

Analisis Multiple Alignment Pada Penyebaran Epidemi Sars Cov E.G 5.1 Menggunakan Metode Neighbor - Joining

Carly Marshanda Arta MS¹, Siti Amiroch², Awawin Mustana Rohmah³

^{1,2,3}Universitas Islam Darul ‘Ulum Lamongan,

¹carly.2020@mhs.unisda.ac.id, ²siti.amiroch@unisda.ac.id, ³awawin.emer@unisda.ac.id

Abstract. SARS CoV-2 is a virus that is still a hot topic in the media and is very interesting to study. Moreover, SARS CoV-2 is increasingly mutating over time and giving rise to new types of variants. Recently, the world was again horrified by the emergence of a new type of SARS CoV-2 variant called the E.G 5.1 variant or commonly called Eris. In Indonesia, the E.G 5.1 variant was first reported in Jakarta Province on 09/03/2023. Based on this, the author wants to know the process of spreading the SARS CoV E.G 5.1 Epidemic that occurred in Indonesia with Multiple Alignment analysis. This analysis has several stages, among others, analysing the topology network system, mutation area network system and mutation mode network system, so that a phylogenetic tree is obtained using the Neighbor-Joining algorithm which is used to determine the origin of the virus spread. The data used are 92 DNA sequences obtained through GISAID. The results of the analysis obtained the beginning of the spread of SARS CoV E.G 5.1 in Indonesia which briefly started from Jakarta 09/03/2023, then spread to Bogor 20/04/23, Medan 11/05/23, Surabaya 03/07/23, Bandung 24/10/23, Riau 07/12/23, and finally spread in Bali Province (Denpasar) on 10/12/23 and 11/12/23.

Keywords: Multiple Alignment, Neighbour-Joining, Spread of SARS-CoV E.G 5.1, Phylogenetic Analysis

Abstrak. SARS CoV-2 merupakan suatu virus yang masih menjadi topik hangat di media dan sangat menarik untuk dikaji. Apalagi SARS CoV-2 semakin bermutasi dari waktu ke waktu dan memunculkan varian jenis baru. Akhir-akhir ini dunia kembali dihebohkan dengan munculnya varian SARS CoV-2 jenis baru yang bernama varian E.G 5.1 atau biasa disebut Eris. Di Indonesia, varian E.G 5.1 pertama kali dilaporkan di Provinsi Jakarta pada 09/03/2023. Berdasarkan hal tersebut, penulis ingin mengetahui proses penyebaran Epidemi SARS CoV E.G 5.1 yang terjadi di Indonesia dengan analisis Multiple Alignment. Analisis ini memiliki beberapa tahap antara lain, melakukan analisis sistem jaringan topologi, sistem jaringan daerah mutasi dan sistem jaringan mode mutasi, sehingga diperoleh pohon filogenetik menggunakan algoritma Neighbor-Joining yang digunakan untuk menentukan awal mula penyebaran virus. Data yang digunakan adalah data 92 sekuen DNA yang diperoleh melalui GISAID. Hasil dari analisis tersebut diperoleh awal mula penyebaran SARS CoV E.G 5.1 di Indonesia yang secara singkat berawal dari Jakarta 09/03/2023, kemudian menyebar ke Bogor 20/04/23, Medan 11/05/23, Surabaya 03/07/23, Bandung 24/10/23, Riau 07/12/23, dan terakhir menyebar di Provinsi Bali (Denpasar) pada tanggal 10/12/23 dan 11/12/23.

Kata Kunci: Multiple Alignment, Neighbor-Joining, Penyebaran SARS-CoV E.G 5.1, Analisis Filogenetik

1 Pendahuluan

Human Coronavirus (hCoV) merupakan virus yang berhubungan dengan infeksi saluran pernapasan atas (*Respiratory Tract Infection*) [1] yang disebabkan oleh Severe Acute Respiratory Syndrome Coronavirus-2 (SARS CoV-2). Virus ini pertama kali terdeteksi di Wuhan, Provinsi Hubei, Tiongkok pada akhir bulan Desember 2019 [2].

SARS CoV-2 merupakan virus yang sangat menular yang kemudian menyebar ke seluruh dunia termasuk Indonesia dan akhirnya menyebabkan epidemi SARS [3][4]. Dari proses penyebaran ini, kemungkinan besar virus akan bermutasi yang mengakibatkan munculnya virus dengan varian baru, sehingga penularan virus bisa menjadi lebih mudah. Salah satu varian SARS CoV yang ada antara lain: varian Alpha (Nomenklatur Pongo: B.1.1.7) yang dilaporkan di Inggris pada bulan September 2020; varian Beta (Nomenklatur Pongo: B.1.351, B.1.351.2, B.1.351.3) yang dilaporkani di Afrika Selatan pada bulan Mei 2020; varian Gamma (Nomenklatur Pongo: P.1, P.1.1, P.1.2) yang dilaporkan di Brazil pada bulan November 2020; varian Delta (Nomenklatur Pongo: B.1.617.2, AY.1, AY.2, AY.3) yang dilaporkan di India pada bulan Oktober 2021; dan varian Omicron (nomenklatur Pongo: B.1.1.529) yang pertama kali dilaporkan di Afrika Selatan [5].

Menurut data WHO pada tanggal 31 Desember 2023, SARS CoV-2 telah menginfeksi manusia di seluruh dunia sebanyak 773.819.856 orang dengan total kematian sebanyak 7.010.568 orang. Gelombang pertama pandemi SARS CoV di Indonesia terjadi pada tanggal 02 Maret 2020 hingga 30 April 2021. Kemudian gelombang kedua, didominasi oleh varian Delta yang mencapai puncaknya pada Juli 2021, dan disusul gelombang ke-tiga yang didominasi oleh varian Omicron yang puncaknya pada Februari 2022 [6]. Baru-baru ini, terdapat varian SARS CoV baru yaitu varian Eris E.G 5 yang terdeteksi di Indonesia pada awal Maret 2023 [7]. Varian ini merupakan subvarian dari turunan Omicron [8]. Pada tanggal 09/03/2023, virus SARS CoV E.G 5.1 pertama kali terdeteksi di Indonesia di Provinsi DKI Jakarta. Virus ini kemudian menyebar ke 5 provinsi di Indonesia, yaitu Provinsi Sumatera Utara, Kalimantan Selatan, Jawa Barat, Jawa Timur, dan Bali. Di Indonesia jumlah kasus yang terinfeksi SARS CoV varian E.G 5 terus meningkat. Dengan menyebarnya varian baru tersebut, maka diperlukan kewaspadaan dalam mengatasi penyebaran virus.

Bioinformatika merupakan bidang ilmu yang menggunakan prinsip-prinsip statistik, matematika, dan ilmu komputer untuk menyelesaikan segala permasalahan yang berkaitan dengan biologi [6][7]. Dalam dunia bioinformatika, terdapat suatu ilmu yang dapat mencari bagaimana penyebaran sekuen virus itu terjadi. Salah satunya adalah penerapan dari analisis penyejajaran ganda. Penyejajaran Ganda atau Multiple Alignment (MA) merupakan penyejajaran dua atau lebih sekuen secara bersamaan [11]. Tujuan dilakukannya penyejajaran ganda adalah untuk membuat pohon filogenetik. Pohon filogenetik merupakan pohon yang menggambarkan kekerabatan antar sekuen yang nantinya digunakan untuk mengetahui penyebaran virus yang terjadi.

Dalam penelitian ini, analisis penyejajaran ganda dilakukan dengan algoritma Needleman–Wuncsh, yaitu algoritma pemrograman dinamik yang menerapkan prinsip penyejajaran global [12]. Sedangkan metode yang digunakan untuk mengonstruksi pohon filogenetik yaitu algoritma Neighbor-Joining dengan *software* Clustal Omega. Algoritma ini memerlukan matriks jarak untuk mengonstruksi pohon filogenetik yang digunakan untuk mengetahui penyebaran virus SARS CoV E.G 5.1 yang terjadi di Indonesia.

2 Metode Penelitian

2.1 Pengumpulan Data

Data yang digunakan dalam penelitian ini adalah data 92 sekuen DNA dari pasien yang terinfeksi virus SARS CoV E.G 5.1 diberbagai provinsi di Indonesia (Jakarta, Sumatera Utara, Kalimantan Selatan, Jawa Barat, Jawa Timur, dan Bali) yang diperoleh melalui website <https://gisaid.org/>. Data ini terdiri dari sekuen ke-1 dengan ID akses GISAID yaitu EPI_ISL_17308775 sampai sekuen ke-92 dengan ID akses EPI_ISL_18673548 yang masing – masing sekuen memiliki panjang sekuen yang berbeda.

2.2 Penyejajaran Sekuen

Menurut Shen (2006) dalam [13], penyejajaran sekuen merupakan metode yang penting dalam analisis posisi dan tipe mutasi yang tersembunyi didalam sekuen biologi dan memungkinkan dilakukan perbandingan yang tepat. Hal yang terpenting dalam penyejajaran sekuen adalah menentukan perpindahan mutasi yang menyebabkan jarak genetik tiap sekuen [14].

Diberikan dua sekuen A_1 dan A_2 yang didefinisikan sebagai berikut:

$$A_1 = (a_{11}, a_{12}, \dots, a_{1n_a}) \quad \text{dan} \quad A_2 = (a_{21}, a_{22}, \dots, a_{2n_a})$$

Penyisipan simbol “-“ ke dalam sekuen A_1 dan A_2 bertujuan untuk membentuk dua buah sekuen yang baru yaitu sekuen A_1' dan A_2' . Selanjutnya, elemen-elemen dari sekuen A_1 dan A_2 memiliki *range* dari $V_5 = \{0, 1, 2, 3, 4\}$ atau $\{a, c, g, t, -\}$. Definisi dari penyejajaran ganda adalah kumpulan dari sekuen yang dinyatakan sebagai:

$$\mathcal{A} = \{A_1, A_2, \dots, A_m\}$$

Untuk setiap A_s merupakan sekuen terpisah yang didefinisikan pada V_q , dan dinyatakan sebagai berikut :

$$A_s = (a_{s,1}, a_{s,2}, \dots, a_{s,n}), \quad S = 1, 2, \dots, m$$

Dengan n_s merupakan panjang sekuen A_s dan m merupakan jumlah sekuen pada masing-masing kelompok.

2.3 Pairwise Alignment

Metode *Pairwise Alignment* sangat penting dalam bioinformatika untuk memproses penyejajaran sekuen. Metode yang digunakan pada *pairwise alignment* adalah metode Needleman-Wunsch yang merupakan algoritma penyelarasan global [12]. Adapun langkah-langkah dalam algoritma Needleman-Wunsch adalah sebagai berikut:

Diketahui barisan $A = (a_1, a_2, \dots, a_n)$, dan $B = (b_1, b_2, \dots, b_n)$ sebagaimana terlihat pada Tabel 1 berikut.

Tabel 1. Tabel Dua Dimensi Sekuen A, B

	a_1	a_2	...	a_n
b_1	$s(0,0)$	$s(1,0)$	$s(2,0)$	$s(n,0)$
b_2	$s(0,1)$	$s(1,1)$	$s(2,1)$	$s(n,1)$
b_n	$s(0,m)$	$s(1,m)$	$s(2,m)$	$s(n,m)$

Sedangkan untuk elemen $s(i, j)$ dapat dihitung oleh tiga elemen yaitu, $s(i - 1, j - 1)$ pada pojok kiri atas, $s(i - 1, j)$ pada sisi kiri dan $s(i, j - 1)$ pada bagian atas. Pertama-tama tentukan skor marginal $s(i, 0)$ dan $s(0, j)$. Sedangkan nilai *virtual simbol* adalah d , dimana *virtual symbol* merupakan panjang dari *virtual symbol* itu sendiri. Sehingga $s(0, j) = -j \times d$, $s(i, 0) = -i \times d$, dan $s(0,0) = 0$. Maka elemen dapat dihitung dengan menggunakan persamaan (1) sebagai berikut:

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

2.4 Matriks Penalti

Matriks penalti menunjukkan tingkat perbedaan untuk tiap-tiap unit molekul, seperti nukleotida atau asam amino dalam sekuen biologi. Matriks penalti dapat dinotasikan sebagai berikut:

$$D = d(a, b)_{a, b \in V_s}$$

Dalam bioinformatika, matriks penalti pada penyejajaran sekuens DNA ditetapkan oleh matriks Hamming. Didefinisikan matriks Hamming untuk V_s adalah sebagai berikut:

$$d_H(a, b) = \begin{cases} 0, & \text{jika } a = b \in V_s \\ 1, & \text{jika } a \neq b \end{cases}$$

2.5 Algoritma Neighbor – Joining

Algoritma Neighbor-Joining merupakan metode berbasis jarak yang digunakan untuk mengonstruksi pohon filogenetik [15][15][14][13]. Algoritma ini memerlukan inputan berupa matriks jarak yang diperoleh dengan cara menyejajarkan setiap sekuen untuk mencari ketidaksamaan barisan yang disejajarkan [15]. Berikut merupakan langkah-langkah metode Neighbor-Joining :

- 1) Menentukan matriks jarak N sekuen.
- 2) Sebuah pohon dengan semua OTU dalam matriks sebagai percabangan dari titik pusat dibuat asumsi, kemudian dibentuk dalam pola.
- 3) Untuk masing-masing OTU, menghitung S, dengan S merupakan jumlah dari jarak (D) antara OTU satu dengan OTU lain, dibagi (N-2), dengan N adalah total banyaknya OTU.

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

- 4) Identifikasi pasangan OTU dengan nilai minimum:

$$M_{ij} = D_{ij} - S_i - S_j \quad (3)$$

- 5) Menggabungkan dua taksa pada satu node dalam sebuah *subtree*.
- 6) Hitung panjang cabang:

$$D_{xi} = \frac{D_{ij}+S_i-S_j}{2}, \quad D_{xj} = \frac{D_{ij}+S_j-S_i}{2} \quad (4)$$

- 7) Menghitung jarak matriks yang baru dengan menghubungkan i dan j , kemudian mengganti dengan node (x) yang menghubungkannya.

$$D_{xk} = \frac{D_{ik}+D_{jk}+D_{ij}}{2} \quad (5)$$

- 8) Ulangi langkah tersebut sampai diperoleh (N-2) sekuen.

3 Hasil dan Pembahasan

Pada tahap ini melakukan analisis penyejajaran ganda untuk mengonstruksi pohon filogenetik. Langkah-langkah yang dilakukan yaitu menganalisis sistem jaringan topologi, menganalisis sistem jaringan daerah mutasi, dan menganalisis sistem jaringan mode mutasi yang akan dijabarkan berikut.

3.1 Analisis Sistem Jaringan Topologi

Sistem jaringan topologi dihasilkan oleh output penyejajaran ganda, yaitu $G(W) = \{M, V, W\}$, dimana W adalah fungsi penalti dari output penyejajaran ganda. Dan matriks penalti diperoleh melalui penyejajaran antar dua sekuen menggunakan algoritma Needleman-Wuncsh. Setelah dilakukan penyejajaran sekuen, akan diketahui perbedaan tiap sekuen. Jumlah perbedaan inilah yang akan digunakan sebagai matriks penalti. Dibawah ini merupakan matriks penalti pada hasil penyejajaran sekuen SARS CoV E.G 5.1.

	S1	S2	S3	S4	S5	...	S88	S89	S90	S91	S92
S1	0	68	53	69	48	...	73	77	91	73	67
S2	68	0	35	32	36	...	44	48	62	45	66
S3	53	35	0	59	59	...	61	71	86	67	88
S4	69	32	59	0	5	...	14	21	14	17	62
S5	48	36	59	5	0	...	15	47	62	44	89
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
S88	73	44	61	14	15	...	0	16	30	21	58
S89	77	48	71	21	47	...	16	0	17	28	65
S90	91	62	86	14	62	...	30	17	0	43	80
S91	73	45	67	17	44	...	21	28	43	0	69
S92	67	66	88	62	89	...	58	65	80	69	0

Dimana S1 mewakili sekuen 1 yaitu hCoV-19/Indonesia/JK-CW62-4064/2023 dengan ID EPI_ISL_17308775 (Jakarta 09/03/2023), S2 mewakili sekuen 2 yaitu hCoV-19/Indonesia/JK-CW62-4073/2023 dengan ID EPI_ISL_17308780 (Jakarta 09/03/2023), S3 mewakili sekuen 3 yaitu hCoV-19/Indonesia/JK-CW62-4086/2023 dengan ID EPI_ISL_17308785 (Jakarta 10/03/2023), dan seterusnya sampai pada S92

yang mewakili sekuen 92 yaitu hCoV-19/Indonesia/JK-BIOKES-W04218/2023 dengan ID EPI_ISL_18721886 (Jakarta Timur 26/12/2023). Setelah diperoleh matriks penalti, langkah selanjutnya yaitu menentukan jarak genetik antar sekuen dengan menggunakan persamaan (6) berikut:

$$\text{Jarak Genetik} = \frac{\text{Beda antar sekuen}}{\text{Panjang sekuen}} \quad (6)$$

Dengan menggunakan persamaan (6) diperoleh jarak genetik antar sekuen yang terlihat pada Tabel 2 berikut:

Tabel 2. Jarak Genetik Masing – masing Sekuen

	S1	S2	S3	S4	S5	...	S88	S89	S90	S91	S92
S1	0	0.00233	0.00182	0.00236	0.00164	...	0.00250	0.00250	0.00264	0.00312	0.00250
S2	0.00233	0	0.00120	0.00110	0.00123	...	0.00151	0.00151	0.00164	0.00212	0.00154
S3	0.00182	0.00120	0	0.00202	0.00202	...	0.00206	0.00209	0.00243	0.00295	0.00230
S4	0.00236	0.00110	0.00202	0	0.00017	...	0.00048	0.00048	0.00072	0.00048	0.00058
S5	0.00164	0.00123	0.00202	0.00017	0	...	0.00051	0.00051	0.00161	0.00212	0.00151
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
S88	0.00250	0.00151	0.00206	0.00048	0.00051	...	0	0.00069	0.00089	0.00140	0.00079
S89	0.00250	0.00151	0.00209	0.00048	0.00051	...	0.00069	0	0.00055	0.00103	0.00072
S90	0.00264	0.00164	0.00243	0.00072	0.00161	...	0.00089	0.00055	0	0.00058	0.00096
S91	0.00312	0.00212	0.00295	0.00048	0.00212	...	0.00140	0.00103	0.00058	0	0.00147
S92	0.00250	0.00154	0.00230	0.00058	0.00151	...	0.00079	0.00072	0.00096	0.00147	0

Setelah diperoleh jarak genetik, langkah selanjutnya yaitu menentukan daerah stabil dan tidak stabil. Daerah stabil merupakan daerah yang memiliki nukleotida yang sama, sedangkan daerah yang tidak stabil merupakan daerah yang memiliki nukleotida yang berbeda. Daerah tidak stabil ini nantinya digunakan untuk mengetahui posisi terjadinya mutasi antar sekuen DNA pasien yang terinfeksi SARS CoV E.G 5.1. Setelah menentukan daerah tidak stabil melalui *multiple alignment*, diperoleh jumlah perbandingan antara daerah yang stabil dan tidak stabil. Berikut perbandingan antara daerah stabil dan tidak stabil pada *multiple alignment* dan posisinya yang terlihat pada Tabel 3.

Tabel 3. Daerah Stabil dan Tidak Stabil Beserta Posisinya

	Posisi Nucleotide ke-	Jumlah	Persentase (%)
Daerah Stabil	1-30, 68-85, 87-94, 96-108, 110-152, 154-183, 433-466, 468, 470-502, 528-536, , 28823-28831, 28833-28866, 28868-28966, 28968-28976	15330	52,52
Daerah Tidak Stabil	31-67, 86, 95, 109, 153, 184-432, 467, 469, 503-527, 537-538, 542-570, 590, 600-607, 1702, 1712, 1729, 1799-1800, 1854, 1912-1913, 1927-1957, 2305-2347, 2368, 2485, 2503, 2519, 2559, 2562, 2694, 2700-3006, 3068, 3359, 3379-3471, 3529, 3550, 3787, 3797, 3851-4114, 4147, 4152, 4185-4186, 4270, 4497, 4618, 4767, 4843, 4865-4878, 5001-5198, 5250, 5301-5302, 5314-5341, 5343, 5452, 5498-5510, 5516-5525, 5581-6548, 6552, 6585, 6630, 6666-7263, 7783, 7955, 7993, 8006, 8053, 8089, 8121, 8172-8536, 8601, 8603, 8607, 8627, 8632-8641, 8661, 8724-8963, 8990, 8992, 9133, 9159, 9320-10209, 10212, 10300, 10474, 10540-10541, 10548-10549, 10639, 10666, 10785, 10845-11930, 11974-11983, 12461, 12562, 12576, 12676, 12781, 12835, 12862, 12977-13219, 13347, 13422, 13428, 13451, 13479, 13542, 13728, 13819, 13875, 13945, 14031, 14041, 14056, 14175-	13860	47,48

14423, 14639, 14754, 14794-15000, 15171-15654, 15799, 15853, 16335, 16415, 16443, 16536, 16581, 16923, 17179, 17195, 17223, 17250, 17320-17338, 17340, 17349-17778, 17793, 17801, 17949, 18033, 18286, 18468-18469, 18543, 18592, 18594, 18667-18668, 18670, 18672, 18675, 18782, 18965-18970, 19175, 19228-19574, 19805, 19866-22431, 22449, 22452, 22463, 22470, 22472, 22475-22476, 22486-22684, 22839, 22896, 22916-24115, 24166, 24247, 24276, 24361, 24504, 24534-24539, 24701, 24796, 24802-24859, 24864, 24946, 24950, 24970, 25029, 25071, 25121, 25124, 25126, 25138, 25175, 25201, 25239, 25267, 25311, 25361, 25373, 25479, 25486, 25517, 25535, 25568, 25590, 25756, 25915, 25918, 26013, 26023-26078, 26087, 26138, 26166, 26203, 26207, 26236, 26293-28156, 28176, 28198, 28202, 28272, 28302, 28306, 28334-28335, 28341, 28372, 28392, 28441-28442, 28527, 28567, 28610, 28636, 28691, 28801, 28822, 28832, 28867, 28967, 28977-29190.		
Jumlah	29190	100 %

Berdasarkan Tabel 3, terlihat bahwa jumlah daerah yang stabil sebanyak 15.330 daerah, dengan posisi nukleotid ke 1 – 30, 68 – 85, 87 – 94, 96 – 108, dan seterusnya sampai pada nukleotid ke 28968 – 28976. Sedangkan untuk daerah yang tidak stabil memiliki daerah sebanyak 13860 daerah, dengan posisi nukleotid ke 31 – 67, 86, 95, 109, 153, 184 – 432, dan seterusnya sampai pada posisi nukleotid ke 28977 - 29190. Perbandingan jumlah daerah stabil dan tidak stabil masing-masing sebesar 52,52 % dan 47,48 % dari total panjang sekuen yaitu 29.190. Berikut hasil secara keseluruhan banyaknya nukleotida pada daerah yang tidak stabil dan posisinya yang terlihat pada Tabel 4.

Tabel 4. Banyaknya Nukleotid pada Daerah yang Tidak Stabil dan Posisinya

Posisi Nukleotid	Jumlah Nukleotid					
	A	C	G	T	N	-
31	0	0	0	91	1	0
32	0	0	91	0	1	0
⋮	⋮	⋮	⋮	⋮	⋮	⋮
28867	0	1	0	91	0	0
⋮	⋮	⋮	⋮	⋮	⋮	⋮
28967	0	90	0	0	2	0
⋮	⋮	⋮	⋮	⋮	⋮	⋮
29188	0	88	0	0	0	4
29189	0	88	0	0	0	4
29190	0	88	0	0	0	4

Dari Tabel 4 berarti bahwa posisi nukleotid pada daerah yang tidak stabil merupakan posisi terjadinya mutasi antar sekuen DNA pasien yang terinfeksi SARS. Perbedaan ini terlihat dari jumlah nukleotid yang tidak sama pada masing-masing posisi nukleotid pada 92 sekuen yang disejajarkan.

3.2 Analisis Sistem Jaringan Daerah Mutasi

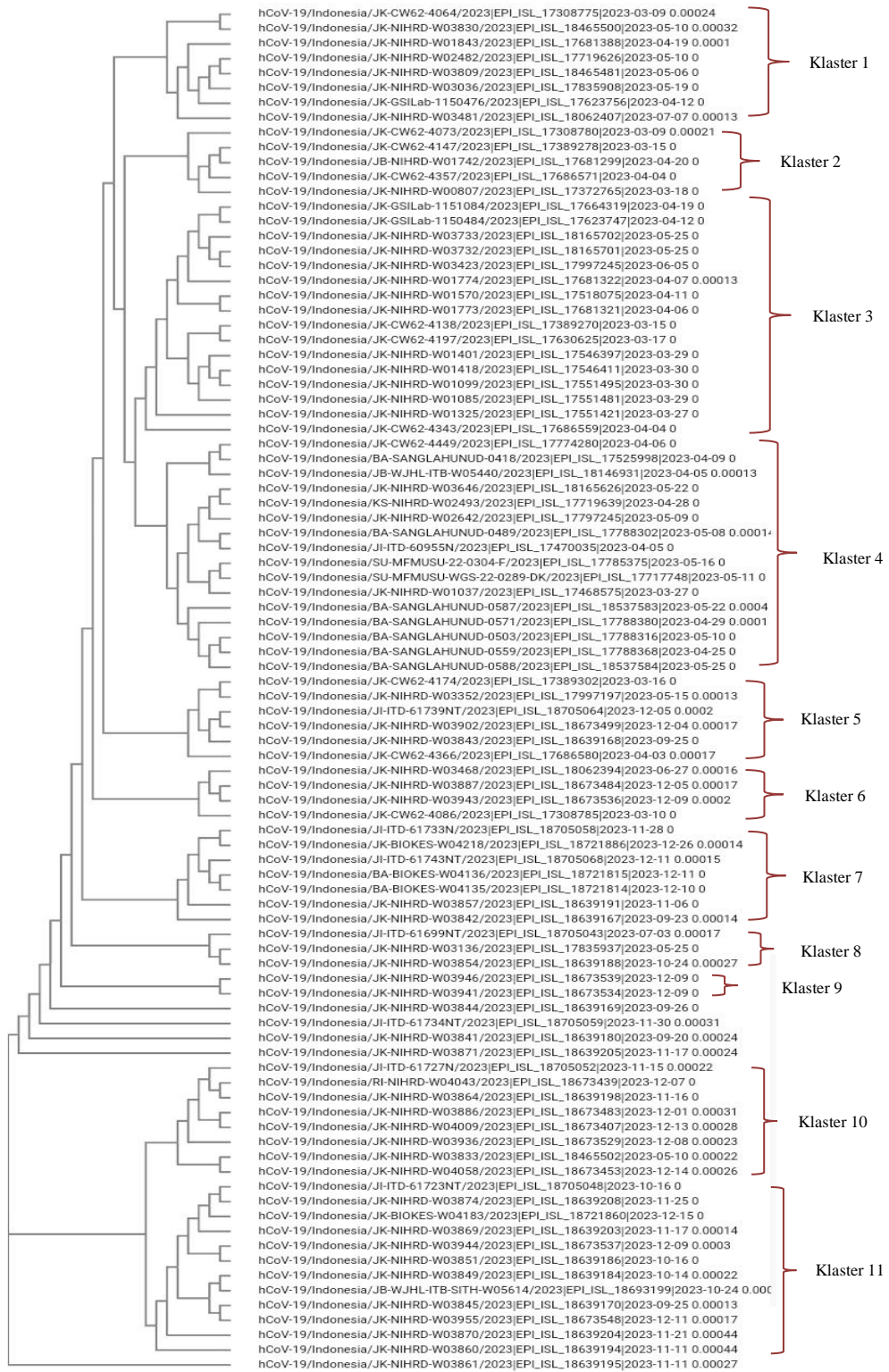
Setelah melakukan analisis jaringan topologi dengan diperoleh daerah tidak stabil, maka langkah selanjutnya yaitu menganalisis sistem jaringan daerah mutasi pada *multiple alignment* SARS CoV E.G 5.1. Langkah untuk melakukan sistem jaringan daerah mutasi adalah dengan membangun graf dan pohon filogenetik dari hasil penyejajaran sekuen SARS CoV E.G 5.1. Pohon filogenetik ini merepresentasikan hubungan evolusi yang menunjukkan penyebaran SARS CoV E.G 5.1 dari sekuen satu ke sekuen lain yang mana masing-masing sekuen diambil dari pasien yang

berbeda. Sehingga hasil akhir dari pohon filogenetik ini merupakan gambaran proses penyebaran SARS E.G 5.1 dari berbagai provinsi di Indonesia. Dalam pembentukan pohon filogenetik, tentunya membutuhkan matriks penalti sebagai inputan. Dalam penelitian ini, penulis membuat pohon filogenetik dengan *software* CLUSTAL Omega. Output yang dihasilkan berupa pohon filogenetik tanpa perhitungan koreksi jarak, karena *software* tersebut tidak dilengkapi dengan perhitungan koreksi jarak tersebut.

Berikut disajikan output pohon filogenetik penyebaran epidemic SARS CoV E.G 5.1 di Indonesia dengan *software* CLUSTAL Omega menggunakan metode Neighbor-Joining tanpa koreksi jarak yang terlihat pada Gambar 1.

Pohon filogenetik pada Gambar 1 menjelaskan tentang hubungan kekerabatan antar virus yang diperkirakan sebagai awal mula penyebaran SARS CoV E.G 5.1 di Indonesia. Tiap *leaf* (daun) merepresentasikan penyebaran SARS-CoV dari satu daerah yang kemudian menyebar ke daerah lain. Pada Gambar 1, terlihat bahwa penyebaran SARS CoV E.G 5.1 di Indonesia terbagi menjadi beberapa klaster, dimana setiap klaster terdapat awal mula penyebaran yang ditentukan berdasarkan percabangan yang paling dekat. Pada klaster ke-1, percabangan yang paling dekat adalah ID EPI_ISL_17308775 (Jakarta 09/03/23) dengan EPI_ISL_18465500 (Jakarta Selatan 10/05/23) dan ID EPI_ISL_17719626 (Jakarta 10/05/23) dengan EPI_ISL_18465481 (Jakarta Timur 06/05/23). Pada klaster ke-2, percabangan yang paling dekat adalah ID EPI_ISL_17389278 (Jakarta 15/03/23) dengan EPI_ISL_17681299 (Bogor 20/04/23). Dan seterusnya sampai pada klaster ke-11 yaitu ID EPI_ISL_18639184 (Jakarta Selatan 14/10/23) dengan ID EPI_ISL_18693199.

Berdasarkan hasil klaster pada pohon filogenetik tersebut, secara keseluruhan penyebaran ini berawal dari Jakarta 09/03/2023, setelah itu pada tanggal 10/05/2023, virus tersebut menyebar secara bersamaan ke Jakarta Selatan, selanjutnya menuju ke Jakarta Timur, Jakarta Utara dan Jakarta. Dari sini, virus terus menyebar sampai Jakarta Utara 19/04/2023 dan Jakarta Barat 07/07/2023. Sementara di daerah lain juga terdeteksi virus di Jakarta 15/03/23 kemudian menyebar ke Bogor 20/04/23 sampai menyebar ke daerah lain yang sama-sama teridentifikasi virus di Jakarta 06/04/23 dan Bali 09/04/23. Selanjutnya virus ini terus menyebar sampai Medan (Sumatera Utara) 11/05/23, Surabaya (Jawa Timur) 03/07/23, Bandung (Jawa Barat) 24/10/23, Riau (Pekanbaru) 07/12/23, dan penyebaran terakhir di Bali (Denpasar) pada tanggal 10/12/23 dan 11/12/23.

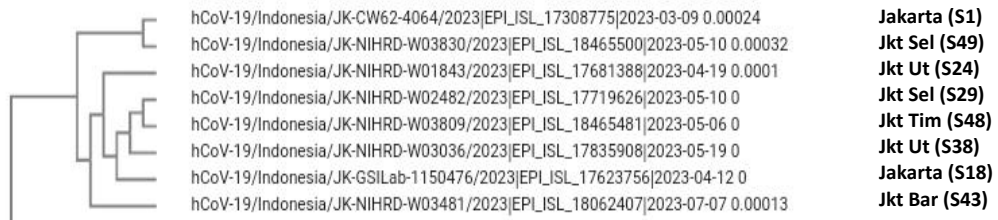


Gambar 1. Pohon Filogenetik Penyebaran Epidem SARS CoV E.G 5.1 di Indonesia

3.3 Analisis Sistem Jaringan Mode Mutasi

Pada tahap ini melakukan visualisasi graf tidak berarah yang dapat ditunjukkan hubungan mutasi antar sekuen. Notasi pada simpul menunjukkan nama sekuen yang diberi kode sebagai kode S1, S2, S3, sampai S92 dengan masing-masing kode mewakili nama sekuennya. Untuk membuat graf jarak yang digunakan adalah matriks penalti antar sekuen yang telah diperoleh sebelumnya.

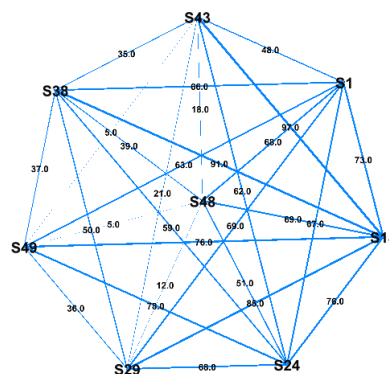
Berikut merupakan hasil kluster pada tahap awal penyebaran epidemi SARS COV E.G 5.1 yang terlihat pada Gambar 2.



Gambar 2. Hasil Kluster Tahap Awal Epidemi SARS COV E.G 5.1

Berdasarkan Gambar 2, dapat dijelaskan bahwa untuk penyebaran pertama pada tahap awal berasal dari Jakarta (S1) pada tanggal 09/03/2023, kemudian menyebar ke Jakarta Selatan (S49) pada tanggal 10/05/2023. Pada hari yang sama, juga terjadi penyebaran di Jakarta Selatan (S29) yang sebelumnya berawal dari Jakarta Timur (S48) pada tanggal 06/05/2023. Dari sini kemudian virus menyebar ke Jakarta Utara (S38), Jakarta (S18), dan Jakarta Barat (S43). Dari hasil penyebaran tahap awal epidemi SARS CoV E.G 5.1 dengan pohon filogenetik pada Gambar 1, dapat disajikan hasil visualisasi graf yang merepresentasikan banyaknya mutasi virus yang terjadi pada tahap awal.

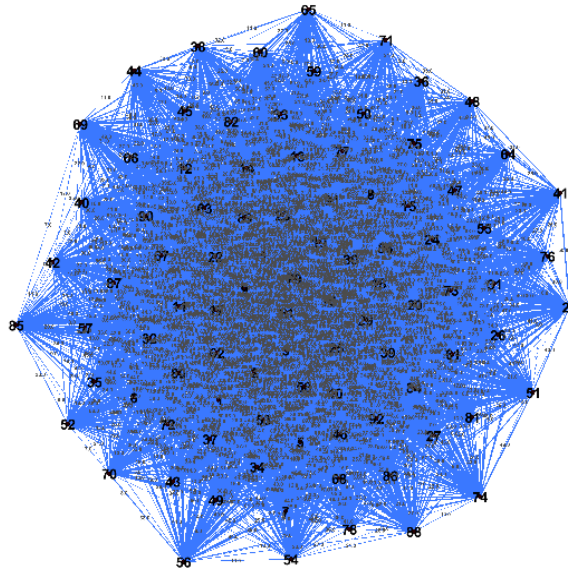
Berikut merupakan visualisasi graf yang merepresentasikan banyaknya mutasi virus yang terjadi di tahap awal yang disajikan pada Gambar 3.



Gambar 3. Visualisasi Graf pada Kluster Awal

Berdasarkan Gambar 3, terlihat hasil visualisasi graf pada kluster awal penyebaran epidemi SARS CoV E.G 5.1. Node S1, S18, S24, S29, S38, S48, dan S49 merepresentasikan sekuen 1, sekuen 18, sekuen 24, sekuen 29, sekuen 38, sekuen 48 dan sekuen 49. Label pada sisi yang menghubungkan antar node merepresentasikan banyaknya mutasi yang terjadi antar sekuen.

Berikut visualisasi graf hubungan mutasi antar sekuen secara keseluruhan yang terlihat pada Gambar 4.



Gambar 4. Visualisasi Graf Secara Keseluruhan pada Mutasi Penyebaran Epidemii SARS CoV E.G 5.1

Pada Gambar 4 terlihat bahwa tiap node merepresentasikan masing-masing sekuen. Label (angka) pada sisi menunjukkan banyaknya mutasi yang terjadi antar sekuen. Semakin tebal (gelap) garis yang merepresentasikannya, maka semakin banyak mutasi yang terjadi antar sekuen tersebut.

4 Kesimpulan

Dari keseluruhan hasil analisis penyejajaran ganda, dapat disimpulkan bahwa penyejajaran ganda dapat diterapkan untuk pengontruksian pohon filogenetik yang digunakan untuk mencari awal mula penyebaran SARS CoV E.G 5.1. Awal mula penyebaran virus ini berasal dari Jakarta 09/03/2023. Dari sini virus menyebar ke Jakarta 12/04/2023, selang 7 hari kemudian menyebar ke Jakarta Utara. Di daerah lain seperti Jakarta 06/04/23 juga terdeteksi virus kemudian menyebar di Bali 09/04/23, Medan (Sumatera Utara) 11/05/23, Surabaya (Jawa Timur) 03/07/23, Bandung (Jawa Barat) 24/10/23, Riau (Pekanbaru) 07/12/23, dan terakhir di Bali (Denpasar) pada tanggal 10/12/23 dan 11/12/23. Secara garis besar penyebab penyebaran virus ini berasal dari Jakarta yang kemudian menyebar ke berbagai provinsi di Indonesia.

5 Daftar Pustaka

- [1] N. Aliabadi, M. Jamaliduost, G. Pouladfar, N. H. Marandi, and M. Ziyaeyan, "Characterization and phylogenetic analysis of Iranian SARS-CoV-2 genomes: A phylogenomic study," *Heal. Sci. Reports*, vol. 6, no. 1, pp. 1–9, 2023, doi: 10.1002/hsr2.1052.
- [2] S. Gaffar, S. A. F. Rahmani, and A. Hardianto, "Mutation and Phylogenetic Analysis of Spike Glycoprotein of Indonesian Isolates of Severe-Acute-

- Respiratory-Syndrome-Coronavirus-2 (SARS-CoV-2),” *Maj. Kedokt. Bandung*, vol. 53, no. 1, 2021, doi: 10.15395/mkb.v53n1.2230.
- [3] A. M.; Rohmah and D. Rahmalia, “SEIR model simulation on the spreading of Ebola virus between two regions,” *J. Phys. Conf. Ser.*, vol. 1, 2021.
- [4] A. M. Rohmah and D. Rahmalia, “Penyelesaian positif model penyebaran virus ebola antar dua wilayah,” *J. Mat. Univ. Udayana*, vol. 2, no. 10, pp. 44–52, 2020.
- [5] A. Susilo *et al.*, “Mutasi dan Varian Coronavirus Disease 2019 (COVID-19): Tinjauan Literatur Terkini,” *J. Penyakit Dalam Indones.*, vol. 9, no. 1, p. 59, 2022, doi: 10.7454/jpdi.v9i1.648.
- [6] R. Bestari *et al.*, “SARS-CoV-2 lineages circulating during the first wave of the pandemic in North Sumatra, Indonesia,” *IJID Reg.*, vol. 8, no. July, pp. S1–S7, 2023, doi: 10.1016/j.ijregi.2023.07.006.
- [7] S. U. Firdaus, J. Wiwoho, and M. A. Hidayat, “The Effectiveness of the Implementation of Occupational Health and Safety (OHS) in Indonesian Medical Personnel During the Covid-19,” *UNIFIKASI J. Ilmu Huk.*, vol. 10, no. 2, pp. 74–87, 2023, doi: 10.25134/unifikasi.v10i2.7938.
- [8] A. Girma, “The Many Mutations of the COVID-19 Variant: Current Perspectives on EG.5/Eris,” *Environ. Health Insights*, vol. 17, pp. 0–4, 2023, doi: 10.1177/11786302231217805.
- [9] A. E. E. D. Rashed, H. M. Amer, M. El-Seddek, and H. E. D. Moustafa, “Sequence Alignment Using Machine Learning-Based Needleman-Wunsch Algorithm,” *IEEE Access*, vol. 9, pp. 109522–109535, 2021, doi: 10.1109/ACCESS.2021.3100408.
- [10] M. S. Pradana and S. Amiroch, “Protein sequence analysis of the Zika virus and the dengue virus using Smith Waterman algorithm,” *AIP Conf. Proc.*, vol. 2084, no. March, 2019, doi: 10.1063/1.5094275.
- [11] 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.
- [12] S. Amiroch, M. I. Irawan, I. Mukhlash, A. Nur, M. Ansori, and A. Nidom, “Identification of the Spread of the Influenza Virus Type A / H9N2 in Indonesia Using the Neighbor-Joining Algorithm with Felsenstein Models 使用费尔森施泰因模型的邻居加入算法识别印度尼西亚一种 / 我们的权利型流感病毒的传播,” 2021.
- [13] S. Amiroch and A. Rohmatullah, “Determining geographical spread pattern of MERS-CoV by distance method using Kimura model,” *AIP Conf. Proc.*, vol. 1825, no. Symomath, pp. 1–8, 2017, doi: 10.1063/1.4978970.
- [14] M. I. Irawan and I. Mukhlash, “Maximum Likelihood Method on The Construction of Phylogenetic Tree for Identification the Spreading of SARS Epidemic,” no. 1, pp. 137–141, 2018.
- [15] S. Amiroch, M. S. Pradana, M. I. Irawan, and I. Mukhlash, “A Simple Genetic Algorithm for Optimizing Multiple Sequence Alignment on the Spread of the SARS Epidemic,” *Open Bioinforma. J.*, vol. 12, no. 1, pp. 30–39, 2019, doi: 10.2174/1875036201912010030.