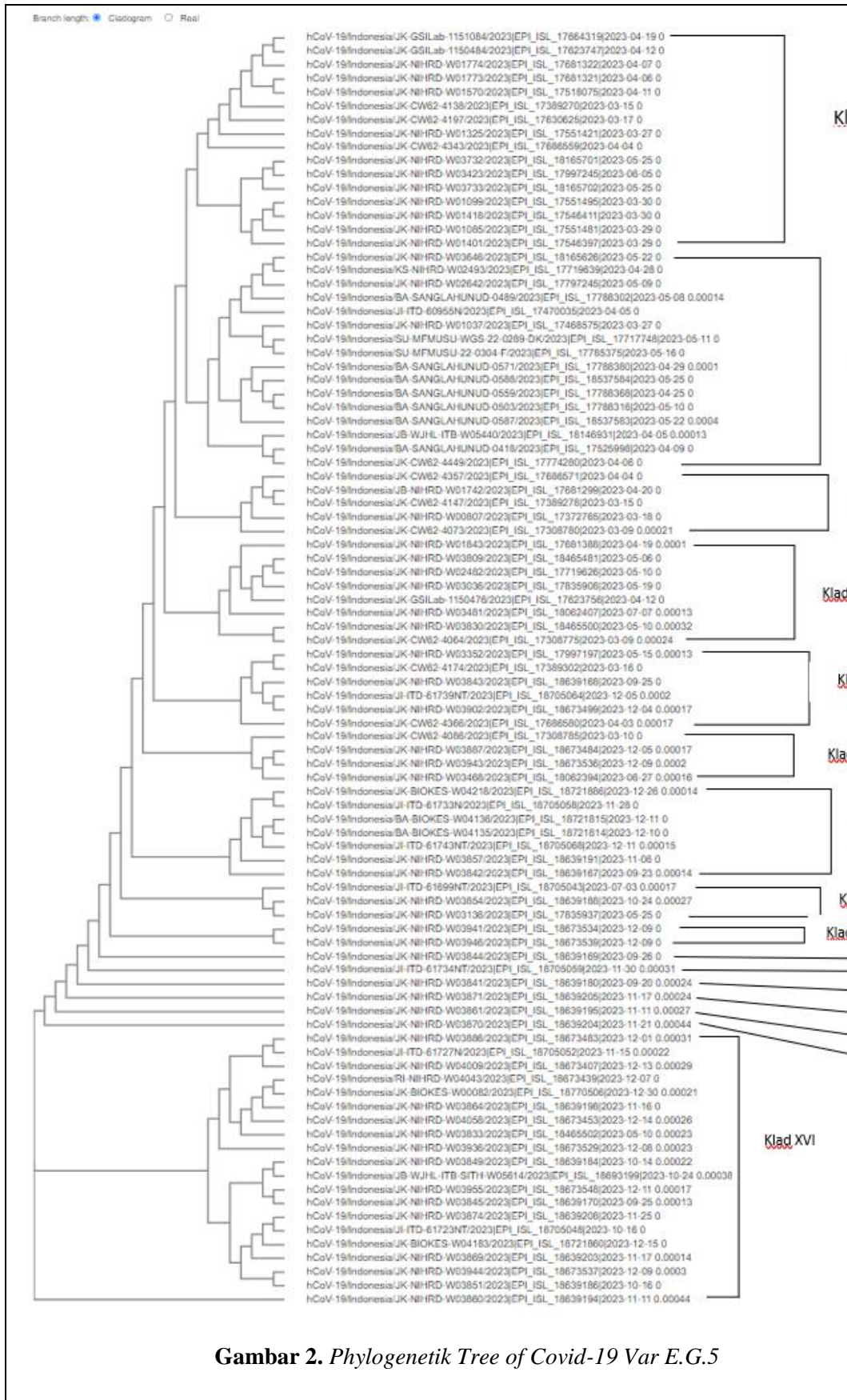
Untuk data *sequence* pasien yang terjangkit virus covid-19 varian E.G.5 yang diambil di file FASTA akan dilakukan proses *alignment* di ClustalX 2.1 dan Clustal Omega yang kemudian dilanjutkan dengan analisis virus menggunakan metode *maximum likelihood* menggunakan Phylip 3.698, sehingga diperoleh pohon genetik dan analisis jaringan topologi, daerah mutasi, serta analisis sistem jaringan mode mutasi.

4. Hasil dan Pembahasan

4.1 Pembentukan Pohon Filogenetik dengan Metode *Maximum likelihood*

Topologi pohon filogenetik yang dihasilkan bersifat monofiletik yang berarti bahwa kelompok virus covid-19 var E.G.5 yang menyebar di Indonesia tersebut memiliki satu leluhur yang mewariskan sifat genetik, morfologi serta biokimia pada semua keturunannya. Hal tersebut membuat anggota monofiletik berkerabat sangat dekat satu dengan yang lain [13]. Pernyataan tersebut sesuai dengan pohon filogenetik yang terbentuk, karena seluruh anggota dalam pohon filogenetik (kecuali *out group*) berasal dari satu genus varian yang sama yakni covid-19 E.G.5.





Gambar 2. Phylogenetic Tree of Covid-19 Var E.G.5

Pohon Filogenetik dengan metode *maximum likelihood* pada Gambar 2 membentuk 16 kelompok *in group* yang terdiri dari 6 klad besar yakni klad I, klad II, klad IV, klad V, klad VII, dan klad XVI. Klad I terdiri dari 16 titik ditemukannya virus yakni di Jakarta (EPI_ISL_17664319/T) var E.G.5.1, Jakarta (EPI_ISL_17623747/Q) var E.G.5.1, Jakarta Selatan (EPI_ISL_17681322/W) var E.G.5, Jakarta Selatan (EPI_ISL_17681321/V) var E.G.5, Jakarta Selatan (EPI_ISL_17518075/J) var E.G.5.1, Jakarta (EPI_ISL_17389270/E) var E.G.5, Jakarta (EPI_ISL_17630625/S) var E.G.5, Jakarta Pusat (EPI_ISL_17551421/N) var E.G.5, Jakarta (EPI_ISL_17686559/Y) var E.G.5, Jakarta Pusat (EPI_ISL_18165701/AT) var E.G.5, dan Jakarta Utara (EPI_ISL_17997245/AO) var E.G.5.

Klad II terdiri dari 16 titik ditemukannya virus yakni virus covid var E.G.5.2 di Jakarta Pusat (EPI_ISL_18165626/AS), Banjarbaru (EPI_ISL_17719629/AD), Jakarta Selatan (EPI_ISL_17797245/AK), Bali (EPI_ISL_17788302/AG), Surabaya (EPI_ISL_17470035/I), Jakarta Selatan (EPI_ISL_17468575/H), Medan (EPI_ISL_17717748/AB), Medan (EPI_ISL_17785375/AF), Bali (EPI_ISL_17788380/AJ), Bali (EPI_ISL_18537584/AZ), Bali (EPI_ISL_17788316/AH), Bali (EPI_ISL_18537583/AY), Bandung (EPI_ISL_18146931/AR), Bali (EPI_ISL_17525998/K), dan Jakarta (EPI_ISL_17774280/AE). Klad III terdiri dari 5 titik ditemukannya virus yakni virus covid var E.G.5 di Jakarta (EPI_ISL_17686571/Z), Bogor (EPI_ISL_17681299/U), Jakarta (EPI_ISL_17389278 /F), Jakarta Barat (EPI_ISL_17372765/D), dan Jakarta (EPI_ISL_17308780/B). Klad III dan klad IV memiliki kesamaan yaitu pada semua titik mempunyai varian virus yang sama.

Klad IV terdiri dari 8 titik yang ditemukan di kota Jakarta yakni di Jakarta Utara (EPI_ISL_17681388/X) var E.G.5.1, Jakarta Timur (EPI_ISL_18465481/AV) var E.G.5.1, Jakarta Selatan (EPI_ISL_17719626/AC) Var E.G.5.2, Jakarta Utara (EPI_ISL_17835908/AL) Var E.G.5.1, Jakarta (EPI_ISL_17623756/R) var E.G.5.1, Jakarta Barat (EPI_ISL_18062407/AQ) var E.G.5.1, Jakarta Selatan (EPI_ISL_18465500/AW) var E.G.5.1.4, dan Jakarta (EPI_ISL_17308775/A) var E.G.5.1. Untuk klad V terdiri dari 6 titik daerah yaitu var E.G.5.1 di kota Jakarta Selatan (EPI_ISL_17997197/AN), Jakarta (EPI_ISL_17389302/G), Jakarta Pusat (EPI_ISL_18639168/BB), dan Jakarta (EPI_ISL_17686580/AA) serta var HV.1 di Surabaya (EPI_ISL_18705064/CI) dan Jakarta Utara (EPI_ISL_18673499/BV). Dan untuk Klad VI terdiri dari 4 titik kota yaitu var E.G.5.1 yang ditemukan di kota Jakarta (EPI_ISL_17308785/C) dan Jakarta Barat (EPI_ISL_18062407/AQ) serta var E.G.5.1.8 yang ditemukan di kota Jakarta Selatan (EPI_ISL_18673484/BU) dan Jakarta Selatan (EPI_ISL_18673536/BY).

Klad VII terdiri dari 7 titik kota yaitu var JG.3 di kota Jakarta Timur (EPI_ISL_18721886/CN), Surabaya (EPI_ISL_18705058/CG), Denpasar (EPI_ISL_18721815/CL), Denpasar (EPI_ISL_18721814/CK), Surabaya (EPI_ISL_18705068/CJ), dan Jakarta Utara (EPI_ISL_18639191/BI) serta var E.G.5.3 yang ditemukan di kota Jakarta Utara (EPI_ISL_18639167/BA). Untuk klad VII terdiri dari 3 titik kota yaitu ditemukannya virus covid-19 var E.G.5.1.1 di kota Surabaya (EPI_ISL_18705043/CD), Jakarta Utara (EPI_ISL_18639188/BH), dan kota Jakarta Utara (EPI_ISL_17835937/AM). Dan untuk Klad IX terdiri dari 2 titik kota yaitu kota Jakarta Selatan (EPI_ISL_18673534/BX) dan Jakarta Selatan (EPI_ISL_18673539/CA) berupa covid-19 var E.G.5.1.1.

Untuk klad X, XI, XII, XIII, XIV, dan XV masing-masing terdiri dari 1 titik kota dan memiliki varian virus covid yang sama yakni var HK3, dimana klad X di kota Jakarta Pusat (EPI_ISL_18639169/BC), klad XI ditemukan di kota Surabaya (EPI_ISL_18705059/CH), klad XII di kota Jakarta Utara (EPI_ISL_18639180/BE), klad XIII di kota Jakarta Timur (EPI_ISL_18639205/BO), klad XIV di kota Jakarta Utara (EPI_ISL_18639195/BK), dan klad XV ditemukan di kota Jakarta Selatan (EPI_ISL_18639204/BN).

Klad XVI merupakan klad yang memiliki titik kota paling banyak yakni sebesar 20 titik, untuk titik kotanya yaitu var HK.3 ditemukan di Jakarta Selatan (EPI_ISL_18673483/BT), Surabaya (EPI_ISL_17308785/C), Pekanbaru (EPI_ISL_18673439/BR), Jakarta Selatan (EPI_ISL_18639205/BO), Jakarta Selatan (EPI_ISL_18639198/BL), Jakarta Barat (EPI_ISL_18639191/BI), Jakarta Barat (EPI_ISL_18639194/BJ), Jakarta Barat (EPI_ISL_18465502/AX), dan Jakarta Utara (EPI_ISL_18639194/BL). Selain itu, ditemukan var HK.3.1 di Jakarta Barat (EPI_ISL_18673407/BQ) dan var HK.2 di Jakarta Selatan (EPI_ISL_18673548/CB), var HK3.2 ditemukan di kota Jakarta Utara (EPI_ISL_18639170/BD), Jakarta Selatan (EPI_ISL_18639208/BP), Surabaya (EPI_ISL_18705048/CE), Jakarta Utara (EPI_ISL_18721860/CM), Jakarta Selatan (EPI_ISL_18639203/BM), Jakarta Selatan (EPI_ISL_18673537/BZ), Jakarta Utara (EPI_ISL_18639186/BG), dan Jakarta Utara (EPI_ISL_18639194/BJ). Serta ditemukan pula var E.G.5.1.1 di kota Jakarta Selatan (EPI_ISL_18673529/BW), Jakarta Selatan (EPI_ISL_18639184/BF), dan Bandung (EPI_ISL_18693199/CC). Meskipun pada klad XVI terdapat tiga varian virus dan tidak berasal dari daerah yang berdekatan, namun titik tersebut merupakan satu kelompok klad yang sama karena kemiripan pada *sequence* nya.

Selanjutnya muncul frekuensi dasar empiris untuk *nucleotide* dan didapatkan nilai In likelihood sebesar -44725.41544 pada metode *maximum likelihood*. Berikut dasar frekuensi empiris:

Tabel 2. Frekuensi Dasar Empiris

<i>Nucleotide</i>	A	C	G	T
<i>Frequencies</i>	0.29887	0.18300	0.19611	0.32202

Ln Likelihood = -44725.41544

Between	And	Length	Approx. Confidence	Lim
-----	---	-----	-----	-----
2	EG51-JKT-A	0.00029	(0.00007,	0.0
2	EG51-JKT-G	0.00006	(zero,	0.0
2	21	0.00000	(zero,	infi
21	48	0.00000	(zero,	infi
48	18	0.00000	(zero,	infi
18	6	0.00000	(zero,	infi
6	20	0.00000	(zero,	infi
20	83	0.00000	(zero,	infi
83	22	0.00000	(zero,	infi
22	44	0.00000	(zero,	infi
44	53	0.00000	(zero,	infi
53	3	0.00000	(zero,	infi
3	4	0.00006	(zero,	0.0
4	8	0.00000	(zero,	infi
8	39	0.00000	(zero,	infi
39	89	0.00000	(zero,	infi
89	13	0.00000	(zero,	infi
13	9	0.00000	(zero,	infi
9	10	0.00000	(zero,	infi
10	37	0.00000	(zero,	infi
37	7	0.00000	(zero,	infi
7	38	0.00000	(zero,	infi
38	EG5-JP-N	0.00000	(zero,	infi
38	1	0.00006	(zero,	0.0
1	EG5-JKTA-B	0.00020	(0.00003,	0.0
1	46	0.00000	(zero,	infi
46	EG5-JB-D	0.00006	(zero,	0.0
46	5	0.00000	(zero,	infi
5	EG5-JKT-Z	0.00010	(zero,	0.0
5	32	0.00006	(zero,	0.0

32	EG5-Bgr-U	0.00006	(zero,	infinite
32	EG5-JKTA-F	0.00004	(zero,	0.0001
7	43	0.00013	(0.00000,	0.0002
43	EG5-JS-V	0.00006	(zero,	0.0001
43	EG5-JS-J	0.00006	(zero,	infinite
37	EG5-JKT-Y	0.00007	(zero,	0.0001
10	EG5-JKT-S	0.00006	(zero,	infinite
9	33	0.00007	(zero,	0.0001
33	40	0.00007	(zero,	0.0001
40	EG52-BL-A	0.00006	(zero,	0.0001
40	42	0.00000	(zero,	infinite
42	EG52-BL-AY	0.00040	(0.00018,	0.0006
42	41	0.00000	(zero,	infinite
41	34	0.00001	(zero,	infinite
34	EG52-BL-AJ	0.00010	(zero,	0.0002
34	EG52-BL-AH	0.00006	(zero,	0.0001
41	EG52-BL-AZ	0.00000	(zero,	infinite
33	50	0.00000	(zero,	infinite
50	51	0.00007	(zero,	0.0001
51	EG52-MDNaf	0.00006	(zero,	infinite
51	EG52-MDNab	0.00003	(zero,	infinite
50	35	0.00000	(zero,	infinite
35	EG52-BL-AG	0.00013	(0.00000,	0.0002
35	52	0.00000	(zero,	infinite
52	EG52-JS-H	0.00030	(0.00011,	0.0005
52	17	0.00000	(zero,	infinite
17	29	0.00006	(zero,	0.0001
29	EG52-JP-AS	0.00007	(zero,	0.0001
29	82	0.00000	(zero,	infinite
82	EG52-JS-AK	0.00007	(zero,	0.0001
82	EG52-BJBad	0.00000	(zero,	infinite
17	11	0.00000	(zero,	infinite
11	EG52-SBY-I	0.00007	(zero,	0.0001
11	47	0.00007	(zero,	0.0001
47	EG52-BDar	0.00013	(0.00000,	0.0002
47	36	0.00000	(zero,	infinite
36	EG52-JKTae	0.00007	(zero,	0.0001
36	EG52-BL-K	0.00000	(zero,	infinite
13	16	0.00006	(zero,	0.0001
16	EG5-JP-L	0.00006	(zero,	infinite
16	15	0.00000	(zero,	infinite
15	14	0.00003	(zero,	0.0001
14	EG5-JS-P	0.00006	(zero,	infinite
14	EG5-JP-M	0.00003	(zero,	infinite
15	EG5-JS-O	0.00000	(zero,	infinite
89	91	0.00011	(zero,	0.0001
91	EG5-JKT-T	0.00006	(zero,	infinite
91	EG5-JKT-Q	0.00005	(zero,	infinite
39	EG5-JS-W	0.00013	(0.00000,	0.0001
8	EG5-JKT-E	0.00007	(zero,	0.0001
4	85	0.00008	(zero,	0.0001
85	EG5-JU-AO	0.00008	(zero,	0.0001
85	30	0.00000	(zero,	infinite
30	EG5-JP-AU	0.00005	(zero,	0.0001
30	EG5-JP-AT	0.00000	(zero,	infinite
3	EG514-JSaw	0.00036	(0.00014,	0.0001
53	EG53-JU-BA	0.00016	(0.00002,	0.0001
44	EG51-JKT-C	0.00006	(zero,	0.0001
22	87	0.00012	(zero,	0.0001
87	EG51-JP-BB	0.00006	(zero,	infinite
87	81	0.00006	(zero,	0.0001
81	HV1-JU-BV	0.00016	(0.00003,	0.0001
81	HV1-SBY-CI	0.00020	(0.00004,	0.0001
83	EG51-JS-AN	0.00013	(0.00000,	0.0001
20	EG51-JB-AQ	0.00013	(0.00000,	0.0001
6	77	0.00003	(zero,	0.0001
77	80	0.00007	(zero,	0.0001
80	EG518-JSby	0.00020	(0.00004,	0.0001
80	EG518-JSbu	0.00017	(0.00002,	0.0001
77	25	0.00003	(zero,	infinite
25	78	0.00015	(0.00000,	0.0001
78	JG3-JU-BI	0.00007	(zero,	0.0001

78	76	0.00000	(zero,	0.0000
76	79	0.00037	(0.00015,	0.00059
79	JG3-BL-CK	0.00006	(zero,	0.00014
79	JG3-BL-CL	0.00007	(zero,	0.00016
76	26	0.00000	(zero,	0.0000
26	JG3-SBY-CJ	0.00019	(0.00003,	0.00039
26	45	0.00007	(zero,	0.00016
45	JG3-JT-CN	0.00013	(0.00000,	0.0002
45	JG3-SBY-CG	0.00007	(zero,	0.00016
25	61	0.00010	(zero,	0.0002
61	72	0.00006	(zero,	0.00014
72	HK3-JS-BT	0.00030	(0.00009,	0.00050
72	HK31-JB-BQ	0.00027	(0.00008,	0.00040
61	88	0.00000	(zero,	infinity
88	HK3-JU-BE	0.00024	(0.00006,	0.0004
88	12	0.00000	(zero,	infinity
12	HK3-JP-BC	0.00010	(zero,	0.0002
12	75	0.00000	(zero,	infinity
75	HK3-JB-BJ	0.00027	(0.00008,	0.00040
75	59	0.00000	(zero,	infinity
59	60	0.00030	(0.00011,	0.00050
60	71	0.00007	(zero,	0.0001
71	HK3-PKB-BR	0.00006	(zero,	0.00014
71	HK3-JS-CO	0.00019	(0.00003,	0.00039
60	HK3-JS-BL	0.00006	(zero,	infinity
59	68	0.00000	(zero,	infinity
68	HK2-JS-CB	0.00020	(0.00004,	0.00030
68	58	0.00000	(zero,	infinity
58	HK3-JS-BN	0.00044	(0.00020,	0.0006
58	57	0.00000	(zero,	infinity
57	HK3-JU-BJ	0.00044	(0.00020,	0.0006
57	56	0.00000	(zero,	infinity
56	EG511-JSbw	0.00024	(0.00006,	0.0004
56	55	0.00000	(zero,	infinity
55	HK3-JT-BO	0.00024	(0.00006,	0.0004
55	54	0.00000	(zero,	infinity
54	HK3-JU-BK	0.00027	(0.00008,	0.00040
54	49	0.00000	(zero,	infinity
49	73	0.00003	(zero,	0.00009
73	EG511-JSbf	0.00021	(0.00004,	0.00038
73	66	0.00003	(zero,	infinity
66	67	0.00007	(zero,	0.00016
67	EG511-JUbh	0.00030	(0.00011,	0.00050
67	HK32-JUbd	0.00007	(zero,	0.00016
66	EG511-BDcc	0.00037	(0.00015,	0.00059
49	28	0.00000	(zero,	infinity
28	HK3-JB-AX	0.00024	(0.00006,	0.0004
28	27	0.00000	(zero,	infinity
27	84	0.00006	(zero,	0.00014
84	EG511-JBam	0.00006	(zero,	0.00014
84	EG511-SBYc	0.00016	(0.00002,	0.000
27	69	0.00000	(zero,	infinity
69	70	0.00025	(0.00007,	0.000
70	EG511-JSca	0.00006	(zero,	infinity
70	EG511-JSbx	0.00006	(zero,	0.000
69	24	0.00000	(zero,	infinity
24	HK3-SBY-CH	0.00030	(0.00011,	0.000
24	23	0.00000	(zero,	infinity
23	62	0.00007	(zero,	0.000
62	63	0.00010	(zero,	0.000
63	HK32-JS-BZ	0.00030	(0.00011,	0.000
63	HK32-JU-BG	0.00006	(zero,	0.000
62	64	0.00006	(zero,	0.000
64	HK32-JSbm	0.00013	(0.00000,	0.000
64	74	0.00010	(zero,	0.000
74	HK32-JS-BP	0.00007	(zero,	0.000
74	65	0.00001	(zero,	infinity
65	HK32-JUcm	0.00010	(zero,	0.000
65	HK32-SBYce	0.00006	(zero,	0.000
23	HK3-SBY-CF	0.00024	(0.00006,	0.000
18	90	0.00002	(zero,	0.000
90	EG51-JKT-R	0.00006	(zero,	0.000
90	31	0.00002	(zero,	infinity
31	EG51-JU-X	0.00010	(zero,	0.000
31	19	0.00007	(zero,	0.000
19	EG52-JS-AC	0.00006	(zero,	0.000
19	86	0.00000	(zero,	infinity
86	EG51-JU-AV	0.00003	(zero,	0.000
86	EG51-JU-AL	0.00000	(zero,	infinity
48	EG51-JKTaa	0.00016	(0.00002,	0.000
21	EG51-JB-AP	0.00017	(0.00002,	0.000
* = significantly positive, P < 0.05				
** = significantly positive, P < 0.01				

Gambar 3. Output dari branch legh dan confidence limits

Berdasarkan Gambar 3, nilai \ln likelihood sebesar -44725.41544 merupakan nilai estimasi optimal (maksimum) likelihood dilakukan sebesar 44725.41544 . Nilai ini menunjukkan besarnya nilai estimasi yang paling maksimum untuk cabang dan dapat digunakan untuk menentukan pohon yang dapat menjelaskan pohon filogenetik lebih baik. Meskipun nilai \ln likelihoodnya negatif, hal itu berarti probabilitas yang sesuai kurang dari 1 karena yang dilihat adalah nilai algoritmanya. Selain itu, pada hasil output tersebut disebutkan antara dua percabangan terdapat panjang cabang (2 dengan EG51-JKT-EPI_ISL_17308775, panjang cabang 0.00029) berada pada *approximasi confidence limits* tertentu, artinya jarak tersebut berada dalam interval konfidensi 0.00007- 0.00051. Nilai positif pada interval konfidensi berarti pada percabangan tersebut tidak perlu di *arrangement*, karena nilai interval yang sempit menunjukkan panjang cabang lebih akurat yang diperkuat dengan nilai $p < 0,01$. Meski demikian, pada beberapa percabangan pohon terdapat aproksimasi kondensi limitnya yang mempunyai signifikansi positif dengan $P < 0,05$.

Analisis metode *maximum likelihood* merekonstruksikan kekerabatan antar spesies berdasarkan panjang garis cabang. Panjang yang berbeda menunjukkan tingkat evolusi masing-masing varian. Garis yang semakin panjang menunjukkan evolusi jarak semakin jauh, sedangkan garis yang lebih pendek menunjukkan dekatnya jarak evolusi varian [14]. Berdasarkan Gambar 2, garis terpanjang dimiliki oleh EPI_ISL_18639194 di kota Jakarta Utara anggota klad 16, hal tersebut menunjukkan bahwa penyebaran virus covid di Jakarta Utara (EPI_ISL_18639194) var HK3 merupakan spesies yang memiliki umur molekuler terlebih dahulu daripada covi-19 E.G.5 varian lainnya. Dan untuk cabang lainnya yang lebih pendek menunjukkan bahwa varian lain memiliki evolusi pendek dengan sifat lebih maju, sehingga dianggap sebagai keturunan dari nenek moyangnya.

4.2 Analisis Penyejajaran Ganda

Dari analisis penyejajaran ganda pada 93 *sequence* DNA pasien terinfeksi virus covid-19 varian E.G.5 yang telah dilakukan, diperoleh analisis sistem jaringan topologi, sistem jaringan daerah mutasi, dan sistem jaringan mode mutasi yang dijelaskan sebagai berikut:

1) Analisis Sistem Jaringan Topologi

Sistem jaringan topologi yang dihasilkan oleh output penyejajaran ganda, yaitu $G(W) = \{M, V, W\}$ dimana W adalah fungsi penalty dari output penyejajaran ganda. Dan matriks penalty diperoleh dari penyejajaran antar 2 *sequence* dengan menerapkan algoritma *Needleman Wunch* yang disimulasikan dalam Matlab sebagai berikut:

Tabel 3. Needleman Wunch

	B	C	D	E	F	G	H	I	CM	CN	CO	
A	0	4	5	5	0	5	5	5	0	0	4	5
B	4	0	10	5	5	10	2	4	3	6	4	5
C	5	10	0	1	2	3	4	3	3	10	11	6
D	5	5	1	0	10	5	6	6	2	5	6	7
E	0	5	2	10	0	1	1	11	6	7	2	6
F	5	10	3	5	1	0	3	1	1	5	1	5
G	5	2	4	6	1	3	0	6	5	7	2	1
H	5	4	3	6	11	1	6	0	4	8	7	12
I	0	3	3	2	6	1	5	4	0	0	11	6
...
...
CM	0	6	10	5	7	5	7	8	0	815	810	...
CN	4	4	11	6	2	1	2	7	11	819	0	815
CO	5	5	6	7	6	5	1	12	6	829	824	0

Dalam analisis sistem jaringan topologi juga diperoleh daerah stabil yang menunjukkan posisi nukleotid yang sama pada penyejajaran ganda dan daerah tidak stabil yang menunjukkan posisi nukleotid yang berbeda. Berikut adalah tabel daerah stabil dan tidak stabil pada penyejajaran ganda epidemi covid-19 E.G.5:

Tabel 4. Daerah Stabil dan Tidak Stabil pada Penyejajaran Ganda Sequence Covid-19 E.G.5

	Posisi nucleotida ke-	Jumlah	Prosentase
Daerah Stabil	323-332, 335-344, 354-356, 369-373, 375-377, 381-393, 740-745, 747, 749-776, 779-814, 816-820, 831, 832, 947-959, 961-1005, 1007-1011, 1013-1017, 1019-1075, 1077, 1181-1297, 1299-1303, 1336-1417, 1419-1439, 1546-1559, 1561-1572, 1574-1600, 1602-1676, 1678-1716, 1718-1798, 1812-1844, 1912-1944, 1946-1948, 1960-2009, 2015, 2020-2037, 2361-2339, 2600-2620, 2622-2686, 2773-2827, 2829-2841, 2843-2845, 2847-2861, 2863-2873, 2875-2900, 3022-3044, 3047-3053, 3055, 3059-385, 3165-3174, 3176-3230, 3232-3283, 3582-3610, 3612-3690, 3692-3693, 3695-3791, 4156-4168, 4173-4402, 4404-4501, 4585-4626, 4628-4651, 4653-4660, 4662-4670, 4672-4690, 4692-4753, 4755-4783, 4785-4801, 4803-4824, 4826-4854, 4856-4862, 4864-4902, 4986-5056, 5058-5108, 5179-5234, 5397-5443, 5445-5461, 5463-5621, 5623-5689, 5820-5918, 5920-5950, 5961-6100, 6105, 6107, 6162-6332, 6334-6374, 6376-6399, 6403-6438, 6440-6498, 6503-6568, 6570-6586, 6703-6778, 6780-6799, 6884-6888, 7656-7661, 7663-7672, 7740-7790, 7792-7813, 7861-7867, 8217-8232, 8234-8259, 8261-8268, 8270-8354, 8359-8411, 8413-8444, 8465-8493, 8495-8520, 8522-8550,	4.289	44,45%

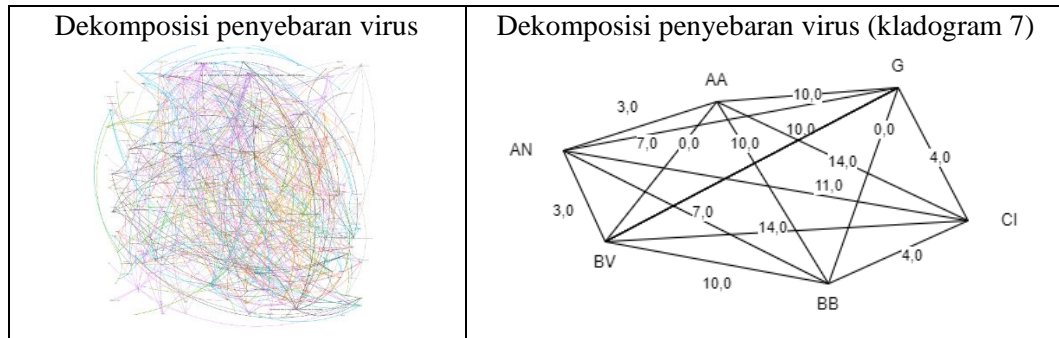
	8552-8556, 8558-8568, 8570-8577, 8579, 8601-8630, 8632-8682, 8684-8688, 8690, 8706, 8708-8762, 8764-8815, 8817-8847, 8849-8851, 8871-8872, 8874-8889, 8900-8941, 9564-9569, 9581-9576, 9579-9580, 9582-9601, 9603-9611, 9614-9621, 9624, 9626-9634, 9636-9657, 9659-9722, 9724-9740, 9742-9784, 9786-9799, 9801-9832, 9834-9836.		
Daerah tidak stabil	1-332, 333-334, 345-353, 357-368, 374, 378-380, 394-739, 746, 748, 777-778, 815, 821-830, 834-946, 960, 1006, 1012, 1018, 1076, 1078-1180, 1298, 1304-1335, 1418, 1440, 1461-1545, 1560, 1573, 1501, 1677, 1717, 1799-1811, 1845-1911, 1945, 1949-1959, 2010-2014, 2016-2019, 2018-2260, 2340-2599, 2621, 2687-2772, 2828, 2842, 2846, 2862, 2874, 2901-3023, 3045-3046, 3054, 3056-3058, 3086-3165, 3175, 3231, 3284-3581, 3611, 3691, 3694, 3792-4155, 4169-4172, 4403, 4502-4584, 4627, 4652, 4661, 4671, 4691, 4754, 4784, 4802, 4825, 4855, 4863, 4903-4985, 5057, 5109-5178, 5235-5396, 5444, 5462, 5622, 5690, 5819, 5919, 5951-5960, 6101-6104, 6106, 6018-6160, 6161, 6333, 6375, 6400-6402, 6439 6499-6502, 6569, 6587-6702, 6779, 6800-6883, 6889-7655, 7662, 7673-7739, 7814-7860, 7868-8216, 8233, 8260, 8269, 8355-8358, 8412, 8445-8464, 8494, 8521, 8551, 8557, 8569, 8578, 8600, 8631, 8683, 8689, 8707, 8763, 8816, 8848, 8852-8870, 8873, 8899, 8942-9563, 9570, 9577, 9578, 9581, 9602, 9612, 9613, 9622, 9623, 9625 9635, 9658, 9723, 9741, 9785, 9800, 9833, 9837-9649.	5.360	55,55%
Panjang <i>sequence</i>		9.649	100%

2) Analisis Sistem Jaringan Daerah Mutasi

Pada analisis sistem jaringan daerah mutasi dilakukan untuk membangun graf dan pohon filogenetik menggambarkan hubungan evolusi yang menunjukkan mutasi dari virus covid-19 yang tersebar di Indonesia. Berdasarkan Gambar 2 diperoleh hasil percabangan yang paling dekat yaitu OTU Jakarta (EPI_ISL_17664319) dan Jakarta (EPI_ISL_17623747) dengan jarak genetik masing-masing leaf yaitu bernilai 0. Karena Jakarta (EPI_ISL_17664319) disinyalir sebagai *host* dari virus covid-19 varian E.G.5, maka *sequence* yang paling dekat dengan *host* dianggap sebagai tempat penyebaran virus untuk pertama kali.

3) Analisis Sistem Jaringan Mode Mutasi

Sebelum melakukan analisis jaringan mode mutasi dari matriks penalti divisualisasi graf tidak berarah yang menunjukkan hubungan mutasi antar *sequence*. Notasi pada simpul menunjukkan nama *sequence* yang dikodekan sebagai huruf A, B, C, ..., CO dengan masing-masing kode mewakili nama *sequence* sebagaimana tercantum pada pembahasan sebelumnya.



Gambar 4. Dekomposisi Jaringan Mutasi Penyebaran Virus Covid-19 Varian E.G.5

Pada Gambar 4, label pada sisi menunjukkan banyaknya mutasi yang terjadi. Semakin tebal garis yang merepresentasikannya, maka semakin banyak mutasi yang terjadi antar simpul tersebut. Sebagaimana diketahui dari pembahasan sebelumnya, terdapat mutasi pada daerah tidak stabil yang berbeda terhadap 93 *sequence* DNA covid-19 varian E.G.5 ini. Terlihat dari gambar terjadi beberapa mutasi terdapat 16 pembagian dalam penyebaran virus covid-19 varian E.G.5 dan ada 374 aktor dan 470 hubungan di dalamnya. Pada dekomposisi penyebaran virus (klad 7) menunjukkan penyebaran mutasi covid-19 di 7 titik kota di Indonesia dengan dua jenis varian baru yakni var JG.3 yang mendominasi dan varian E.G.5.1.

5. Kesimpulan

Konstruksi filogenetik menggunakan maksimum likelihood ini dapat diterapkan pada kasus penyebaran covid-19 varian E.G.5. Metode berbasis probabilitas ini memungkinkan untuk akuisisi banyak topologi pohon dan karenanya memerlukan metode pencarian heuristik untuk mempersingkat waktu dalam memperoleh pohon filogenetik dengan nilai kemungkinan tertinggi. Hasil penelitian ini menunjukkan bahwa nilai maksimum likelihood sebesar -44725.41544 dan terbagi menjadi 16 klad dengan 6 klad besar dalam *cluster* penyebaran virus covid-19 var E.G.5. Nilai ini digunakan untuk menentukan pohon mana yang dapat menjelaskan penyelarasan *sequence* dengan baik. Gambar 2 menunjukkan penyebaran awal epidemi covid-19 var E.G.5 dimulai dari Jakarta, karena jarak genetik antara Jakarta (EPI_ISL_17664319) dengan Jakarta (EPI_ISL_17623747) yaitu bernilai 0,0000 th. Dan untuk analisis *multiple alignment* menunjukkan terdapat daerah tidak stabil sebesar 5.360, dimana terjadi mutasi dengan nomor nukleotida yang berbeda. Hasil analisa jaringan wilayah mutasi menunjukkan bahwa penyebaran epidermi Covid-19 E.G.5 berasal dari Jakarta Selatan dan mulai menyebar ke daerah sekitar Jakarta dan berlanjut ke Bali, Surabaya, Bandung, dan Pekanbaru.

6. Daftar Pustaka

- [1] K. Dowell, "Molecular Phylogenetics: An introduction to computational methods and tools for analyzing evolutionary relationships," *Mol Phylogenetics*, pp. 1–19, 2008.
- [2] D. Baum, "Reading a phylogenetic tree: the meaning of monophyletic groups," *Nature Education*, vol. 1, no. 1, p. 190, 2008.
- [3] S. Shen, *Theory and Mathematical methods in Bioinformatics*. Springer Science & Business Media, 2008.
- [4] A. R. Tantri, "Phylogenetic Affinities of SARS-CoV-2 Virus Genetic Variation Delta and Omicron in The Indonesian Population." Universitas Gadjah Mada, 2023.
- [5] N. G. A. Andreanto, "dinamika covid-19 di indonesia berdasarkan mutasi spike (s) selama satu tahun wabah/nicholas gerry andreanto." Universitas Negeri Malang, 2021.
- [6] S. Amiroch, M. S. Pradana, M. I. Irawan, and I. Mukhlash, "Multiple alignment analysis on phylogenetic tree of the spread of SARS epidemic using distance method," *J Phys Conf Ser*, vol. 890, no. 1, 2017, doi: 10.1088/1742-6596/890/1/012080.
- [7] S. Amiroch, M. Syaiful Pradana, M. Isa Irawan, and I. Mukhlash, "Maximum Likelihood Method on the Construction of Phylogenetic Tree for Identification the Spreading of SARS Epidemic," in *IEEE, IEEE*, 2018, pp. 137–141. doi: 10.1109/SAIN.2018.8673334.
- [8] A. Isaev, *Introduction to mathematical methods in bioinformatics*. Springer, 2004.
- [9] M. A. Larkin *et al.*, "Clustal W and Clustal X version 2.0," *bioinformatics*, vol. 23, no. 21, pp. 2947–2948, 2007.
- [10] J. L. Thorne, H. Kishino, and J. Felsenstein, "An evolutionary model for maximum likelihood alignment of DNA sequences," *Journal of Molecular Evolution*, vol. 33, pp. 114–124, 1991.
- [11] P. Lemey, M. Salemi, and A.-M. Vandamme, *The phylogenetic handbook: a practical approach to phylogenetic analysis and hypothesis testing*. Cambridge University Press, 2009.
- [12] T. Hidayat and A. Pancoro, "Ulasan Kajian filogenetika molekuler dan peranannya dalam menyediakan informasi dasar untuk meningkatkan kualitas sumber genetik anggrek," 2019.
- [13] F. Fitmawati, A. Suwita, N. Sofiyanti, and H. Herman, "Eksplorasi dan Karakterisasi Keanekaragaman Plasma Nutfah Mangga (*Mangifera*) di Sumatera Tengah," *Prosiding SEMIRATA 2013*, vol. 1, no. 1, 2013.