

**IMPLEMENTASI METODE *MAXIMUM LIKELIHOOD*
DALAM ANALISIS FILOGENETIKA VIRUS MPXV
(*MONKEYPOX*)**

(Studi Kasus : Kekerabatan Jenis Virus *Monkeypox*)

TUGAS AKHIR

Diuji sebagai Salah Satu Syarat Untuk Memperoleh Gelar Sarjana Program
Studi Statistika



Disusun Oleh:

Sodik

20611198

**PROGRAM STUDI STATISTIKA
FAKULTAS MATEMATIKA DAN ILMU PENGETAHUAN ALAM
UNIVERSITAS ISLAM INDONESIA
YOGYAKARTA
2024**

**HALAMAN PERSETUJUAN PEMBIMBING
TUGAS AKHIR**

Judul : Implementasi Metode *Maximum Likelihood*
dalam Analisis Filogenetika Virus MPXV
(*Monkeypox*)
(Studi Kasus : Kekerabatan Jenis Virus
MonkeyPox)

Nama Mahasiswa : Sodik

NIM : 20611198

**TUGAS AKHIR INI TELAH DIPERIKSA DAN DISETUJUI UNTUK
DIUJIKAN**

Mengetahui

Ketua Prodi Statistika



Dr. Atina Ahdilqa, S.Si., M.Si.

Yogyakarta 7 Februari 2025

Dosen Pembimbing



Dr.techn. Rohmatul Fajriyah, S.Si, M.Si

**HALAMAN PENGESAHAN
TUGAS AKHIR**

**IMPLEMENTASI METODE *MAXIMUM LIKELIHOOD* DALAM
ANALISIS FILOGENETIKA VIRUS MPXV (*MONKEYPOX*)**
(Studi Kasus : Kekerabatan Jenis Virus *MonkeyPox*)

Nama Mahasiswa : Sodik

NIM : 20611198

**PROPOSAL TUGAS AKHIR INI TELAH DIUJIKAN
PADA TANGGAL : 29 April 2025**

Nama Penguji		Tanda Tangan
1.	Dina Tri Utari, S.Si., M.Sc.	
2.	Dr. Siti Amiroch S.Si., M.Si	
3.	Dr.techn. Rohmatul Fajriyah, S.Si, M.Si.	

Mengetahui,

Dekan Fakultas Matematika dan Ilmu Pengetahuan Alam




Prof. Riyanto, S.Pd., M.Si., Ph.D.



KATA PENGANTAR



Assalamu 'alaikum Wr.Wb

Puji syukur penulis panjatkan kehadirat Allah SWT yang telah melimpahkan rahmat, taufik, dan hidayah-Nya berupa keimanan, kekuatan, kesabaran, kelancaran, dan keselamatan dalam pembuatan Seminar Proposal. Shalawat serta salam penulis haturkan kepada Nabi Muhammad SAW beserta keluarga, sahabat, dan umatnya. Penelitian ini tersusun sebagai hasil Tugas Akhir dipenuhi dalam langkah menyelesaikan jenjang Strata Satu (S1) di Program Studi Statistika, Universitas Islam Indonesia. Penelitian ini berjudul **“Implementasi Metode *Maximum Likelihood* dalam Analisis Filogenetika Virus *MonkeyPox*”**

Penulis menyadari bahwa penulisan laporan ini banyak memperoleh bantuan dari berbagai pihak, baik yang berupa saran, kritik, bimbingan maupun bantuan lainnya. Oleh karena itu, pada kesempatan ini penulis menyampaikan ucapan terima kasih kepada:

1. Allah SWT atas rahmat, kesehatan, karunia, dan petunjuk-Nya yang telah diberikan;
2. Bapak Prof. Riyanto, S.Pd., M.Si., Ph.D., selaku Dekan Fakultas Matematika dan Ilmu Pengetahuan Alam, Universitas Islam Indonesia;
3. Ibu Dr. Atina Ahdika, S.Si., M.Si., selaku Ketua Program Studi Statistika, Fakultas Matematika dan Ilmu Pengetahuan Alam, Universitas Islam Indonesia;
4. Ibu Dr.techn. Rohmatul Fajriyah, S.Si, M.Si, selaku dosen pembimbing yang telah memberikan bimbingan dan arahan sejak awal hingga penyusunan Proposal ini selesai;
5. Dosen-dosen Statistika Universitas Islam Indonesia yang selalu berbagi ilmu baik dalam bidang akademik maupun non akademik;
6. Dewan Penguji yang memberikan saran dalam penulisan Proposal ini.

7. Orang tua penulis Bapak Sidik dan Ibu Yayah serta keluarga besar bapak rusli yang selalu memberikan dukungan, semangat, dan doa;
8. Teman – Teman seperjuangan Paringga, Luthfi, Ika, Indah Fifeh, Valin, Widya, Hafis, Muna, Yudis, Ferry, Niam, Farhan, Husain, Hafis A, Izul, Hendra dan Roni dan semuanya yang memberikan support dalam penyelesaian Tugas Akhir ini.
9. Para senior yang luar biasa yang selalu memberikan motivasi Kak mongga selaku mentor dan *Best* kating selama kuliah, Kak widi, Mba Nisya, yang selalu memberikan motivasi untuk *study abroad*.
10. Teman-teman Organisasi seperjuangan JAG 2021/2022 dan Kafe Prancis KUI 2024/2025 yang selalu memberikan semangat selama kuliah.
11. Teman-teman ALKARA (Statistika 2020) yang sudah banyak memberikan semangat dan bantuan dalam penyusunan Proposal ini;
12. Semua pihak yang tidak dapat penulis sebutkan yang telah membantu dalam bentuk apapun.

Penulis menyadari sepenuhnya bahwa Tugas Akhir ini masih jauh dari sempurna karena keterbatasan kemampuan dan pengetahuan. Oleh karena itu, segala kritik dan saran yang sifatnya membangun selalu penulis harapkan. Semoga Proposal TA ini dapat bermanfaat bagi penulis khususnya dan bagi semua pihak yang membutuhkan. Akhir kata, semoga Allah SWT selalu melimpahkan rahmat serta hidayah-Nya kepada kita semua, Aamiin ya robbal ‘alamiin

Wassalamu ‘alaikum Wr.Wb

Yogyakarta, 20 Januari 2025

Sodik

DAFTAR ISI

HALAMAN JUDUL.....	i
HALAMAN PERSETUJUAN PEMBIMBING TUGAS AKHIR	ii
HALAMAN PENGESAHAN TUGAS AKHIR.....	iii
KATA PENGANTAR.....	iv
DAFTAR ISI	vi
DAFTAR TABEL	viii
DAFTAR GAMBAR.....	ix
DAFTAR LAMPIRAN	x
PERNYATAAN	xi
INTISARI.....	xii
ABSTRACT	xiii
BAB I PENDAHULUAN	1
1.1. Latar Belakang Masalah.....	1
1.2. Rumusan Masalah	4
1.3. Batasan Masalah.....	5
1.4. Tujuan Penelitian	5
1.5. Manfaat Penelitian	5
BAB II TINJAUAN PUSTAKA	6
BAB III LANDASAN TEORI	19
3.1 Biologi Komputasi	19
3.2 Filogenetika.....	20
3.3 Asam Deokrisibonukleat (DNA).....	21
3.4 <i>Multiple Sequence Alignment</i>	22
3.5 MAFFT	23
3.6 <i>Concensus dan Occupancy</i>	23
3.7 <i>Align Crustal W</i>	24
3.8 <i>Align by MUSCLE</i>	24
3.9 <i>Maximum Likelihood</i>	25
3.10 <i>Phylogenetic Likelihood</i>	28
3.11 <i>Likelihood Computation</i>	30
3.12 <i>Summarizing Bootstrapped Phylogenies</i>	31
3.13 <i>Maximum Likelihood and more complex models of evolution</i>	32
3.14 <i>General Time Reversible (GTR)</i>	33
3.15 <i>Bootstrapping</i>	35
3.16 <i>MonkeyPox (Cacar Monyet)</i>	35
3.17 <i>Molecular Phylogenetics</i>	37
BAB IV METODOLOGI PENELITIAN.....	42
4.1. Populasi dan Sampel Penelitian	42
4.2. Tempat dan Waktu Penelitian	43
4.3. Definisi Operasional Variabel Penelitian.....	43
4.4. Alat dan cara organisir data.....	44
4.5. Metode Penelitian.....	44
BAB V PEMBAHASAN	46
5.1. Statistika Deskriptif.....	46
5.2. Analisis Philogenetics	55

BAB VI KESIMPULAN	79
6.1 Kesimpulan	79
6.2 Saran.....	81
DAFTAR PUSTAKA.....	82
LAMPIRAN	89

DAFTAR TABEL

Tabel 2.1 Daftar Penelitian Sebelumnya	10
Tabel 3.1 Penerapan Model GTR.....	33
Tabel 3.2 Kelebihan dan kekurangan setiap metode	39
Tabel 4.1 Data Variabel	43
Tabel 5.1 Data pasien berdasarkan gender	47

DAFTAR GAMBAR

Gambar 3.1 <i>Maximum Likelihood</i> Topologi (Johanna Lindgren, 2019).....	30
Gambar 3.2 <i>MonkeyPox Research Timeline</i> (Matthew et al, 2020).....	36
Gambar 3.3 <i>Moleculer Phylogenetics</i> (Suresh Kaushik et.al, 2021).....	38
Gambar 4.1 Tahapan Penelitian	45
Gambar 5.1 Persentase <i>Monkeypox</i> berdasarkan <i>Gender</i>	46
Gambar 5.2 <i>Clade/Lineage</i> berdasarkan negara.....	47
Gambar 5.3 Visualisasi <i>monkeypox</i> dengan <i>Treemap</i>	51
Gambar 5.4 Peta penyebaran <i>Monkeypox</i> berdasarkan Negara dan Benua.....	52
Gambar 5.5 <i>Monkeypox</i> Berdarkan kategori umur.....	53
Gambar 5.6 <i>Monkeypox</i> berdarkan <i>Gender</i> dan Kondisi Pasien.....	54
Gambar 5.7 (<i>Multiple Sequence Alignment, MSA by MAFFT</i>)	55
Gambar 5.8 <i>Consensus Adenin</i>	57
Gambar 5.9 <i>Occupancy Adenin</i>	58
Gambar 5.10 <i>Concencus (Timin)</i>	58
Gambar 5.11 <i>Occupancy (Timin)</i>	58
Gambar 5.12 <i>Concencus (Guanine)</i>	59
Gambar 5.13 <i>Occupancy (Guanine)</i>	60
Gambar 5.14 <i>Concencus (Cytosyn)</i>	60
Gambar 5.15 <i>Occupancy (Cytosyn)</i>	60
Gambar 5.16 <i>MX Information</i>	61
Gambar 5.17 <i>Phylogenetics Tree By Maximum Likelihood</i>	62
Gambar 5.18 <i>Phylogenetics Tree circular sample</i>	63

DAFTAR LAMPIRAN

Lampiran 1.....	89
Lampiran 3.....	93
Lampiran 3.....	94

PERNYATAAN

Dengan ini saya menyatakan bahwa dalam Tugas Akhir ini tidak terdapat karya karya yang sebelumnya pernah diajukan untuk memperoleh gelar kesarjanaan di suatu Perguruan Tinggi dan sepanjang pengetahuan saya juga tidak terdapat karya atau pendapat yang pernah ditulis atau diterbitkan orang lain, kecuali yang diacu dalam naskah ini dan disebutkan dalam daftar pustaka.

Yogyakarta, 20 Januari 2025



Sodik

INTISARI

IMPLEMENTASI METODE *MAXIMUM LIKELIHOOD* DALAM ANALISIS FILOGENETIKA VIRUS MPXV (*MONKEYPOX*)

(Kekerabatan Jenis Virus *Monkeypox*)

Sodik

Program Studi Statistika, Fakultas Matematika dan Ilmu Pengetahuan Alam
Universitas Islam Indonesia

MonkeyPox merupakan salah satu masalah global belakangan ini. Penyebabnya adalah penyebaran virus dari infeksi monyet yang menyalur kepada manusia secara global. Penelitian ini menggunakan $n = 106$ DNA genom virus dari berbagai negara, termasuk sampel dari Indonesia. Penelitian ini bertujuan untuk mengkaji hubungan evolusioner antar genom virus MPXV (*Monkeypox*) melalui rekonstruksi pohon filogenetika, guna memahami penyebaran dan perkembangan berbagai varian di sejumlah benua. Dalam penelitian ini, analisis filogenetika yang dilakukan yaitu dengan metode *maximum likelihood* dalam konteks filogenetika molekuler, dengan tujuan memberikan wawasan biologis terkait diversitas varian MPXV. Data yang digunakan merupakan data sekunder yang diperoleh dari basis data GISAID. Proses analisis melibatkan sejumlah perangkat lunak bioinformatika, termasuk *Jalview*, *MAFFT*, *MEGA-X*, *Microsoft Excel*, *Power BI*, dan *iTOL v5*. Penelitian ini menggunakan metode *maximum likelihood*, dengan model substitusi untuk DNA adalah GTR+G+I. Hasil dari analisis penelitian ini yaitu rekonstruksi pohon filogenetika yang di tambahkan dengan *phenotype*, adapun hasil dengan menggunakan metode *maximum likelihood* dengan model substitusi DNA GTR+G+I. Hasil analisis menunjukkan bahwa terdapat Clade I, II, dan IIb yang menjadi induk *clade* yang terbagi menjadi tiga klaster yang di tandai dengan berbagai varian warna. Selain itu, hasil juga menunjukkan adanya *lineage* (garis turunan dari induk *clade*) seperti pada *clade* I terdapat sub-*clade* Ia dan Ib dan *lineage* garis keturunan dari *clade* II dan IIb yang paling banyak tersebar, seperti IIa, IIb, IIb B.1, IIb C.1, IIb 1.B.20, IIb C.1.1, IIb A, IIb A.2, IIb B.1.6, IIb B.1.7, IIb B.1.2, IIb C.1, IIb B.1.22, IIb A2, IIb A.1, IIb A.2.1, IIb B.1.1, IIb B.1.10, IIb B.1.17, IIb B.1.3, IIb A.2.2, IIb B.1.1, IIb B.1.10, IIb B.1.11, IIb B.1.20, IIb B.1.3, IIb B.1.6 dan IIb. Dari beberapa sampel menunjukkan adanya garis keturunan yang berasal dari Indonesia yaitu IIb C.1 yang berasal dari *clade* IIb.

Kata Kunci : filogenetika molekuler, *Maximum-likelihood*, *Monkeypox*.

ABSTRACT

MAXIMUM LIKELIHOOD METHOD IMPLEMENTATION IN PHYLOGENETIC ANALYSIS OF MPXV VIRUSES (MONKEYPOX)

Case Study : Relationship types of monkeypox virus

Sodik

Department of Statistics, Faculty of Mathematics and Natural Sciences
Universitas Islam Indonesia

Monkeypox has recently emerged as a significant global health concern. This is primarily due to the widespread transmission of the MPXV (Monkeypox) viruses, originally zoonotic from monkeys to humans, across various regions of the world. This study utilized $n = 106$ complete MPXV genome DNA sequences from multiple countries, including samples from Indonesia. The objective of this research is to investigate the evolutionary relationships among MPXV genomes through phylogenetic tree reconstruction, in order to better understand the spread and development of various viral variants across continents. Phylogenetic analysis was conducted using the maximum likelihood method within the framework of molecular phylogenetics, aiming to provide biological insight into MPXV variant diversity. The dataset used in this study is secondary data retrieved from the GISAID database. Several bioinformatics tools were employed during the analysis, including Jalview, MAFFT, MEGA-X, Microsoft Excel, Power BI, and iTOL v5. The maximum likelihood method was applied using the GTR+G+I nucleotide substitution model. The analysis resulted in a reconstructed phylogenetic tree annotated with phenotypic data. The tree revealed three main clades: Clade I, Clade II, and Clade Iib, each further subdivided into three major clusters, distinguished by various color codes. Additionally, the analysis identified multiple lineages (descendant branches) derived from the main clades. Clade I consisted of sub-clades Ia and Ib, while Clades II and Iib included the most diverse and widespread lineages, such as Iia, Iib, Iib B.1, Iib C.1, Iib B.1.20, Iib C.1.1, Iib A, Iib A.2, Iib B.1.6, Iib B.1.7, Iib B.1.2, Iib B.1.22, Iib A.1, Iib A.2.1, Iib B.1.1, Iib B.1.10, Iib B.1.17, Iib B.1.3, Iib A.2.2, and others. Notably, one of the samples, belonging to lineage Iib C.1 under Clade Iib, was identified as originating from Indonesia.

Keywords: *Molecular phylogenetics, maximum-likelihood, computational biology.*

BAB I

PENDAHULUAN

1.1. Latar Belakang Masalah

Cacar monyet adalah penyakit *zoonosis*, yaitu penyakit menular yang menyerang manusia dari hewan. Penyakit cacar yang disebabkan oleh virus *monkeypox* ini, pertama kali diidentifikasi pada tahun 1958 pada koloni monyet yang dipelihara di laboratorium Denmark untuk tujuan penelitian. Kasus cacar monyet yang pertama kali dilaporkan menginfeksi manusia terjadi di Republik Demokratik Kongo pada tahun 1970. Cacar monyet menginfeksi manusia yang melakukan kontak dekat dengan hewan yang terinfeksi, umumnya di lokasi hutan hujan tropis. Cacar monyet merupakan penyakit endemik di 10 negara Afrika, terutama di Afrika Barat dan Afrika Tengah (WHO, 2022).

Sejak 1 Januari 2022, lonjakan kasus telah terjadi di luar negara-negara Afrika, dengan sebagian besar kasus terdeteksi pada orang yang tidak memiliki riwayat perjalanan dari Afrika. Pada Juli 2022, WHO (*World Health Organization*) mengumumkan hampir 14.000 kasus yang dikonfirmasi di sekitar 70 negara yang biasanya tidak ditemukan. Hingga 23 Juli 2022, direktur jenderal WHO *Tedros Adhanom Ghebreyesus*, melalui akun twitternya mengumumkan bahwa wabah cacar monyet adalah masalah kesehatan masyarakat darurat yang menjadi perhatian internasional.

Endemi berarti adanya epidemi (definisi: penyakit menular yang menyebar dengan cepat di wilayah yang luas dan menyebabkan banyak korban) yang terjadi yang terjadi di seluruh dunia, terjadi di wilayah geografis yang sangat luas, melintasi batas-batas negara dan biasanya mempengaruhi banyak orang (Phina, 2009). Cacar monyet didefinisikan sebagai endemi karena kasusnya telah terjadi di berbagai benua atau negara dan telah menjadi masalah bagi seluruh warga dunia. Fakta juga menunjukkan bahwa cacar monyet telah terjadi di berbagai belahan dunia berdasarkan WHO Region, seperti di Eropa (EURO), Amerika (AMRO), Asia Tenggara (SEARO), Mediterania Timur (EMRO), dan Pasifik (AFRO). Afrika (AFRO), dan Pasifik Barat (WPRO). Saat ini, WHO (*World Health Organization*) mencatat setiap pembaruan harian tentang kasus cacar monyet yang dikonfirmasi

di berbagai wilayah dengan jumlah kasus terkonfirmasi tertinggi adalah Kasus dengan jumlah kasus terkonfirmasi tertinggi adalah Eropa dan Amerika (WHO, 2022).

Menurut *World Health Organization* (WHO), kasus cacar monyet (*monkeypox*) dikategorikan berdasarkan tingkat keparahan dan konteks epidemiologis. Kasus suspek melibatkan seseorang dengan ruam vesikular atau pustular yang tidak diketahui penyebabnya, disertai demam, pembengkakan kelenjar getah bening, sakit kepala, nyeri otot, dan nyeri punggung. Kasus itu mencakup individu dengan kriteria klinis yang memiliki riwayat kontak dengan kasus terkonfirmasi, riwayat perjalanan ke daerah endemik, atau kontak dengan bahan terkontaminasi dalam 21 hari sebelum gejala muncul. Kasus terkonfirmasi didiagnosis melalui hasil laboratorium positif untuk virus *monkeypox*, menggunakan tes PCR atau isolasi virus. Kasus terinfeksi adalah kasus yang diidentifikasi pada individu yang baru saja bepergian dari daerah endemik. Pengelolaan kasus meliputi isolasi pasien terkonfirmasi, perawatan suportif untuk gejala, dan penggunaan antibiotik jika terjadi infeksi bakteri sekunder. Pelacakan kontak penting dilakukan untuk memantau orang-orang yang memiliki kontak dekat dengan pasien terkonfirmasi selama 21 hari sejak paparan terakhir. Pencegahan melibatkan vaksinasi cacar (*smallpox*), praktik kebersihan yang baik, dan pendidikan kesehatan untuk meningkatkan kesadaran tentang gejala dan pencegahan *monkeypox* di masyarakat. Panduan ini penting untuk mengidentifikasi, mengelola, dan mencegah penyebaran cacar monyet di komunitas.

Pada 20 Agustus 2022, juru bicara Kementerian Kesehatan Republik Indonesia (Kemenkes RI), Dr. Mohammad Syahril, mengumumkan kasus cacar monyet pertama yang terkonfirmasi di Indonesia. Seorang pasien berusia 27 tahun asal DKI Jakarta yang pernah melakukan perjalanan ke luar negeri mengeluhkan gejala cacar monyet berupa ruam di wajah, telapak tangan, dan kaki, serta menyebar ke area genital. Setelah menjalani pemeriksaan, pasien diharuskan menjalani perawatan dengan isolasi mandiri. Jumlah total suspek cacar monyet per tanggal 20 Agustus 2022 sebanyak 22 suspek yang saat ini sudah dipulangkan yang tersebar di Jawa Tengah, Pontianak, Banten, Riau, dan 1 kasus positif cacar monyet di DKI Jakarta yang dilaporkan melalui konferensi pers (KemenkesRI) (20/8/2022).

Dampak nyata dari endemi adalah meningkatnya jumlah orang terinfeksi virus cacar monyet yang berdampak pada permintaan pasokan di sektor kesehatan (fasilitas khusus; ruang isolasi di rumah sakit dan layanan dari tenaga medis) dan sektor farmasi untuk merawat pasien cacar monyet dan memproduksi obat-obatan yang dapat dikonsumsi oleh pasien untuk mengurangi rasa sakit yang disebabkan oleh virus yang menjadi inti dari cacar monyet.

Penelitian mengenai hubungan evolusi yang terkait dengan cacar monyet dan variannya masih berlangsung dan belum selesai. Inisiatif Global untuk Berbagi Data Flu Burung (GISAID), sebuah database nirlaba dari organisasi genom virus dari beberapa negara yang bersifat terbuka (*open-access*) (Shu & McCauley, 2017). GenBank juga dapat menjadi referensi untuk mencari bahan penelitian yang berkaitan dengan genom virus cacar monyet. GenBank, sebuah platform akses terbuka yang menyediakan sekuens nukleotida yang dapat diakses oleh publik. GenBank juga menyediakan anotasi biologi dan bibliografi untuk memudahkan para peneliti atau mereka yang memiliki akses untuk memahami nukleotida nukleotida (Benson et al., 2015). Para peneliti yang tertarik untuk mempelajari cacar monyet dapat mengakses GISAID dan GenBank. Indonesia adalah salah satu negara yang mendistribusikan sekuens genom utuh dari berbagai daerah/provinsi/kota/kabupaten yang dapat digunakan oleh para peneliti Indonesia untuk menelusuri atau mengkorelasikan hubungan evolusi antar varian. Salah satu lembaga yang berperan penting dalam mendistribusikan sekuens genom utuh Indonesia ke GISAID adalah Lembaga Biologi Molekuler Eijkman. Selain itu, laboratorium dan institusi lain juga mendistribusikan sekuens genom utuh Indonesia ke GISAID. Genom utuh Indonesia yang telah didistribusikan ke GISAID saat ini berjumlah 8229 virus.

Seiring berjalannya waktu, pada era ini sudah memasuki Revolusi Industri 4.0. Sebuah era dimana perkembangan *domain Science, Technology, Engineering, and Matematika* (STEM). Contoh nyatanya adalah adanya kecerdasan buatan (*Artificial Intelligence*), *Big Data*, *Internet of Things (IoT)*, dan lain-lain. Perkembangan ini Perkembangan pesat ini diimbangi dengan masalah-masalah yang membutuhkan teknologi untuk memecahkan dan membantu para peneliti atau masyarakat. Teknologi untuk menyelesaikan dan membantu peneliti atau

masyarakat menyelesaikan permasalahan yang ada. Semua bidang ilmu disatukan untuk memecahkan suatu masalah. Seperti biologi komputasi yang menerapkan ilmu komputer, biologi, statistika, matematika terapan, kimia, biokimia, biokimia, biologi molekuler, genetika, biofisika, genomik, evolusi dan ekologi dalam memecahkan masalah biomedis dan ilmu hayati.

Dengan perkembangan biologi komputasi, para peneliti dapat berperan dalam meneliti, menganalisis, dan memberikan hipotesis terkait perkembangan virus cacar monyet yang menyebabkan *Monkey Pneumonia*. Dalam penelitian ini, peneliti berupaya mengimplementasikan biologi komputasi dalam filogenetika molekuler untuk melacak mutasi mutasi pada berbagai *clade Monkeypox*. Untuk analisis filogenetika, digunakan metode *maximum likelihood* dikarenakan memiliki berbagai keunggulan diantaranya, bisa digunakan untuk mencari model evolusi yang lebih kompleks dan representatif. Metode ini dapat mempertimbangkan berbagai aspek proses evolusi, seperti perbedaan laju substitusi *nukleotida*, serta memberikan estimasi probabilistik terhadap pohon filogenetik. Hal ini memungkinkan penilaian tingkat kepercayaan terhadap pohon yang dihasilkan, sehingga hasil analisis menjadi lebih akurat dan informatif. Metode ini menggunakan informasi dari semua situs dalam *alignment* secara lengkap dan bersamaan, sehingga lebih efisien dalam memanfaatkan data yang tersedia. *Maximum likelihood* bisa memungkinkan penggunaan berbagai model evolusi yang bisa disesuaikan dengan karakteristik data yang dianalisis. Hal ini memungkinkan analisis yang lebih akurat karena model yang digunakan bisa disesuaikan dengan kenyataan biologis serta mempermudah mencari jenis kekerabatan dari varian virus *monkeypox*. Maka dari itu diharapkan penelitian ini dapat memberikan gambaran tentang perkembangan varian virus dari MPXV (*Monkeypox*).

1.2. Rumusan Masalah

Berdasarkan latar belakang tersebut, maka permasalahan yang akan dikaji dalam penelitian ini adalah sebagai berikut:

1. Bagaimana distribusi epidemiologi dan pola penyebaran MPXV (*monkeypox*) dapat dijelaskan melalui analisis berdasarkan variabel gender, clade virus per negara, kategori umur, serta riwayat medis pasien?

2. Bagaimana hubungan evolusi antar genom virus MPXV (*Monkeypox*) di berbagai negara serta perkembangan mutasinya dapat dianalisis melalui rekonstruksi pohon filogenetika dengan penerapan metode *maximum likelihood* untuk mengidentifikasi pola evolusioner virus tersebut?

1.3. Batasan Masalah

1. Data yang digunakan dalam penelitian ini adalah data sekuen DNA dari pasien *Monkeypox*.
2. Penelitian ini dilakukan dengan menggunakan program *MAFFT (Multiple Analysis Fast Furior)*, *iTOL (Interactive Tree of Life) v.5*, *Jalview*, *Excel*, *Mega X* dan *PowerBI*.

1.4. Tujuan Penelitian

1. Mengetahui distribusi/penyebaran kasus *Monkeypox* melalui data-data pendukung seperti (gender, umur, negara, riwayat pasien, *clade* dan *sub clade*).
2. Mengetahui hubungan evolusi antara genom virus dan perkembangan varian-varian MPXV melalui hasil rekonstruksi pohon filogenetika dengan menerapkan metode *maximum likelihood* dalam filogenetika molekuler. Memberikan wawasan biologis melalui genom virus yang diteliti mengenai varian cacar monyet sehingga dapat dipahami oleh masyarakat global.

1.5. Manfaat Penelitian

Manfaat yang diharapkan Penulis dalam penelitian ini adalah sebagai berikut.

1. Dapat membantu masyarakat memberikan pemahaman yang lebih baik tentang bagaimana virus *Monkeypox* berevolusi dari waktu ke waktu.
2. Memberikan kontribusi pada pemahaman, tentang evolusi virus secara umum dan metode analisis filogenetika. Hasil-hasilnya dapat menjadi bahan referensi untuk penelitian lebih lanjut dalam bidang *virologi*, *filogenetika*, *genomic medicine* dan *epidemiologi*.

BAB II

TINJAUAN PUSTAKA

Penelitian tentang analisis filogenetik yang terkait dengan *MonkeyPox* telah dilakukan oleh beberapa peneliti sebelumnya. Salah satunya penelitian yang cukup mirip walau berbeda objek yang berjudul "*Full-length genome* panjang genom dan analisis filogenetik Covid-19 galur dari Yogyakarta dan Jawa Tengah, Indonesia" pada tahun 2020". Penelitian ini bertujuan untuk mengetahui hubungan evolusi antara sampel virus yang diperoleh peneliti dengan sampel virus SARS-CoV-2. Covid-19 dari negara lain dengan memanfaatkan analisis filogenetika. Peneliti menggunakan metode *maximum-likelihood* dan menggunakan model substitusi DNA (GTR+G+I) dengan 1000 kali pengulangan *bootstrapping* yang dapat membantu peneliti untuk menentukan parameter apa saja yang mempengaruhi perkembangan SARS-CoV-2 di Provinsi Jawa Tengah dan Daerah Istimewa Yogyakarta (DIY). Penelitian ini juga menggunakan MAFFT untuk penyelarasan sekuen genom virus yang akan dianalisis. Dari hasil penelitian tersebut, dapat terlihat bahwa para peneliti menemukan fakta-fakta seperti: 1) Mutasi S-D614G mendominasi sampel Indonesia, 2) Adanya P323L pada clade GH dan 3) Terdapat lima clade yang terdeteksi di Indonesia hingga September 2020, yaitu L, G, O, GR, dan GH. (tabel) (Lumongga, 2022).

Penelitian yang telah dilakukan oleh L. L. CAVALLI-SFORZA AND A. W. F. EDWARDS (1967) mereka mencoba memahami jalur evolusi dengan menggunakan model yang menganggap evolusi seperti gerakan acak bercabang (*Brownian-motion process*). Model ini berlaku jika perbedaan antar populasi hanya disebabkan oleh penyimpangan genetik acak dan seleksi yang bervariasi. Dalam model tersebut, populasi tidak punah, tetapi pecah menjadi dua populasi baru secara acak yang identik dengan induknya. Tantangannya adalah memperkirakan bentuk dan ukuran pohon evolusi yang paling mungkin. Karena metode *maximum likelihood* sulit diterapkan, pendekatan alternatif digunakan, dan sebuah contoh disediakan untuk menunjukkan cara kerja serta batasannya.

Penelitian yang berjudul "Distribusi Covid-19 dan Konstruksi Pohon Filogenetika *MonkeyPox* di Indonesia" dilakukan oleh Turista et al. pada tahun 2020. Tujuan dari penelitian ini adalah untuk mendeskripsikan distribusi kasus

cacar monyet di Indonesia dan merekonstruksi pohon filogenetik untuk mengetahui hubungan antara SARS-CoV-2 dari isolat Indonesia dengan isolat dari beberapa negara lain. studi kasus ini diperoleh dari GISAID dan GenBank. Alat yang digunakan untuk merekonstruksi pohon filogenetik dari studi kasus ini adalah *MEGA X (Molecule Evolutionary Genetics Analysis)*. Metode yang digunakan adalah *maximum-likelihood* dengan 1000 kali pengulangan bootstrapping menggunakan model *substitusi* Tamura-Nei. Berdasarkan hasil penelitian dengan studi kasus yang digunakan oleh peneliti, terdapat beberapa provinsi di Indonesia yang memiliki angka kematian tertinggi yaitu Provinsi Daerah Khusus Ibukota (DKI) Jakarta, Jawa Barat, dan Jawa Timur. Sementara itu, Provinsi Kepulauan Riau, Bali, dan Gorontalo adalah merupakan provinsi dengan tingkat kesembuhan tertinggi. Isolat *MonkeyPox* dari Indonesia memiliki kekerabatan yang sangat dekat dengan isolat *MonkeyPox* dari negara lain. Isolat yang dianalisis memiliki genom yang sangat mirip meskipun ada sedikit perubahan genetik dalam proses penularan.

Studi berjudul "*Analisis Filogenetik Virus Flu Burung H5N1 Menggunakan Metode Maksimum likelihood*" yang dilakukan oleh Liu et al. pada tahun 2021 menunjukkan bagaimana alat bioinformatika digunakan untuk penyelarasan dan analisis filogenetik. Dalam penelitian tersebut, Liu et al. menggunakan MAFFT (v.7.475) dengan parameter default untuk melakukan proses penyelarasan sekuen nukleotida yang diperoleh dari database GenBank. Selain itu, pensejajaran protein dilakukan menggunakan Clustal Omega dengan konfigurasi default untuk memastikan keseragaman hasil. Untuk membangun pohon filogenetik, para peneliti menggunakan RAxML (v.8.2.12) dengan metode (Maximum Likelihood). Model substitusi yang digunakan adalah GTR + G + I, dan analisis bootstrap dilakukan sebanyak 1000 kali untuk memastikan reliabilitas cabang pohon filogenetik yang dihasilkan. Hasil penelitian menunjukkan bahwa virus H5N1 dari wilayah Asia Tenggara memiliki kekerabatan genetik yang tinggi dengan strain yang ditemukan di China, dengan tingkat kesamaan sekuen mencapai 98,7%. Penelitian ini menegaskan pentingnya MAFFT sebagai alat penyelarasan sekuen yang efisien dalam analisis filogenetik untuk memahami pola evolusi virus

Studi lain oleh Happi et al. (2022) menunjukkan bahwa penggabungan data filogenetika dengan informasi epidemiologi, seperti riwayat perjalanan pasien,

sangat penting untuk menentukan pola penyebaran virus. Analisis ini menggunakan data genomik dari 212 *strain MPXV*, menunjukkan bahwa wabah 2022 memiliki hubungan erat dengan strain dari Nigeria pada 2017–2019. Pohon filogenetik yang dihasilkan menunjukkan adanya diversifikasi dalam clade IIB, yang mendukung hipotesis bahwa wabah baru merupakan perluasan dari wabah lokal sebelumnya.

Penelitian berjudul "Analisis bioinformatika sekuen genom utuh SAR-COV2 dari Indonesia" yang dilakukan pada tahun 2021 oleh Ulfah dan Helianti. Penelitian ini bertujuan untuk melaporkan variasi genom SAR-COV2, terutama mutasi pada protein spike dan protease yang tersebar di Indonesia. Sedangkan nukleotida nukleotida disejajarkan dengan menggunakan MAFFT versi 7. Selain itu, analisis filogenetik dari penelitian ini memberikan gambaran dari berbagai daerah di Indonesia yang secara umum dikelompokkan ke dalam dua kelas utama, yaitu: G dan L. Klade yang mendominasi berdasarkan studi kasus penelitian ini adalah Clade G (meliputi kelompok GH dan GR dengan persentase sebesar 59%). Daerah-daerah di Indonesia yang didominasi oleh sampel yang diklasifikasikan sebagai clade G adalah: Bandung, Sidoarjo, Yogyakarta, Jakarta, Surabaya, Jawa Tengah, dan Jawa Barat.

Selain itu, ada juga Robert Lanfear et al (2016) melakukan penelitian berjudul "*PartitionFinder 2: New Methods for Selecting Partitioned Models of Evolution for Molecular and Morphological Phylogenetic Analyses*" dari hasil *research* tersebut ialah Fitur baru *PartitionFinder 2* meliputi algoritme yang lebih cepat seperti *k-means* dan *relaxed-clustering*, model evolusi terkini seperti LG4X, pohon awal berbasis *Maximum Likelihood*, kemampuan menganalisis dataset morfologi, dan format keluaran baru untuk meningkatkan kompatibilitas perangkat lunak. Peningkatan efisiensi lainnya mencakup parser penyelarasan baru, pemanfaatan prosesor ganda, dan pengurangan file yang dibaca/digunakan selama analisis, yang memungkinkan pengolahan dataset besar lebih cepat. Yang mana dari hal itu, *PartitionFinder 2* jauh lebih cepat dibandingkan versi sebelumnya, terutama untuk dataset besar. Misalnya, analisis dataset serangga selesai dalam 35 jam menggunakan versi 2.0.0, sementara versi 1.1.1 hanya menyelesaikan kurang dari 1% dalam waktu yang sama.

Penelitian oleh Isidro et al. (2022) menganalisis filogenetika *strain Monkeypox* dari berbagai negara selama wabah global. Dengan menggunakan metode *Maximum Likelihood*, analisis ini mengungkapkan bahwa garis keturunan B.1 (clade IIb) adalah yang dominan dalam wabah 2022. Hal ini juga mengidentifikasi mutasi genetik spesifik yang memengaruhi protein penting dalam *genome monkeypox*. Temuan ini memberikan indikasi bahwa adaptasi genetik mungkin berperan dalam peningkatan penyebaran virus.

Dalam sebuah studi pada tahun 2014 di Institute of Theoretical Informatics, Karlsruhe Institute of Technology, Jerman ada seorang peneliti yang bernama Alexandros Stamatakis dengan penelitiannya yang berjudul “*RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies*” dari hasil tersebut menganalisis *phylogenetics* dengan RAxML (*Randomized Axelerated Maximum Likelihood*) atau sebuah program populer untuk analisis filogenetik dataset besar dengan metode *maximum likelihood* yang mana ntuk menangani dataset yang semakin besar dan memenuhi kebutuhan penelitian, maka dari itu mencakup banyak fitur baru, seperti penambahan model substitusi, dukungan untuk jenis data baru, teknik untuk mengurangi penggunaan memori, dan alat untuk analisis lanjutan pada kumpulan pohon filogenetika.

Pada studi bidang *genomic medicine* di university of oxford , penelitian dilakukan oleh Andrew Rambaut et al. (2018). Inferensi filogenetika Bayesian menggunakan rantai *Markov Monte Carlo* (MCMC) sangat penting dalam memahami sejarah evolusi dari data urutan molekul. Untuk menganalisis hasil dari MCMC harus menggunakan perangkat lunak yang digunakan adalah Tracer (versi 1.7). Tracer memungkinkan pengguna untuk melihat dan menganalisis jejak MCMC dari inferensi filogenetik Bayesian. Fitur yang disediakan oleh Tracer termasuk estimasi kepadatan kernel, visualisasi multivariat, rekonstruksi lintasan demografi, dan ringkasan distribusi posterior bersyarat.

Dalam studi filogenetika *bioinformatics* pada tahun 2021, Ivica Letunic dan Peer Bork melakukan analisis filogenetika dengan tujuan memper upgrade hasil dari pohon filogenetika dengan menggunakan (*iTOL*) v5, tools tersebut bisa menghadirkan tampilan pohon menjadi lebih menarik dan penambahan beberapa

fitur seperti menambahkan anotasi manual dengan alat menggambar langsung, memberi label, atau menggambar bentuk pada pohon.

Penelitian ini akan mengimplementasikan metode parsimony dalam analisis garis keturunan *monkeypox*, dengan sampel dari Indonesia dan negara lain. Metode ini berbasis karakter, memberikan solusi yang efisien untuk menganalisis data genetik dengan tingkat kompleksitas tinggi. *Maximum parsimony* memiliki kelebihan dalam menghasilkan pohon yang mudah diinterpretasi, meski memerlukan validasi tambahan seperti *bootstrapping* untuk meningkatkan keandalan pohon filogenetik.

Beberapa penelitian sebelumnya telah dilakukan oleh para peneliti tentang analisis molekuler filogenetika untuk membangun pohon filogenetika dengan *sequence monkeypox* dari berbagai sampel sebagai studi kasus. Penulis menggunakan penelitian-penelitian berikut sebagai referensi dalam melakukan penelitian ini.

Tabel 2.1 Daftar Penelitian Sebelumnya

Tahun	Nama	Judul	Hasil Penelitian
2014	Alexandros Stamatakis	<i>RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies</i>	Dari hasil tersebut menganalisis <i>phylogenetics</i> dengan RAxML (<i>Randomized Axelerated Maximum Likelihood</i>) atau sebuah program populer untuk analisis filogenetik dataset besar dengan metode <i>maximum likelihood</i> yang mana ntuk menangani dataset yang semakin besar dan memenuhi kebutuhan penelitian, maka dari itu mencakup banyak fitur baru, seperti penambahan model substitusi, dukungan untuk jenis data baru, teknik untuk mengurangi penggunaan

			memori, dan alat untuk analisis lanjutan pada kumpulan pohon filogenetika.
Tahun	Nama	Judul	Hasil Penelitian
2018	Andrew Rambaut et al	<i>Analysis phylogenetics use Markov Monte Carlo (MCMC)</i>	Pada studi bidang <i>genomic medicine</i> di university of oxford , penelitian dilakukan oleh Andrew Rambaut et al. 2018. Inferensi filogenetika Bayesian menggunakan rantai Markov Monte Carlo (MCMC) sangat penting dalam memahami sejarah evolusi dari data urutan molekul. Untuk menganalisis hasil dari MCMC harus menggunakan perangkat lunak yang digunakan adalah Tracer (versi 1.7). Tracer memungkinkan pengguna untuk melihat dan menganalisis jejak MCMC dari inferensi filogenetik Bayesian. Fitur yang disediakan oleh Tracer termasuk estimasi kepadatan kernel, visualisasi multivariat, rekonstruksi lintasan demografi, dan ringkasan distribusi posterior bersyarat.
2020	Liu et al.	<i>Phylogenetic Analysis of the H5N1 Avian Influenza Virus Using the Maximum Parsimony Method</i>	Hasil penelitian menunjukkan bahwa virus H5N1 dari wilayah Asia Tenggara memiliki kekerabatan genetik yang tinggi dengan strain yang ditemukan di China, dengan tingkat kesamaan sekuen mencapai 98,7%. Penelitian ini menegaskan pentingnya MAFFT sebagai alat penyelarasan sekuen yang efisien dalam analisis filogenetik untuk memahami pola evolusi virus.
2020	Turista et al.	<i>Distribution of Covid-19 and Phylogenetic Tree Construction of SARS-CoV-2 in Indonesia</i>	Beberapa provinsi di Indonesia dengan angka kematian tertinggi yaitu Provinsi Daerah Khusus Ibukota (DKI) Jakarta, Jawa Barat, dan Timur

			Jawa. Provinsi Kepulauan Riau, Bali, dan Gorontalo adalah beberapa di antaranya provinsi dengan kesembuhan tertinggi tarif. Isolat SARS-CoV-2 dari Indonesia mempunyai kerabat yang sangat dekat dengan Isolasi SARS-CoV-2 dari yang lain negara. Isolat yang dianalisis memiliki genom yang sangat mirip meskipun ada sedikit perubahan genetik di dalamnya proses transmisi.
Tahun	Nama	Judul	Hasil Penelitian
2020	Gunadi et al	Full-length genome characterization and phylogenetic analysis of SARS-CoV-2 virus strains from Yogyakarta and Central Java Indonesia	Penelitian ini bertujuan untuk mengetahui hubungan evolusi antara sampel virus yang diperoleh peneliti dengan sampel virus SARS-CoV-2. Peneliti menggunakan metode <i>maximum-likelihood</i> dan menggunakan model substitusi DNA (GTR+G+I) dengan 1000 kali pengulangan <i>bootstrapping</i> yang dapat membantu peneliti untuk menentukan parameter apa saja yang mempengaruhi perkembangan SARS-CoV-2 di Provinsi Jawa Tengah dan Daerah Istimewa Yogyakarta (DIY). Penelitian ini juga menggunakan MAFFT untuk penyelarasan sekuen genom virus yang akan dianalisis. Dari hasil penelitian tersebut, peneliti menemukan fakta-fakta seperti: 1) Mutasi S-D614G mendominasi sampel Indonesia, 2) Adanya P323L pada clade GH dan 3) Terdapat lima clade yang terdeteksi di Indonesia hingga September 2020, yaitu L, G, O, GR,
2020	Forster et al	<i>Phylogenetic network analysis of the SARS-CoV-2 genome.</i>	Hasil penelitian menunjukkan bahwa perbedaan sejarah migrasi, peristiwa pendirian, dan ukuran sampel dapat mempengaruhi pola filogeografi.

			Kehadiran riwayat perjalanan pasien dapat membantu dalam menentukan pola filogenetik dari SARS-CoV-2.
Tahun	Nama	Judul	Hasil Penelitian
2021	Ivica Letunic dan Peer Bork	<i>Analysis phylogenetics with (iTOL) v5.</i>	Dari hasil penelitian ini, analisis filogenetika dengan tujuan memper upgrade hasil dari pohon filogenetika dengan menggunakan (<i>iTOL</i>) v5, tools tersebut bisa menghadirkan tampilan pohon menjadi lebih menarik dan penambahan beberapa fitur seperti menambahkan anotasi manual dengan alat menggambar langsung, memberi label, atau menggambar bentuk pada pohon.
2021	Ulfah dan Helianti	Analisis bioinformatika sekuen genom utuh SAR-COV2 dari Indonesia"	Penelitian ini bertujuan untuk melaporkan variasi genom SAR-COV2, terutama mutasi pada protein spike dan protease yang tersebar di Indonesia. Sedangkan nukleotida disejajarkan dengan menggunakan MAFFT versi 7. Selain itu, analisis filogenetik dari penelitian ini memberikan gambaran dari berbagai daerah di Indonesia yang secara umum dikelompokkan ke dalam dua kelas utama, yaitu: G dan L. Klade yang mendominasi berdasarkan studi kasus penelitian ini adalah

			Clade G (meliputi kelompok GH dan GR dengan persentase sebesar 59%. Daerah-daerah di Indonesia yang didominasi oleh sampel yang diklasifikasikan sebagai clade G adalah: Bandung, Sidoarjo, Yogyakarta, Jakarta, Surabaya, Jawa Tengah, dan Jawa Barat.
Tahun	Nama	Judul	Hasil Penelitian
2022	Happi et al	<i>Phylogenetics with Epidemiological information.</i>	Analisis ini menggunakan data genomik dari 212 starin MPXV, Menunjukkan wabah pada tahun 2022 memiliki hubungan erat dengan strain dari Nigeria pada tahun 2017-2019. Pohon filogenetika yang dihasilkan menunjukkan adanya diversifikasi dalam clade Iib, yang mendukung hipotesis bahwa wabah baru merupakan perluasan dari wabah lokal sebelumnya.
2022	Lincon Mazumder, Md. Rakibul Hasan, Kanij Fatema	<i>Identification of B and T Cell Epitopes for Designing Epitope-Based Vaccines Epitope-Based Peptide Vaccines against Cell Surface Binding Proteins of the Monkey Pox Virus: An Immunoinformatics Study</i>	Hasil. Analisis mengungkapkan bahwa protein pengikat permukaan sel dari virus cacar monyet mengandung 30 sel B dan 19 epitop sel T dalam parameter yang diberikan. Di antara epitop sel T, epitope "ILFLMSQRY" ditemukan sebagai salah satu kandidat vaksin peptida yang paling potensial. Analisis docking mengungkapkan sebuah afinitas pengikatan yang sangat baik dari epitop ini dengan reseptor manusia HLA-B*15:01 dengan energi pengikatan yang sangat rendah (-7,5 kkal / mol).

Tahun	Nama	Judul	Hasil Penelitian
2022	Hoang et al.	<i>MP Boot for Algorithm Parsimony</i>	Algoritma <i>parsimony</i> dapat dipercepat menggunakan optimasi algoritmik seperti bootstrap dan pendekatan hill-climbing. Studi ini relevan dalam menganalisis data dengan ukuran besar, seperti <i>genome</i> patogen selama wabah global. Hasilnya memberikan pola penyebaran virus dalam populasi yang terhubung secara geografis.
2022	Isidro et al.	<i>Analyzed the phylogenetics of Monkeypox strains from different countries during a global outbreak , by using the Maximum Likelihood method.</i>	Analisis ini mengungkapkan bahwa garis keturunan B.1 (Clade Iib) adalah yang dominan dalam wabah 2022. Penelitian ini juga mengidentifikasi mutasi genetik spesifik yang memengaruhi protein penting dalam monkeypox. Dalam hal ini memberikan indikasi bahwa adaptasi genetik yang mungkin berperan dalam peningkatan penyebaran virus.
2022	S. N. Shchelkunov et al	<i>Genome Analysis of Monkey Pox Virus</i>	Hasil penelitian menunjukkan virus cacar monyet (MPV) termasuk dalam genus <i>orthopoxvirus</i> dari keluarga <i>Poxviridae</i> , endemik di beberapa bagian Afrika, dan menyebabkan penyakit pada manusia yang menyerupai cacar. Genom MPV 196.858 bp dianalisis sehubungan dengan struktur fitur dan bingkai pembacaan terbuka. Setiap ujung genom mengandung terminal 6379-bp yang identik tetapi berorientasi berlawanan pengulangan terbalik, yang mirip dengan virus orthopoxvirus lainnya, termasuk urutan resolusi telomer yang diduga dan pendek pengulangan tandem. Analisis berbantuan komputer digunakan untuk mengidentifikasi 190 bingkai bacaan terbuka yang

Tahun	Nama	Judul	Hasil Penelitian
2022	Rokshana Parvin, Ahmed Ali et al.	<i>Monkey pox virus: A comprehensive review of taxonomy, evolution, epidemiology, diagnosis, prevention, and control regimens.</i>	mengandung 60 asam amino residu. Hasil Penelitian menunjukkan bahwa virus cacar monyet (MPXV) muncul kembali dan menginfeksi lebih dari 780 kasus manusia di lebih dari 20 negara di seluruh dunia pada saat penulisan ulasan ini. Telah ditunjukkan sebelumnya bahwa tidak hanya monyet yang berperan dalam penularan MPXV ke manusia, tetapi juga hewan pengerat (<i>Cricetomys gambianus</i> , dan <i>Graphiurus murinus</i>) dan tupai (<i>Heliosciurus spp</i> , <i>Funisciurus spp</i>). Selain itu, tindakan antropogenik, seperti penggundulan hutan, perubahan iklim perubahan iklim, interaksi antara hewan dan manusia, globalisasi, dan penghentian vaksinasi cacar, dapat berkontribusi pada kemunculan kembali MPXV. Di sini, kami meninjau nomenklatur saat ini, epidemiologi, keragaman genetik virus cacar, khususnya MPXV, dan rejimen kontrol, termasuk antivirus dan vaksin yang tersedia. Kami selanjutnya menyoroti keterkaitan genetik dari virus MPXV yang muncul untuk mengetahui garis keturunan virus menggunakan analisis filogenetik.
2022	Lumongga	<i>The Implementasion Of Computational Statistics in Molecular Phylogenetics to Identify Lineage Diversity Thorough Viral Genome Sequences</i>	Penelitian ini menggunakan 60 genom virus dari berbagai negara, termasuk Indonesia, untuk meneliti hubungan evolusi dan perkembangan varian SARS-CoV-2 melalui analisis filogenetik. Metode maximum like-lihood digunakan dengan perangkat lunak seperti Jalview, MAFFT, MEGA-X, RAxML

			<p>GUI 2.0, dan iTOL v.5, dengan data sekunder dari GISAID. Analisis menunjukkan bahwa Clade L, S, dan O serta Clade G dan GV terkait dalam cluster tertentu, sementara cluster lainnya termasuk GH dan GR. Ciri khas Clade GK juga terdapat pada Clade O, keduanya merupakan bagian dari varian VOC Delta. Clade L dapat digunakan sebagai referensi untuk sampel dengan clade yang berbeda, sementara Clade G dapat dianggap sebagai induk dari clade lainnya seperti GK, GR, GV, dan lainnya. Beberapa sampel menunjukkan keturunan dari Indonesia, menimbulkan kemungkinan kasus lokal, impor, atau kasus impor yang kemudian menyebar.</p>
--	--	--	---

Beberapa tinjauan pustaka di atas menjadi referensi bagi penulis untuk mengembangkan penelitian ini. Dari berbagai penelitian yang telah dilakukan, metode *maximum likelihood* digunakan untuk menganalisis hubungan evolusi analisis filogenetika selain itu juga bisa digunakan untuk menyusun pohon filogenetika berdasarkan data molekuler. *Maximum likelihood* (ML) mengadopsi model probabilistik yang mempertimbangkan proses evolusi secara lebih rinci, memberikan hasil yang lebih akurat meski dengan beban komputasi yang lebih tinggi dengan mempertimbangkan model evolusi molekuler, sehingga dapat memperkirakan hubungan evolusi berdasarkan probabilitas tertinggi yang sesuai dengan data sekuens genetik. Metode ini sering digunakan untuk meningkatkan ketepatan inferensi filogenetika, terutama dalam studi yang melibatkan data molekuler dalam jumlah besar, seperti data yang akan dilakukan pada penelitian ini dengan data korban pasien *monkeypox*, dari 106 *sequence*. Dengan tingkat mutasi yang tinggi dan kompleksitas evolusi virus *Monkeypox* menuntut pendekatan filogenetika yang lebih cermat dalam merekonstruksi hubungan antar sekuens. Dalam analisis filogenetika berbasis 106 sekuens virus *Monkeypox*, metode *Maximum Likelihood Estimation* (MLE) menjadi pilihan utama karena

kemampuannya dalam menangani dinamika substitusi nukleotida dengan lebih akurat dibandingkan metode lain seperti *Maximum Parsimony*. MLE bekerja dengan mengevaluasi kemungkinan berbagai pohon filogenetika berdasarkan model evolusi tertentu, seperti *General Time Reversible (GTR)*, sehingga dapat memperhitungkan perbedaan laju substitusi antar nukleotida, bias frekuensi basa, serta variasi tingkat evolusi antar situs dalam genom virus. Selain itu, MLE lebih andal dalam mengatasi berbagai tantangan analisis filogenetika, termasuk *homoplasies* di mana mutasi yang sama dapat muncul secara independen pada garis keturunan yang berbeda serta *long-branch attraction (LBA)*, yang sering menjadi masalah dalam metode *Parsimony*. Dengan memanfaatkan model probabilistik, MLE memungkinkan rekonstruksi panjang cabang yang lebih akurat, yang pada akhirnya memberikan estimasi waktu divergensi yang lebih realistis untuk memahami evolusi virus *Monkeypox*.

Keunggulan lain dari pendekatan MLE adalah kemampuannya untuk dikombinasikan dengan analisis bootstrap, yang memberikan tingkat kepercayaan pada topologi pohon filogenetika yang dihasilkan. Selain itu, penggunaan *Likelihood Ratio Test (LRT)* memungkinkan pemilihan model evolusi terbaik yang paling sesuai dengan dataset yang digunakan. Hal ini sangat penting mengingat virus seperti *Monkeypox* dapat mengalami seleksi alam yang beragam serta tingkat mutasi yang berbeda-beda pada berbagai bagian genomnya. Dengan mempertimbangkan faktor-faktor tersebut, penerapan MLE dalam analisis filogenetika virus *Monkeypox* dengan 106 sekuens memberikan pendekatan yang lebih robust, akurat, dan dapat diandalkan dalam memahami dinamika evolusi virus, mendeteksi pola mutasi, serta mengidentifikasi hubungan genetik antar isolat virus yang dapat digunakan untuk studi epidemiologi molekuler dan pengendalian wabah di masa depan.

BAB III

LANDASAN TEORI

3.1 Biologi Komputasi

Biologi komputasi adalah cabang ilmu pengetahuan alam yang menerapkan berbagai disiplin ilmu seperti ilmu komputer, statistik, biostatistik, basis data, dan matematika untuk menyatukan, mengelola, serta menganalisis koleksi besar informasi yang berkaitan dengan biologi, kesehatan, dan medis. Sumber informasi yang digunakan dapat mencakup data statistik pasien, jurnal ilmiah, spesimen jaringan, studi genetik dan molekuler, serta uji klinis. Bidang ini sering disebut sebagai bioinformatika (*National Cancer Institute*).

Sejarah biologi komputasi mencatat bahwa istilah "bioinformatika" pertama kali digunakan pada tahun 1968, sementara definisinya baru dirumuskan pada tahun 1978. Bidang ini mencakup dua fokus utama: (1) pengembangan perangkat lunak, algoritma, dan basis data, dan (2) analisis serta interpretasi data biologis menggunakan perangkat lunak dan algoritma tertentu (Zhang & Liu, 2013). Berbeda dari bioinformatika, yang cenderung fokus pada pengelolaan dan analisis data biologis, biologi komputasi lebih banyak berkaitan dengan pemodelan dan simulasi sistem biologis. Ruang lingkup biologi komputasi meliputi berbagai cabang ilmu biologi, termasuk genomik, proteomik, biologi sel, genetika, dan evolusi molekuler. Data yang sering digunakan antara lain urutan DNA, urutan asam amino, serta struktur dan fungsi protein. Implementasi biologi komputasi dapat diwujudkan dalam berbagai cara, seperti:

1. Merekonstruksi pohon filogenetik untuk mempelajari hubungan evolusi antarspesies.
2. Membuat alat berbasis web atau desktop untuk memecahkan masalah biologis.
3. Menggunakan metode pembelajaran mesin (*machine learning*) untuk menganalisis ekspresi gen.

Bioinformatika dan biologi komputasi menjadi disiplin ilmu yang sangat penting dalam era big data di bidang biologi.

3.2 Filogenetika

Pohon filogenetika adalah sebuah diagram yang digunakan untuk menggambarkan hubungan evolusi antara berbagai gen dan organisme dalam suatu hubungan kekerabatan yang erat. Hal tersebut berfungsi sebagai alat untuk memahami bagaimana suatu kelompok organisme atau gen berevolusi dari nenek moyang yang sama, serta bagaimana mereka bercabang menjadi spesies atau varian yang berbeda dengan seiring berjalannya waktu. Disebut sebagai pohon filogenetika karena bentuknya menyerupai struktur pohon bercabang, di mana setiap elemen dalam diagram ini memiliki kesamaan dengan bagian-bagian pohon yang nyata. Dalam representasi, akar pohon menggambarkan nenek moyang bersama yang menjadi asal dari seluruh taksa dalam diagram, sedangkan cabang-cabangnya melambangkan jalur evolusi yang menghubungkan berbagai organisme atau gen dalam suatu garis keturunan. Selain itu, istilah-istilah yang digunakan dalam diagram ini mengacu pada berbagai bagian dari pohon, termasuk akar, cabang, node, dan daun, yang masing-masing memiliki peran penting dalam interpretasi hubungan evolusi. Salah satu komponen utama dalam pohon filogenetika adalah *external node* atau daun, yang merepresentasikan taksa yang sedang dipelajari. Taksa ini dikenal sebagai *Operational Taxonomic Units (OTUs)*, yang merupakan unit klasifikasi yang dapat mencakup berbagai tingkat hierarki biologis, seperti individu, spesies, strain virus, atau bahkan kelompok organisme yang lebih luas seperti famili atau genus. Dengan demikian, OTUs memungkinkan para peneliti untuk membandingkan berbagai jenis organisme atau entitas biologis yang memiliki hubungan evolusi tertentu, baik dalam satu spesies maupun antar spesies yang berbeda. Selain OTUs, ada pula *node internal*, yang berfungsi sebagai titik percabangan dalam pohon dan merepresentasikan leluhur bersama dari beberapa OTUs. Node ini sering disebut sebagai *Hypothetical Taxonomic Units (HTUs)* untuk menekankan bahwa mereka merupakan leluhur yang bersifat hipotetis dan tidak selalu memiliki representasi nyata dalam dunia biologis. HTUs berperan dalam menunjukkan titik divergensi di mana suatu kelompok organisme bercabang menjadi spesies atau varian yang berbeda, memberikan wawasan mengenai sejarah evolusi dan pola spesiasi yang telah terjadi selama jutaan tahun (Amiroch, 2020).

3.3 Asam Deoksisibonukleat (DNA)

DNA (Asam Deoksiribonukleat) adalah sejenis asam nukleat yang tergolong molekul utama yang menyimpan intruksi genetika setiap organisme dan virus. Pada manusia, DNA tersusun dalam 23 pasang kromosom di dalam nukleus, yang dilengkapi dengan genom mitokondria yang ukurannya lebih kecil. Genom diploid manusia, yang merupakan kombinasi kromosom dari kedua orang tua, memainkan peran penting dalam perkembangan normal dan keberlangsungan hidup. Penelitian awal terhadap DNA dimulai oleh Friedrich Miescher pada tahun 1869, diikuti oleh pengembangan istilah "asam nukleat" oleh Richard Altmann, dan identifikasi komponen utamanya—purin, pirimidin, gula, serta fosfat—oleh Albrecht Kossel.

Pada dekade 1930-an, ilmuwan berhasil mengidentifikasi empat basa nitrogen utama DNA dan menemukan keberadaan gula deoksiribosa sebagai salah satu komponennya. Molekul ini kemudian dinamai *asam deoksiribonukleat* (DNA). Sebuah penemuan penting oleh Erwin Chargaff menunjukkan adanya keteraturan jumlah basa nitrogen dalam DNA, yaitu kesetaraan antara sitosin (C) dan guanin (G), serta adenin (A) dan timin (T), yang bervariasi antarspesies. Sebagai contoh, genom manusia terdiri dari 20% sitosin, 20% guanin, 30% adenin, dan 30% timin (Minchin & Lodge, 2019).

Secara struktural, DNA berbentuk heliks ganda yang terdiri atas dua rantai nukleotida. Setiap nukleotida mengandung gugus fosfat, gula deoksiribosa, dan basa nitrogen (adenin, timin, sitosin, dan guanin). Basa nitrogen berpasangan melalui ikatan hidrogen secara spesifik, yaitu A dengan T dan G dengan C.

Fungsi utama DNA meliputi:

1. Replikasi: Kemampuan untuk menggandakan diri guna memastikan pewarisan informasi genetik selama pembelahan sel.
2. Transkripsi: Berfungsi sebagai cetakan dalam proses sintesis RNA, yang selanjutnya digunakan untuk produksi protein.
3. Mutasi: DNA dapat mengalami perubahan urutan, yang berkontribusi pada variasi genetik dan, dalam beberapa kasus, menyebabkan gangguan fungsional.

Dalam ranah penelitian, DNA menjadi subjek kajian penting dengan berbagai aplikasi praktis, seperti:

1. Diagnostik Medis: Deteksi mutasi genetik yang berhubungan dengan penyakit tertentu.
2. Forensik: Identifikasi individu berdasarkan profil DNA.
3. Bioteknologi: Pengembangan organisme transgenik serta produksi protein rekombinan.
4. Evolusi Molekuler: Analisis hubungan genetik antarspesies untuk memahami proses evolusi.

Kemajuan teknologi modern seperti pengurutan genom (*genome sequencing*) telah memungkinkan pemetaan genom secara komprehensif, sementara alat seperti CRISPR-Cas9 memberikan kemampuan untuk mengedit gen secara presisi. Penelitian tentang DNA tidak hanya memperdalam pemahaman tentang biologi molekuler tetapi juga membuka jalan bagi inovasi di bidang medis, forensik, dan bioteknologi.

3.4 Multiple Sequence Alignment

Penyelarasan sekuen *Multiple Sequence Alignment* (MSA) adalah salah satu tahapan terpenting dalam rekonstruksi pohon filogenetik, studi domain, dan aplikasi lainnya, yaitu penyejajaran sekuen (MSA). Penjajaran sekuen dapat diimplementasikan pada DNA, RNA, atau sekuens protein. Penjajaran sekuens mencakup sekuens seperti menemukan jawaban dalam bentuk algoritmik dengan penyetaraan yang berhubungan dengan evolusi dengan memperkirakan peristiwa evolusi seperti mutasi, insersi, penghapusan, dan penataan ulang dalam keadaan tertentu (Chatzou et al., 2016). Ada beberapa contoh metode penyelarasan sekuens seperti *ClustalW*, *T-Coffee*, *ProbCons*, *MUSCLE*, *MAFFT*, dan *Clustal Omega*.

Multiple Sequence Alignment (MSA) digunakan untuk menyelaraskan lebih dari dua sekuens DNA, RNA, atau protein secara bersamaan, guna mengidentifikasi kesamaan evolusi dan pola konservasi. MSA sangat penting dalam analisis filogenetika karena memungkinkan perbandingan langsung antar sekuens, yang membantu dalam rekonstruksi pohon filogenetik dan penentuan hubungan kekerabatan antar taksa.

3.5 MAFFT

Multiple Alignment Fast Fourier Transform (MAFFT) adalah salah satu metode dalam penyelarasan urutan. MAFFT dibuat dengan tujuan menangani besar (hingga sekitar 5.000 sekuens) dan panjang (sekitar 2.000 atau sekitar 5.000 nt) data pada PC desktop waktu standar.

Penjajaran sekuens berganda (MSA) adalah tugas yang paling umum dalam bioinformatika. *Multiple alignment fast Fourier transform* (MAFFT) adalah program MSA tercepat di antara program-program MSA lainnya yang keakuratan pensejajaran yang dihasilkan dapat dibandingkan dengan program MSA yang paling akurat. Dalam makalah ini, kami memodifikasi skema komputasi korelasi MAFFT untuk peningkatan efisiensi lebih lanjut dalam tiga aspek. Pertama, ekspresi asam amino dan nukleotida berbasis bilangan kompleks baru digunakan dalam korelasi yang dimodifikasi. Kedua, konvolusi linier dengan batasan diusulkan untuk menghitung korelasi asam amino dan urutan nukleotida. Ketiga, merancang algoritma *Fast Fourier Transformation* (FFT) untuk menghitung konvolusi linier. Algoritma FFT didasarkan pada FFT *split-radix* pasangan konjugasi dan tidak memerlukan permutasi orde, dan merupakan hal yang baru karena hanya bagian riil dari output akhir yang diperlukan. Hasil simulasi menunjukkan bahwa kecepatan skema yang dimodifikasi adalah 107,58 hingga 365,74 persen lebih cepat daripada MAFFT asli untuk satu kali eksekusi fungsi `Falign()` pada MAFFT, yang mengindikasikan realisasinya yang lebih cepat (Zheng et al., 2017).

3.6 *Concensus dan Occupancy*

Dalam analisis filogenetika, *Concensus* merujuk pada metode yang digunakan untuk mengintegrasikan berbagai pohon filogenetik yang dihasilkan dari analisis data sekuens DNA atau protein. Metode ini bertujuan untuk memperoleh representasi tunggal yang mencerminkan hubungan evolusi yang paling mungkin. Konsensus sangat penting dalam mengidentifikasi pola kekerabatan yang konsisten di antara berbagai pohon yang mungkin berbeda akibat variasi dalam data atau metode analisis yang digunakan.

Sementara itu, istilah *Occupancy* dalam konteks analisis filogenetika tidak umum digunakan. Namun, dalam analisis data sekuens, *occupancy* dapat merujuk

pada proporsi posisi dalam alignment sekuens yang ditempati oleh nukleotida atau asam amino tertentu. Faktor ini berperan dalam menentukan kelengkapan data yang digunakan dalam rekonstruksi filogenetik dan dapat memengaruhi akurasi pohon filogenetik yang dihasilkan.

Penelitian yang dilakukan oleh Darupamenang et al. (2022) menerapkan metode konsensus dalam analisis filogenetik tanaman *Alocasia Jacklyn* berdasarkan gen *rbcL*. Penerapan metode ini bertujuan untuk mendapatkan pohon filogenetik yang lebih stabil dan dapat diandalkan, sehingga meningkatkan akurasi dalam menentukan hubungan evolusi antar taksa. Selain itu, occupancy dalam analisis data sekuens juga berkontribusi dalam menentukan tingkat kepercayaan terhadap hasil filogenetik yang diperoleh.

3.7 Align Crustal W

Clustal W merupakan sebuah perangkat lunak yang digunakan untuk penyelarasan (*alignment*) beberapa urutan biologis, seperti DNA, RNA dan protein. Perangkat ini dirancang untuk menganalisis kesamaan antar urutan dengan menerapkan metode berbasis algoritma progresif. *Clustal W* mengembangkan pendekatan berbobot yang memperhitungkan kontribusi spesifik setiap sekuens berdasarkan tingkat keanekaragamannya, sehingga memberikan hasil yang lebih akurat. Program ini memulai proses dengan menyelaraskan pasangan-pasangan sekuens secara bertahap (*pairwise alignment*) dan kemudian membangun pohon panduan untuk menghasilkan keselarasan keseluruhan (*multiple sequence alignment*). (Thompson, 2015)

Penggunaan *Crustal W* adalah hal pokok dalam penelitian biologi molekuler, seperti analisis untuk membangun pohon filogenetika, prediksi struktur protein berdasarkan daerah konservatif dan identifikasi homologi antar gen atau protein. Perangkat ini pada umumnya biasa didukung dalam format FASTA, *Crustal W* telah menjadi andalan dalam analisis bioinformatika. (Higgins, 1988).

3.8 Align by MUSCLE

MUSCLE (*Multiple Sequence Comparison by Log-Expectation*) adalah perangkat lunak yang digunakan untuk *multiple sequence alignment* (MSA), yaitu penyelarasan urutan DNA, RNA, atau protein dalam skala besar. MUSCLE

dirancang untuk memberikan hasil yang sangat akurat dengan kecepatan yang lebih tinggi dibandingkan dengan penyelarasan algoritma lainnya.

MUSCLE menggunakan pendekatan berbasis dua tahap untuk menyelaraskan urutan secara efektif. Tahap pertama ialah *progressive alignment*, dimana urutan urutan diselaraskan bertahap berdasarkan kesamaan pasangan urutan. Adapun untuk tahap kedua menggunakan optimasi berulang dari hasil penyelarasan menggunakan metode berbasis log-expectation yang lebih sensitif dalam menangani urutan yang lebih panjang dan variatif. MUSCLE sering digunakan dalam analisis yang memerlukan akurasi tinggi dalam penyelarasan urutan yang sangat besar, seperti dalam studi filogenetik yang melibatkan banyak spesies.

Salah satu fitur unggulan MUSCLE adalah kemampuannya bekerja dengan sequence dalam jumlah besar (hingga ribuan) tanpa kehilangan akurasi penyelarasan. Hal ini menjadikannya pilihan populer dalam analisis urutan genom atau protein dalam skala besar, serta dalam pengembangan pohon filogenetik untuk memahami hubungan evolusi antar spesies. Selain itu, MUSCLE memungkinkan pengguna untuk mengontrol berbagai parameter, seperti pengaturan celah penalti dan metode pemotongan yang lebih fleksibel, untuk meningkatkan hasil penyelarasan (Liu, 2015).

MUSCLE juga lebih cepat dibandingkan program penyelarasan lainnya, seperti *ClustalW* dan *T-Coffee*, berkat penggunaan algoritma berbasis ekspektasi log yang dioptimalkan. Hal ini memungkinkan peneliti mendapatkan hasil penyelarasan yang lebih baik dalam waktu yang lebih singkat, menjadikannya alat yang sangat berguna dalam bioinformatika.

3.9 Maximum Likelihood

Maximum Likelihood adalah metode estimasi yang penting dalam analisis statistika. Tujuan utama dari estimasi *maximum likelihood* adalah untuk menemukan parameter yang memaksimalkan kemungkinan gabungan dari suatu data. Misalkan X_1, X_2, \dots, X_n memiliki kepadatan bersama $f_{\theta}(X_1, X_2, \dots, X_n) = f(X_1, X_2, \dots, X_n | \theta)$ dalam hal ini $X_1 = x_1, X_2 = x_2, \dots, X_n = x_n$ adalah nilai observasi sehingga fungsi likelihood dari θ adalah :

$$L(\theta) = \prod_{i=1}^n (X_i | \theta) \quad (1)$$

Untuk memaksimalkan fungsi likelihood, logaritma umumnya digunakan atau lebih baik dikenal sebagai fungsi *loglikelihood*:

Dalam konteks ini:

- $L(\theta)$: *Likelihood* dari parameter θ
- $\prod_{i=1}^n$: Menunjukkan n suku , yang dalam hal ini adalah distribusi probabilitas dari setiap data X_i yang diberikan parameter θ
- $(X_i | \theta)$: Probabilitas atau densitas probabilitas dari X_i diberikan parameter θ .

$$l(\theta|X_1, X_2, \dots, X_n) = \sum_{i=1}^n \ln f(X_n | \theta) \quad (2)$$

Setiap nilai yang memaksimalkan fungsi *likelihood* juga memaksimalkan fungsi *log-likelihood* $\ln L(\theta)$ dan menghasilkan nilai $\frac{\partial}{\partial \theta} \ln L(\theta) = 0$.

Adapun penjelasan untuk notasinya sebagai berikut :

- $l(\theta|X_1, X_2, \dots, X_n)$: *log-likelihood* dari parameter θ , dimana θ biasanya mewakili parameter-parameter model evolusi (misalnya , laju mutasi dan waktu divergensi).
- $\sum_{i=1}^n$:Menunjukkan penjumlahan dari n suku , yang dalam hal ini adalah logaritma dari fungsi kepadatan probabilitas (atau fungsi probabilitas) dari setiap data X_i yang diberikan parameter θ .
- $\ln f(X_i | \theta)$: Logaritma dari fungsi kepadatan probabilitas dari X_i yang diberikan parameter θ .

Metode *maksimum likelihood* menggunakan nilai-nilai dalam ruang parameter Ω sebagai sebagai penaksir nilai parameter yang tidak diketahui. Dalam aplikasinya $L(\theta)$ menunjukkan fungsi kepadatan bersama probabilitas dari sampel acak karena ruang parameter Ω adalah interval terbuka maka $L(\theta)$ adalah fungsi n yang dapat diturunkan dan diasumsikan maksimum pada Ω . Dalam persamaan (2), jelas bahwa fungsi ini mencoba memaksimalkan nilai *likelihood*.

Mai et al. (2014) mengatakan vektor acak X dengan jumlah observasi N termasuk Y adalah kontinu. $f(Y|\theta)$ adalah fungsi densitas probabilitas Y dengan

parameter θ . Dengan mengasumsikan bahwa pengamatan bersifat independen, maka densitas gabungan yang diperoleh adalah:

$$f(\theta|X_1, X_2, \dots, X_n) = \prod_{i=1}^n f(X_i|\theta) \quad (3)$$

Dalam hal ini berfokus pada parameter θ yang tidak diketahui dalam pengamatan oleh karena itu :

$$l(\theta|X_1, X_2, \dots, X_n) = \ln f(X_1, X_2, \dots, X_n|\theta) \quad (4)$$

Persamaan (4) adalah fungsi *likelihood* dengan parameter θ . Dengan memaksimalkan fungsi *likelihood* dari θ disebut sebagai penaksir *maximum likelihood* untuk mengamati data DNA yang mewakili *sequence*.

Dalam filogenetika, data X_1, X_2, \dots, X_n biasanya merupakan sekuens DNA, karakteristik morfologis, atau data lainnya yang digunakan untuk menyusun pohon filogenetik. Model evolusi dengan parameter θ mencoba menjelaskan bagaimana data ini berubah dari nenek moyang yang sama maka dari itu :

$$\hat{\theta} = \arg \max_{\theta \in \Theta} [l(\theta|X_1, X_2, \dots, X_n)] \quad (5)$$

Yang dimana notasinya adalah :

- $\hat{\theta}$: Estimasi *maximum likelihood* (MLE) dari parameter θ .
- $\text{Argmax } \theta \in \Theta$: mencari nilai θ dalam ruang parameter Θ yang memaksimalkan fungsi *likelihood*.

$L(\theta|X_1, X_2, \dots, X_n)$: Fungsi *likelihood* dari parameter θ yang diberikan data X_1, X_2, \dots, X_n

Dalam praktiknya, akan lebih mudah untuk bekerja dengan logaritma natural dari fungsi *likelihood* atau yang biasa dikenal sebagai *loglikelihood*.

$$l(\theta|X_1, X_2, \dots, X_n) = \sum_{i=1}^n \ln f(X_i|\theta) \quad (6)$$

Misalkan X_1, X_2, \dots, X_n memiliki kepadatan peluang bersama $f_{\theta}(X_1, X_2, \dots, X_n) = f(X_1, X_2, \dots, X_n|\theta)$ dengan nilai yang diamati $X_1 = x_1, X_2 = x_2, \dots, X_n = x_n$ sehingga kemungkinan θ adalah:

$$l(\theta = (x_1, x_2, \dots, x_n)|\theta) \quad (7)$$

Pada persamaan (7) diasumsikan bahwa sampel pengamatan adalah

terdistribusi secara independen dan identik (iid). Kemudian dapat membentuk distribusi gabungan dari N pengamatan dengan parameter $(\theta_1, \theta_2, \dots, \theta_p)$.

$$\begin{aligned} L(\theta_1, \theta_2 \dots \theta_p | X_1, X_2, \dots, X_N) &= L(\theta | X_1, X_2, \dots, X_N) \\ &= L(\theta | X_1) \times L(\theta | X_2) \times \dots \times L(\theta | X_N) \\ &= \prod_{i=1}^n L(\theta | Y_i) \end{aligned} \tag{8}$$

- Parameter θ : Bisa mencakup laju mutasi, waktu divergensi, dan parameter lainnya dari model evolusi.
- Data : X_1, X_2, \dots, X_n : Bisa berupa sekuens DNA, protein, atau karakteristik lainnya dari berbagai spesies atau individu.
- *Likelihood* $L(\theta | X_i)$: Menggambarkan seberapa baik parameter θ diberikan semua data X_1, X_2, \dots, X_n adalah produk dari likelihood individu.

3.10 *Phylogenetic Likelihood*

Dalam konteks *Phylogenetics likelihood* memfokuskan perhatian pada data urutan DNA untuk kemudahan, meskipun metode yang dijelaskan juga cocok dengan data karakter diskrit lainnya. Misalnya, jika memeriksa m = sekuens DNA yang disusun (mungkin terkait dengan m = spesies yang berbeda), di mana setiap sekuens mengandung pengamatan nukleotida di n = lokasi yang berbeda. Lalu, celah dalam penjajaran sering kali dianggap sebagai data yang hilang, tetapi ada metode yang lebih tepat untuk memperlakukan penyisipan dan penghapusan (Redelings dan Suchard, 2021) hal ini mempertimbangkan kemungkinan hasil lemparan koin sebagai fungsi dari parameter p yang tidak diketahui, di sini akan mengeksplorasi kemungkinan data urutan DNA sebagai fungsi dari topologi pohon dan panjang cabang yang tidak diketahui. Dengan topologi pohon yang diberikan

dan panjang cabangnya, dapat menggunakan model substitusi untuk menghitung probabilitas perubahan keadaan di sepanjang cabang-cabang pohon.

Berikut adalah contoh dua asumsi yang dibuat yang sangat penting untuk analisis selanjutnya (Felsenstein, 2014):

1. Evolusi pada lokasi yang berbeda (pada pohon tertentu) adalah independen.
2. Bergantung pada keadaan simpul internal, evolusi berlangsung secara independen pada cabang-cabang filogenik:

$L(\tau, t, \theta)$ adalah *likelihood* atau kemungkinan dari suatu matriks penyelarasan sekuens DNA y berdasarkan suatu topologi pohon τ . panjang cabang t , dan parameter model substitusi θ . Hal tersebut adalah ukuran seberapa baik topologi pohon dan parameter yang diberikan menjelaskan data sekuens yang diamati

Misal $L(\tau, t, \theta)$ menjadi kemungkinan yang berhubungan dengan matriks pensejajaran sekuens DNA $m \times n$ y untuk suatu topologi pohon τ dengan vektor panjang cabang t dan vektor parameter model substitusi θ . maka dapat ditulis *likelihood* sebagai berikut :

$$L(\tau, t, \theta) = \Pr(y; \tau, \theta) = \prod_{i=1}^n P(y_i; \tau, \theta) = \prod_{i=1}^n L_i(\tau, t, \theta), \quad (9)$$

Yang dimana y_i adalah vektor $m \times 1$ dari *nukleotida* yang diamati pada situs ke- i dan $L_i(\tau, t, \theta)$ adalah kemungkinan situs ke- i . Faktorisasi ini mengikuti secara langsung dari asumsi independensi pertama yang diberikan di atas. Dengan demikian, hal ini dapat menemukan kemungkinan dari seluruh matriks urutan dengan menemukan kemungkinan untuk masing-masing n situs. Misalkan contohnya mengamati vektor *nukleotida* (A, T, C, T) di situs tertentu (dengan asumsi hanya ada $m = 4$ yang disejajarkan urutan yang sejajar).

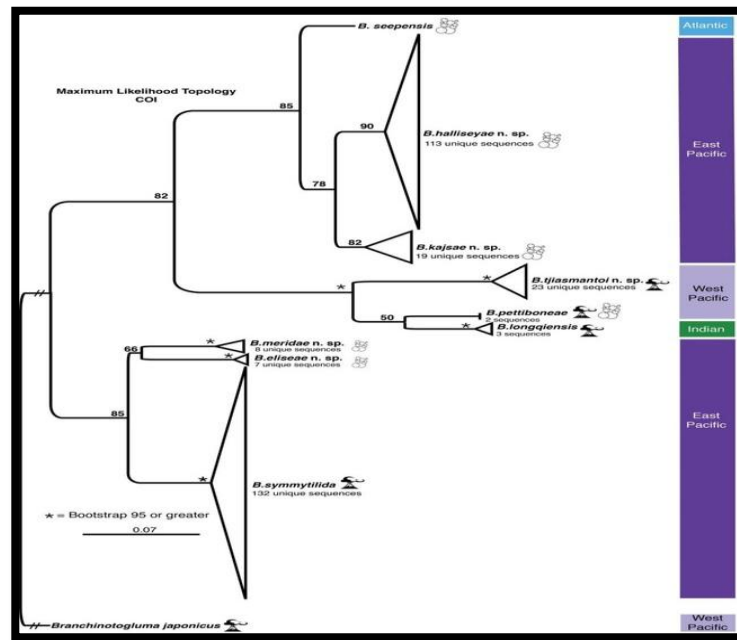
Dengan menggunakan pohon τ_{ex} , hal ini dapat mendekonstruksi kemungkinan vektor nukleotida ini dengan cara berikut:

$$\begin{aligned} \Pr(A, T, C, T; \tau_{ex}, t_{ex}, \theta) &= \sum_x \sum_y \sum_z \Pr(A, T, C, T, x, y, z; \tau_{ex}, t_{ex}, \theta) \\ &= \sum_x \sum_y \sum_z \pi_x P_{xy}(t_1) P_{xz}(t_2) P_{yA}(t_3) P_{yT}(t_4) P = C(t_5) P = T(t_6), \end{aligned} \quad (10)$$

Dalam konteks ini, penjumlahan dilakukan atas elemen-elemen dalam himpunan $\{A, G, C, T\}$. Persamaan (10) diperoleh dengan mengkondisikan pada kondisi simpul internal dan menggunakan asumsi bahwa evolusi adalah independen di seluruh cabang. Kami mencatat bahwa distribusi awal pada akar filogeni

diasumsikan sebagai $\pi = (\pi_A, \pi_G, \pi_C, \pi_T)^T$, yang merupakan distribusi stasioner dari model substitusi (Page dan Holmes, 2019).

ketika m bertambah, komputasi menjadi bermasalah karena penjumlahannya akan melibatkan 4^{m-1} suku-suku pada $m - 1$ simpul internal. Pada bagian selanjutnya, akan menjelaskan sebuah algoritma yang dapat menghitung kemungkinan situs ini secara efisien dengan menghindari komputasi yang berlebihan.



Gambar 3.1 Maximum Likelihood Topologi (Johanna Lindgren, 2019)

3.11 Likelihood Computation

Dalam konteks estimasi filogenetik, penting untuk memahami dan mengukur ketidakpastian yang terkait dengan rekonstruksi filogenika yang diperoleh. Teknik statistik yang umum digunakan untuk mengukur variabilitas sampling dari suatu estimator, seperti *Maximum Likelihood Estimation* (MLE), adalah bootstrap nonparametric. Sebelum menjelaskan prosedur bootstrap, membayangkan cara ideal secara statistik untuk mengestimasi ketidakpastian filogenetik. Bayangkan jika bisa mengulang-ulang "menjalankan kembali" evolusi sepanjang sejarah filogenetik yang sama, dan mendapatkan replika-replika penyelarasan urutan molekuler. Dengan mengestimasi filogeni dari setiap replika ini melalui MLE, dapat memahami variabilitas sampling dari estimasi diri sendiri. Misalnya, jika filogenetika yang diperoleh dari replika-replika ini hampir sama, itu akan

menunjukkan bahwa estimasi filogenetika sangat akurat, dengan variabilitas sampling yang rendah.

Namun, dalam praktiknya, tidak mungkin untuk melakukan evolusi ulang berulang kali seperti ini. Oleh karena itu, bootstrap menyalahi kendala ini dengan mengambil sampel ulang dari data urutan genetik yang diamati. Prosesnya adalah sebagai berikut:

1. Sampel Ulang (*Bootstrap Resampling*): Dari dataset urutan genetik yang ada, secara acak dan dengan penggantian memilih sampel data yang sama ukurannya dengan dataset asli.
2. Estimasi Filogenetika: Untuk setiap sampel bootstrap, bisa menggunakan MLE untuk mengestimasi filogeni (pohon filogenetik) yang paling mungkin berdasarkan sampel data tersebut.
3. Kumpulkan Hasil: Proses ini diulang banyak kali (biasanya ratusan atau ribuan kali, tergantung pada kebutuhan) untuk membangun distribusi dari estimasi filogenetik yang mungkin.
4. Analisis Variabilitas: Dengan menggunakan distribusi hasil dari langkah sebelumnya, yang dapat menghitung statistik seperti interval kepercayaan atau standar deviasi untuk mengukur ketidakpastian dalam estimasi filogenetik.

Teknik bootstrap nonparametrik ini sangat berharga karena tidak bergantung pada asumsi tertentu tentang distribusi data, tetapi secara langsung memanfaatkan data yang diamati untuk mengestimasi ketidakpastian. Hal ini memungkinkan para peneliti untuk membuat kesimpulan yang lebih kuat tentang kepastian filogenetik yang dihasilkan dari analisis MLE.

3.12 *Summarizing Bootstrapped Phylogenies*

Dalam analisis filogenetik menggunakan teknik *bootstrap*, penting untuk memahami dan meringkas kemiripan di antara pohon-pohon bootstrap. Tingkat kemiripan yang tinggi antara pohon-pohon bootstrap menunjukkan tingkat konsistensi atau kestabilan dalam pemisahan filogenetik yang direkonstruksi. Sebagai contoh, jika sebuah bipartisi tertentu dari himpunan spesies muncul dalam sebagian besar pohon *bootstrap*, hal ini mengindikasikan bahwa pemisahan tersebut mungkin mencerminkan struktur filogenetik yang sebenarnya dengan baik. Salah

satu metode untuk menggambarkan pohon-pohon bootstrap adalah dengan membangun pohon konsensus aturan mayoritas. Pohon konsensus ini dibangun dengan mengumpulkan semua pemisahan (split) yang terjadi pada pohon-pohon bootstrap, dan hanya mempertahankan split yang muncul dalam lebih dari 50% dari pohon-pohon tersebut. Dengan kata lain, hanya split-splits yang konsisten diterima oleh mayoritas pohon *bootstrap* yang akan dimasukkan ke dalam pohon konsensus.

Prosedur membangun pohon konsensus ini memastikan bahwa tidak ada dua pemisahan yang bertentangan (*conflicting splits*) yang akan ditempatkan bersama dalam pohon yang sama. Ini karena setidaknya satu pohon *bootstrap* di mana kedua split tersebut hidup berdampingan, sehingga mereka tidak dapat dikonsensuskan menjadi satu pemisahan tunggal. Setelah pohon konsensus dibangun, umumnya setiap split diberi label dengan persentase pohon bootstrap yang mendukung split tersebut. Hal tersebut membantu untuk mengevaluasi kekuatan atau kelemahan interpretasi setiap bagian dari pohon konsensus. Misalnya, split yang didukung oleh banyak pohon bootstrap (dengan persentase tinggi) cenderung lebih dapat diandalkan daripada split yang didukung oleh persentase yang lebih rendah.

3.13 *Maximum Likelihood and more complex models of evolution*

Pada dasarnya, estimasi *maximum likelihood* filogenetik dalam analisis evolusi molekuler sering kali dimulai dengan model yang relatif sederhana, tetapi kompleksitas model dapat ditingkatkan untuk memperhitungkan aspek-aspek seperti heterogenitas tingkat substitusi di seluruh situs-situs dalam urutan molekuler. Model-model yang lebih kompleks, seperti yang diakui secara luas, mempertimbangkan variasi dalam tingkat substitusi di seluruh situs. Misalnya, model yang memungkinkan untuk tingkat substitusi yang berbeda antar situs dapat lebih akurat menggambarkan evolusi sekuens molekuler yang sesungguhnya. Hal ini penting karena situs-situs tertentu dalam urutan genetik dapat mengalami perubahan nukleotida dengan tingkat yang berbeda-beda, yang dapat dipengaruhi oleh berbagai faktor seperti tekanan seleksi atau struktur genetik. Implementasi dari estimasi kemungkinan maksimum filogenetik sering kali mencakup penggunaan model yang dapat menangani heterogenitas ini. Misalnya, model yang melonggarkan asumsi independensi situs atau yang membatasi tingkat substitusi di seluruh cabang filogeni memungkinkan untuk

menggambarkan lebih akurat dinamika evolusi molekuler. Namun, peningkatan kompleksitas model juga dapat memunculkan tantangan baru, seperti kemungkinan bahwa estimasi kemungkinan maksimum tidak lagi menghasilkan solusi yang unik. Hal ini menunjukkan bahwa pemilihan dan identifikasi model yang tepat menjadi sangat penting dalam analisis filogenetik yang lebih kompleks. Studi lanjutan dalam mengidentifikasi model yang cocok dapat membantu dalam memahami lebih baik evolusi molekuler dan hubungan filogenetik antar organisme.

3.14 *General Time Reversible (GTR)*

General Time Reversible (GTR) merupakan model substitusi nukleotida yang mendefinisikan proses *Markov homogeneous* waktu dan *reversible*, dengan menggunakan enam laju substitusi ($\Gamma_{ac}, \Gamma_{ag}, \Gamma_{at}, \Gamma_{cg}, \Gamma_{ct}, \Gamma_{gt}$) dan menggunakan empat frekuensi kesetimbangan basa ($\pi_a, \pi_c, \pi_g, \pi_t$). Fungsi model *General Time Reversible* bisa mengansumsikan probabilitas transisi dari satu basa ke basa lain bersifat tidak simetris, namun laju substitusi dan frekuensi kesetimbangan memenuhi kondisi reversibilitas dari $\pi_i \cdot \Gamma_{ij} = \pi_j \cdot \Gamma_{ji}$,

Sehingga menghasilkan matriks laju Q yang simetris setelah dikalikan dengan vector frekuensi dari π , selain hal itu GTR juga mengestimasi panjang cabang dan struktur pohon filogenetik dengan mempertimbangkan kemungkinan perubahan berbeda antar pasangan basa (Casanelas, 2023).

Proses penerapan metode *Maximum Likelihood* dalam analisis filogenetika dengan model GTR.

Tabel 3.1 Penerapan Model GTR

Sequence	1	2	3	4	5
1	A	A	C	G	T
2	A	C	T	G	G

Langkah pertama yang dilakukan adalah mencari π (*Base frequencies*) yaitu proporsi yang diharapkan dari A, C, G, T dalam urutan.

$$A = 3 \rightarrow \pi_A = \frac{3}{10} = 0.3$$

$$C = 2 \rightarrow \pi_C = \frac{2}{10} = 0.2$$

$$G = 3 \rightarrow \pi_G = \frac{3}{10} = 0.3$$

$$T = 2 \rightarrow \pi_T = \frac{2}{10} = 0.2$$

Langkah kedua yang dilakukan adalah mencari r (*Rate matrix Q*) dengan enam tingkat substitusi independent.

$$r_{AC} = \frac{1+1}{2} = 1$$

$$r_{CT} = \frac{1+1}{2} = 1$$

$$r_{TG} = \frac{1+1}{2} = 1$$

$$r_{AG} = r_{AT} = r_{CG} = 0$$

Langkah ketiga yang dilakukan adalah menghitung berapa kali satu nukleotida berubah menjadi nukleotida lain di seluruh urutan.

$$Q = \begin{matrix} & \begin{matrix} A & C & G & T \end{matrix} \\ \begin{matrix} A \\ C \\ G \\ T \end{matrix} & \begin{bmatrix} - & r_{AC}\pi_C & r_{AG}\pi_G & r_{AT}\pi_T \\ r_{CA}\pi_A & - & r_{CG}\pi_G & r_{CT}\pi_T \\ r_{GA}\pi_A & r_{GC}\pi_C & - & r_{GT}\pi_T \\ r_{TA}\pi_A & r_{TC}\pi_C & r_{TG}\pi_G & - \end{bmatrix} \end{matrix}$$

$$Q = \begin{matrix} & \begin{matrix} A & C & G & T \end{matrix} \\ \begin{matrix} A \\ C \\ G \\ T \end{matrix} & \begin{bmatrix} - & 0.2 & 0 & 0 \\ 0.2 & - & 0 & 0.2 \\ 0 & 0 & - & 0.2 \\ 0 & 0.2 & 0.2 & - \end{bmatrix} \end{matrix}$$

Langkah keempat yang dilakukan adalah mencari *Transition probability matrix* P(t) dihitung menggunakan eksponensiasi matriks Q.

$$P_{XY}(t) = \pi_Y + r_{XY} \cdot \pi_Y \cdot t$$

$$P_{AC}(0.1) = 0.2 + (1)(0.2)(0.1) = 0.2 + 0.02 = 0.22$$

Untuk langkah kelima yang dilakukan adalah mencari *Likelihood* di masing-masing jalur.

$$\begin{array}{ll}
L_1 = \pi_A \cdot P_{ac}(t) & L_2 = \pi_A \cdot P_{AA}(t) \\
= 0.3 \times 0.22 & = 0.3 \times 0.3 \\
= 0.66 & = 0.09
\end{array}$$

Untuk langkah keenam adalah menghitung *overall likelihood*.

$$\begin{array}{ll}
L = L_1 \times L_2 & \log(l) = \log(0.0594) \\
= 0.66 \times 0.09 & = -1.22 \\
= 0.0594 &
\end{array}$$

3.15 *Bootstrapping*

Bootstrapping dalam filogenetika merupakan metode statistik berbasis *reseampling* untuk mengukur tingkat kepercayaan terhadap cabang-cabang dalam pembentukan pohon filogenetika. Proses *bootstrapping* dilakukan dengan cara mengambil sampel ulang dari data sekuens (dengan pengembalian) sebanyak ratusan hingga ribuan kali, lalu membangun pohon untuk setiap sampel ulang tersebut. Setiap *bootstrap replicate* dibentuk dengan secara acak memilih kolom-kolom dari *alignment* asli untuk menghasilkan dataset baru yang memiliki panjang sama, namun dengan kolom-kolom yang mungkin terduplikasi. Pohon filogenetik dibangun dari tiap replika tersebut menggunakan metode inferensi (*misalnya maximum likelihood*), lalu nilai *bootstrap support* dihitung sebagai proporsi frekuensi kemunculan klad/cabang tertentu dalam seluruh replikasi pohon. Proporsi kemunculan kembali suatu cabang atau klad di antara seluruh pohon yang dihasilkan digunakan sebagai nilai *bootstrap support*, biasanya dinyatakan dalam persentase. Secara teknis, nilai dukungan *bootstrap* (*bootstrap percentage*, BP) pada setiap node pohon merepresentasikan estimasi frekuensi empiris dari node tersebut jika eksperimen diulang berkali-kali dengan sampel data serupa. Nilai BP $\geq 70\%$ dianggap menunjukkan dukungan moderat, sedangkan $\geq 95\%$ dianggap sangat kuat. Fungsi utama dalam *bootstrapping* dalam filogenetika adalah untuk mengevaluasi kestabilan topologi pohon serta memberikan indikator sejauh mana data mendukung keberadaan klad tertentu (Hoang et al., 2018; Minh et al., 2020)

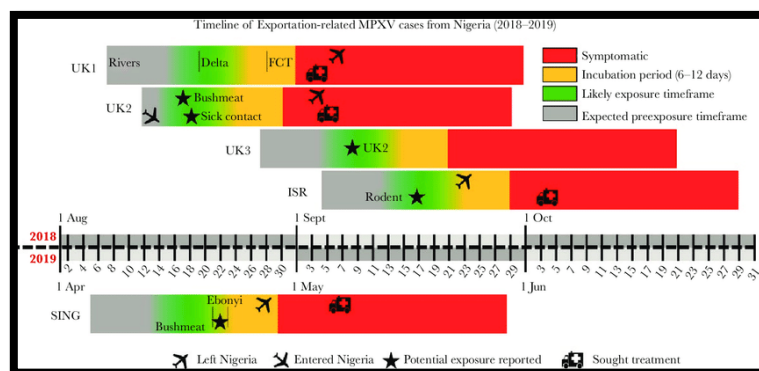
3.16 *MonkeyPox* (*Cacar Monyet*)

Virus cacar monyet, virus zoonosis langka yang dapat menular dari hewan ke manusia, telah diketahui menyebabkan wabah di berbagai belahan dunia, termasuk Afrika Barat dan Amerika Serikat. Kasus cacar monyet pada manusia pertama kali

dilaporkan pada tahun 1970 di Republik Demokratik Kongo, Sejak saat itu, virus ini menjadi endemik di beberapa bagian Afrika Barat, dan wabah telah dilaporkan di lebih dari 60 negara non-endemik .Peristiwa-peristiwa penting dalam wabah cacar monyet pada tahun 2022 meliputi:

- Mei 2022: Kasus cacar monyet pertama kali dilaporkan di Amerika Serikat
- 28 Juni 2022: Pemerintahan Biden mengumumkan fase pertama dari strategi vaksin cacar monyet nasionalnya
- 23 September 2022: Data awal AS menunjukkan bahwa vaksin tersebut efektif, dan WHO merilis taksonomi untuk menargetkan informasi yang salah tentang cacar monyet
- 28 September 2022: Inggris mengumumkan kelayakan dosis kedua untuk vaksinasi cacar monyet
- 1 Oktober 2022: Pusat Pengendalian dan Pencegahan Penyakit AS (CDC) melaporkan bahwa wabah domestik melambat, tetapi cacar monyet kemungkinan akan terus beredar pada tingkat yang rendah di Amerika Serikat.

Virus cacar monyet pertama kali diisolasi pada tahun 1958 selama wabah penyakit mirip cacar di antara monyet *cynomolgus* di Kopenhagen Reservoir alami dari virus ini tidak diketahui, tetapi diyakini adalah mamalia kecil seperti tupai dan monyet. Virus ini dapat menyebar dari orang ke orang melalui kontak langsung dengan kulit yang terinfeksi atau bahan lain, seperti seprai yang terkontaminasi.



Gambar 3.2 *MonkeyPox Research Timeline* (Matthew et al, 2020)

Gambar garis waktu kasus virus *Monkeypox* yang diekspor dari Nigeria pada musim gugur (Agustus-Oktober) 2018 dan musim semi (April-Juni) 2019, masing-

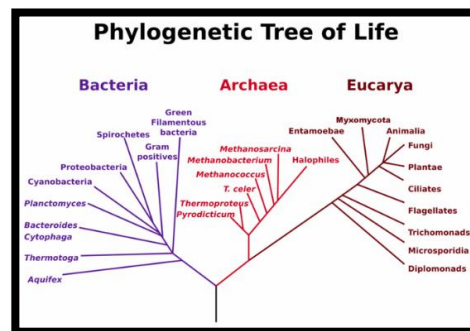
masing ditunjukkan di atas dan di bawah garis waktu. Penggambaran antara jangka waktu paparan dan jangka waktu inkubasi dibuat secara bertahap, karena variasi yang terjadi. Tanggal perjalanan antar negara bagian tidak diketahui untuk UK2. Garis hitam vertikal dalam garis waktu untuk UK1 dan SING menunjukkan perjalanan antar negara bagian dan/atau Wilayah Ibu Kota Federal. Gejala awal didasarkan pada kasus yang melaporkan adanya lesi atau prodrome demam. Waktu inkubasi didokumentasikan berkisar antara 6 hingga 12 hari, dengan batas atas 16 hari. Angka-angka ini digunakan untuk menentukan kemungkinan waktu pemaparan (hijau), serta waktu yang diperkirakan sebelum pemaparan (abu-abu). Singkatan: FCT, Wilayah Ibu Kota Federal; ISR, Israel; MPXV, virus cacar monyet; SING, Singapura; UK, Inggris.

3.17 Molecular Phylogenetics

Dalam menilai hubungan evolusi spesies digunakan protein atau urutan gen yang diterapkan pada analisis filogenetik molekuler. Sejarah evolusi spesies dapat dijelaskan melalui filogeni. Filogenetika adalah cabang ilmu biologi yang mempelajari dan menganalisis klasifikasi yang menunjukkan hubungan antara suatu spesies dengan nenek moyangnya dan hubungan evolusi antara organisme atau spesies. Ilmu ini dapat diaplikasikan dalam bidang ekologi, biologi evolusi, epidemiologi manusia, dan bidang-bidang lainnya. Analisis filogenetik tidak dapat dipisahkan dari evolusi biologis. Evolusi didefinisikan sebagai proses organisme secara bertahap yang memungkinkan spesies sederhana menjadi lebih kompleks melalui perubahan tambahan selama beberapa generasi. Dalam hal tersebut dapat menghitung jarak genetik berdasarkan sekuens DNA dengan jumlah basa polimorfik yang berbeda dari suatu lokus gen di setiap populasi. Memprediksi hubungan evolusi dapat dilakukan dengan analisis filogenetik. Hasil analisis yang diperoleh biasanya berupa pohon filogenetik atau diagram bercabang.

Dalam menentukan pohon filogenetik, terdapat dua metode berdasarkan yaitu berdasarkan karakter dan berdasarkan jarak. Beberapa metode yang dikenal berdasarkan jarak (*distance-based*) seperti: *neighbor-joining* dan UPGMA. Sedangkan metode yang berdasarkan karakter (*character-based*) antara lain: *maximum-parsimony* *maximum-likelihood*, dan *Bayesian Markov Chain Monte Carlo* (BMCMC). Analisis dan pembuatan pohon filogenetik dapat dilakukan

melalui beberapa alat bantu seperti seperti MEGA (*Molecular Evolutionary Genetics Analysis*) X (Kumar et al., 2018), iTOL (*Interactive Tree of Life*) (Letunic & Bork, 2021), RAxML (Randomized Percepatan Maksimum Kemungkinan) (Stamatakis, 2014), BEAST (*The Bayesian Evolutionary Analysis by Sampling Trees*) (Suchard et al., 2018), dan alat lainnya. Mengutip publikasi review dari (Yang & Rannala, 2012), berikut ini adalah lampirkan kelebihan dan kekurangan dari metode filogenetik secara keseluruhan.



Gambar 3.3 *Molecular Phylogenetics* (Suresh Kaushik et.al, 2021)

Tabel 3.2 Kelebihan dan kekurangan setiap metode

Metode	Kelebihan	Kekurangan
Distancebased /Berbasis jarak (UPGMA, Neighbor- Joining, etc).	<ul style="list-style-type: none"> - Memiliki kemampuan komputasi yang cepat. - Mudah diaplikasikan pada semua jenis data selama jarak genetik ditentukan. - Mudah dalam memilih model yang disesuaikan dengan data untuk perhitungan jarak. 	<ul style="list-style-type: none"> - Tidak ada pertimbangan varians estimasi jarak pada beberapa metode berbasis jarak (contohnya, <i>Neighbor- Joining</i>, etc) - Ada masalah dalam menghitung jarak ketika ada beberapa celah yang terlibat dalam penyejajaran dan sekuens yang berbeda . - Ada masalah dalam menghitung jarak ketika ada beberapa celah yang terlibat dalam penyelarasan dan urutan yang berbeda - Panjang cabang negatif, tidak memiliki definisi tersirat. - Tidak mempertimbangkan model evolusi.
<i>Maximum Likelihood</i>	<ul style="list-style-type: none"> - Mendekati fakta-fakta biologis menggunakan substitusi kompleks model. - Ada kerangka kerja yang kuat yang dapat digunakan untuk pengujian hipotesis dan estimasi parameter. Serta metode ini akurat berbaris probabilitas mempertimbangkan evolusi cocok untuk data set besar. 	<ul style="list-style-type: none"> - komputasi yang dilakukan berat dan memakan waktu yang lama.

Metode	Kelebihan	Kekurangan
<i>Maximum parsimony</i>	<ul style="list-style-type: none"> - Ada yang intuitif daya tarik dan kesederhanaan. - Cepat sederhana, tidak perlu model evolusi. 	<ul style="list-style-type: none"> - Asumsi dari metode ini metode ini kurang dapat dimengerti dan bersifat implisit - Ada kekurangan dari model yang tidak menggabungkan pengetahuan tentang evolusi urutan - Secara substansial kurang dihargai cabang panjang sebagai substitusi meningkat. - Rentan terhadap <i>long-branch attractions</i>, kurang akurat untuk data set besar.
<i>Bayes Markov Chain Monte Carlo</i>	<ul style="list-style-type: none"> - Dapat menggunakan model substitusi (sama seperti model kemungkinan-maksimum) - Kemungkinan untuk menggabungkan ahli pengetahuan dan informasi dari probabilitas sebelumnya. - Probabilitas posterior membuatnya lebih mudah untuk menafsirkan pohon dan <i>clade</i>. 	<ul style="list-style-type: none"> - Ada yang berat komputasi dalam penggunaan Rantai Markov Monte Carlo (MCMC). - Kesulitan yang mungkin terjadi mengidentifikasi dan memperbaiki Konvergensi MCMC dan masalah pencampuran dalam kumpulan data yang besar - Kemungkinan adanya pengaruh yang tidak semestinya pada posterior tanpa pengetahuan peneliti dari

		<p>multidimensi prioritas.</p> <ul style="list-style-type: none"> - Sering terlihat bahwa probabilitas posterior terlalu tinggi - Ada tantangan komputasi tantangan komputasi dalam model dalam pemilihan model (Lartillot & Philippe, 2006; Xie et al., 2011).
--	--	---

BAB IV

METODOLOGI PENELITIAN

4.1. Populasi dan Sampel Penelitian

Populasi dalam penelitian ini adalah genom virus *Monkeypox* seluruh warga dunia yang telah tersimpan dalam GISAID. Sedangkan sampel data yang digunakan adalah genom virus *Monkeypox* dari berbagai benua Asia, Eropa, Amerika, Amerika Latin, Afrika dan Australia yang mana dari setiap benua akan diwakilkan dengan negara yang terbanyak terkena kasus dari *monkeypox*. Sampel yang digunakan untuk penelitian ini adalah $n = 106$ genom virus *monkeypox* dalam bentuk fasta data DNA. Dari total data sampel yang digunakan oleh penulis melibatkan genom virus penyintas dari beberapa penyintas dari negara lain.

Data yang digunakan dalam penelitian ini adalah data sekunder yang diperoleh dari GISAID. Semua sampel ini diambil oleh rumah sakit, institusi, atau lembaga penelitian. Pengambilan sampel dilakukan dengan berbagai cara, seperti melalui swab nasofaring dan usap nasofaring dan *orofaring*, cairan bilasan *bronchoalveolar*, dahak, air liur, dan usap tenggorokan. Sampel yang digunakan untuk penelitian ini adalah $n = 106$ orang yang terkena MPXV. adapun Data dalam bentuk DNA (FASTA). Dari keseluruhan sampel data yang digunakan, penulis melibatkan genom virus penyintas dari beberapa penyintas dari berbagai benua dengan total 106 sampel. Penelitian ini menggunakan dua data yang terpisah, yaitu: 1) Data yang berisi informasi mengenai genom virus dari para penyintas seperti (clade, status pasien, tahun kejadian, negara asal penyintas, jenis kelamin, dan lain-lain), dan 2) Data genom virus cacar monyet yang selamat dalam bentuk DNA. Pada data yang pertama, penulis akan membuat visualisasi terkait informasi yang tersedia. Sedangkan pada data kedua, penulis akan mengembangkan pohon filogenetik yang berharap untuk mengidentifikasi garis keturunan dari genom virus dari penyintas *MonkeyPox*.

4.2. Tempat dan Waktu Penelitian

Penelitian ini dilakukan di Yogyakarta tepatnya di Universitas Islam Indonesia secara rutin dari bulan November 2023 - Mei 2024. Adapun waktu dan tempat pengumpulan data dilakukan dan mulai pada Desember 2024 melalui GISAIID.

4.3. Definisi Operasional Variabel Penelitian

Tabel 4.1 Data Variabel

Variabel	Deskripsi
Clade dan Subclade	<p><i>Clade</i> adalah kelompok organisme yang mencakup satu nenek moyang bersama beserta semua keturunannya. <i>Clade</i> sering disebut juga sebagai monofiletik, yang berarti semua anggota dalam kelompok tersebut berasal dari satu leluhur yang sama dan tidak ada spesies di luar kelompok yang termasuk di dalamnya.</p> <p>I, II dan IIb.</p> <p><i>Subclade</i> adalah bagian dari clade yang lebih besar, yaitu kelompok organisme yang memiliki nenek moyang bersama tetapi merupakan bagian dari cabang yang lebih kecil dalam pohon filogenetika.</p> <p>Ia, Ib, IIb C.1, IIb A.2, IIb B.1.6, IIb B.1, IIb dan B.1.17.</p>
Jenis Kelamin	Perempuan, Laki-laki, dan <i>Unknown</i>
Status Pasien	Hidup, Rawat Inap, Tidak diketahui, Meninggal, Dipulangkan, Tanpa Gejala, Sembuh, Diisolasi di Rumah, dan dipindahkan ke fasilitas lain.

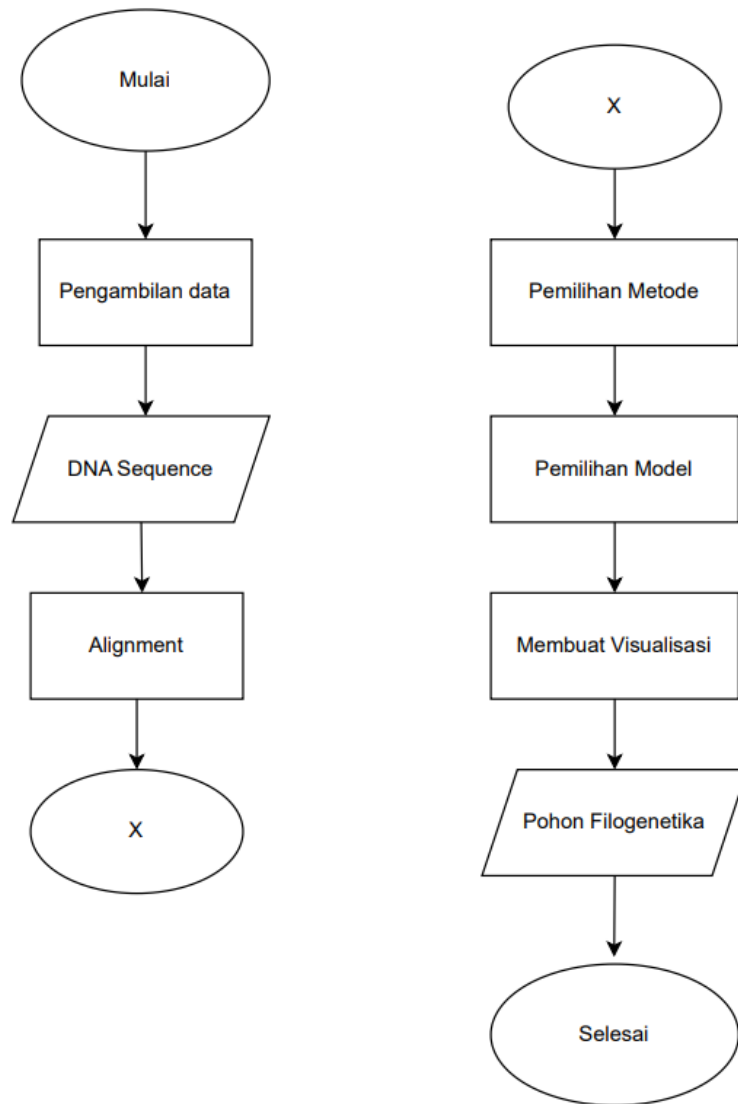
Variabel	Deskripsi
Kategori Umur	0-4 , 5- 9, 10-14, 15-19, 20-24, 25-29, 30-34, 35-39, 40-44, 45-49, 50-54, 55-59, 59> dan <i>Unknown</i>
Tahun	2020 - 2024
Benua dengan perwakilan setiap Negara	Asia, Eropa , Amerika Utara , Amerika Selatan , Afrika dan Australia.
DNA	DNA yang mengandung <i>genome</i> virus <i>MonkeyPox</i> pasien yang dapat dianalisis.

4.4. Alat dan cara organisir data

Terdapat dua data yang dibagi menjadi dua hasil penelitian, yaitu data untuk visualisasi dan data untuk membangun pohon filogenetika. Pengelolaan data menggunakan *Microsoft Excel*. Untuk membuat dashboard menggunakan *Powe BI*. Untuk penyalarsan Genome menggunakan *Jalview*. Untuk proses analisis filogenetika menggunakan *Jalview*, *MEGA-X*, dan Untuk membuat pohon filogenetik menggunakan *iTOL v.5*.

4.5. Metode Penelitian

Metode yang digunakan dalam proses rekonstruksi pohon filogenetik ini adalah metode *maximum likelihood* dan metode *maximum parsimony* dengan model substitusi DNA, general *timereveseable* (GTR); GTR+G+I dan menerapkan 1200 perulangan. Metode *maximum-likelihood* memiliki kelebihan yaitu metode ini mendekati fakta biologi dengan memanfaatkan penggunaan model substitusi yang kompleks dan untuk mengestimasi dan memprediksi parameter dan menguji hipotesis. Adapun metode *Maximum parsimony* berfokus pada pencarian pohon dengan jumlah perubahan karakter minimal, memberikan pendekatan yang sederhana dan intuitif. Hasil akhir dari penelitian ini adalah pohon filogenetika dan wawasan biologi terkait varian *MonkeyPox* yang dapat diterjemahkan oleh penulis dan dipahami oleh semua kalangan, terutama kalangan non-STEM.



Gambar 4.1 Tahapan Penelitian

Berdasarkan diagram diatas, tahapan yang dilakukan dapat dijabarkan sebagai berikut :

1. Mengunjungi GISAID untuk mengumpulkan *genome* virus (Data DNA).
2. Menggunakan *Jalview* untuk mengumpulkan informasi sekuen.
3. Proses penyejajaran sekuen Berganda (*Alignment*) dengan MAFFT.
4. MEGA-X digunakan untuk pemilihan metode dan pengujian model
5. Memvisualisasikan pohon filogenetik menggunakan iTOL v.5.
6. Mencari hubungan kekerabatan dan mengintrepetasikan hasil rekontruksi pohon filogenetika.

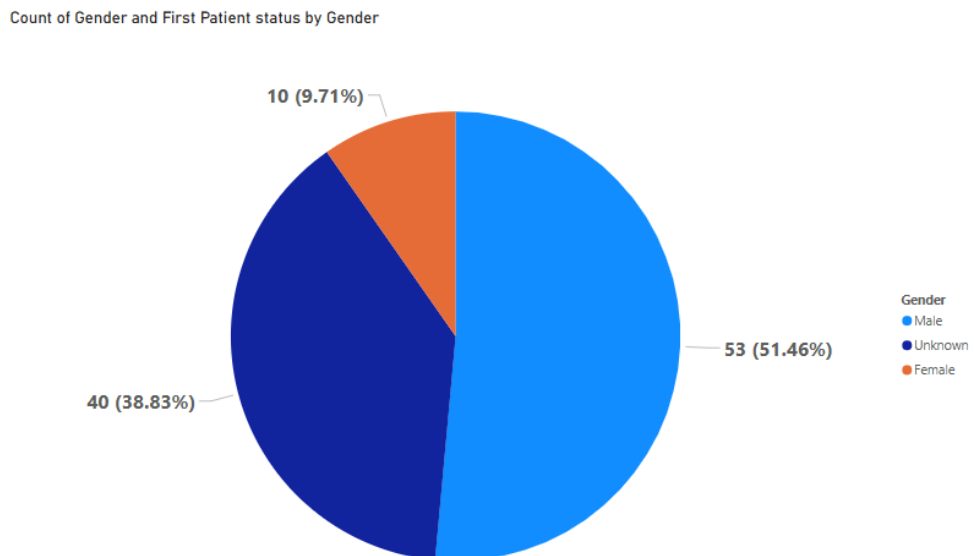
BAB V

PEMBAHASAN

Pada bab kelima ini, penelitian akan membahas hasil dan penjelasan terkait analisis komputasi biologi. Adapun untuk proses ini akan melibatkan beberapa penggunaan software yang diantaranya ialah, *Mega X* untuk membuat kontruksi pohon filogenetika, *Cliplot Tree* untuk visualisasi pohon filogenetika, *jalview* untuk penyalarsan *genome squence* dan *Power BI* untuk pembuatan visualisasi data dalam bentuk dasbor.

5.1. Statistika Deskriptif

Dalam bagian eksplorasi analisis pada bagian 5.1 , penulis akan menampilkan statistika deskriptif sebelum dilakukanya analisis philogenetics, adapun pembuatan dasbor menggunakan perangkat lunak Power BI. Seperti yang dijelaskan dalam bab empat, terdapat beberapa variabel yang digunakan untuk menjelaskan studi kasus *monkeypox* yang tersebar di beberapa negara dan benua.



Gambar 5.1 Persentase *Monkeypox* berdasarkan *Gender*

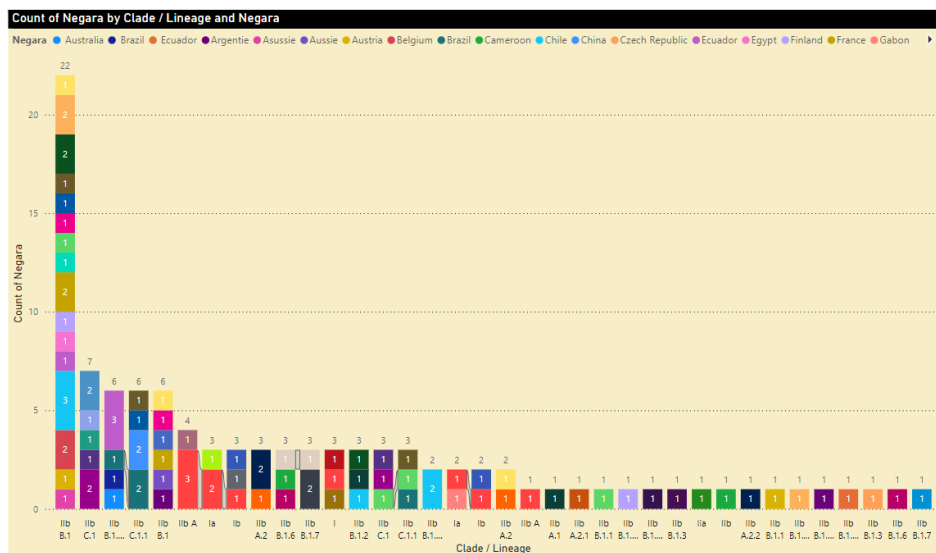
Berdasarkan **gambar 5.1** diatas menampilkan hasil visualisasi persentase angka pasien *monkeypox*. Dalam kasus diatas visualisasi menampilkan tiga kategori gender: Laki- laki dengan persentase 51.46% , perempuan dengan persentase sebesar 9.71% dan yang terakhir adalah pasien yang tidak diketahui jenis kelamin nya (*unknown*) yaitu sebesar 38.83%. Dari grafik diatas bisa di dapatkan

kesimpulan bahwa persentase angka pasien *monkeypox* tertinggi adalah kategori laki-laki dengan persentase sebesar 51.66%.

Tabel 5.1 Data pasien berdasarkan gender

Benua	Female	Male	Unknown
ASIA	-	17 Pasien	3 Pasien
AFRIKA	9 Pasien	6 Pasien	9 Pasien
AUSRALIA	-	1 Pasien	3 Pasien
EUROPA	1 Pasien	8 Pasien	24 Pasien
AMERIKA	-	22 Pasien	-

Adapun dalam kasus *monkeypox* yang di alami oleh wanita pada umum nya virus tersebut menular secara kontak *heteroseksual*, termasuk pekerja sex yang rentan terkena wabah virus *monkeypox*, seperti hal nya yang terjadi pada tahun 2022 banyak dari gender wanita terkena wabah *monkeypox* terutama di negara Republik Demokratik Kongo (DRC), pada saat itu banyak sekali pria yang berkerja untuk menggali pertambangan dan banyak wanita sebagai pelayan sex.



Gambar 5.2 Clade/Lineage berdasarkan negara

Pada **gambar 5.2** grafik diatas menampilkan hasil dari visualisasi persebaran *Clade/ garis keturunan (lineage)* dari *monkeypox* yang tersebar di berbagai negara hingga benua. Grafik ini menunjukkan adanya 24 jenis garis keturunan yang tersebar

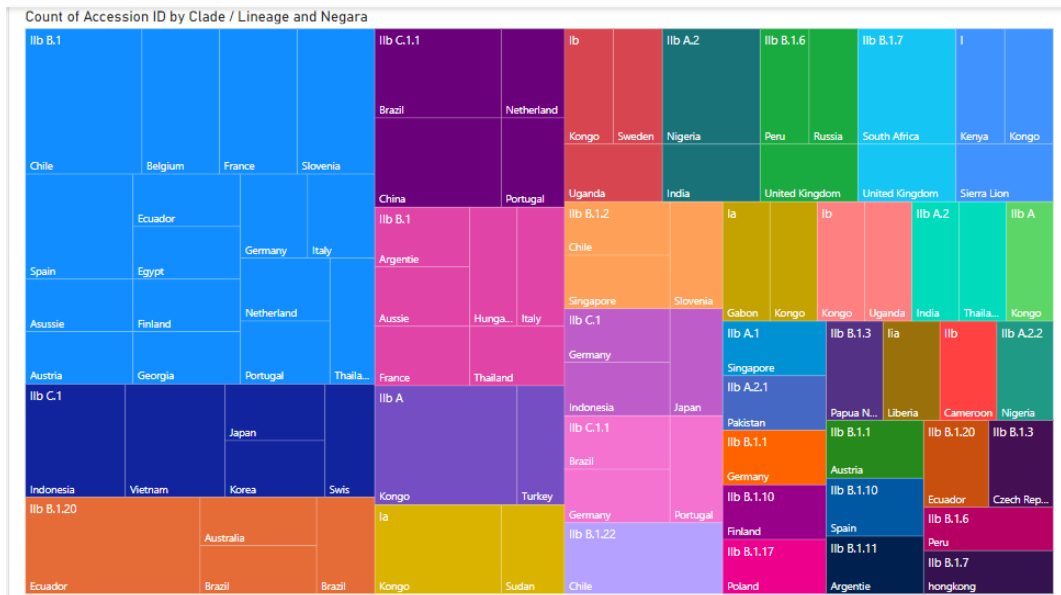
di berbagai negara, berikut adalah garis keturunan yang teridentifikasi di antaranya ialah :

- **Iib B.1** ditemukan di 16 negara, menjadikan garis keturunan dengan cakupan yang paling luas dalam studi kasus ini. Adapun persebarannya terdapat pada negara Thailand (2 kasus), Spain (2 kasus), Slovenia (2 kasus), Portugal (1 kasus) , Netherland (1 kasus), Italy (2 kasus), German (1 kasus), Gerogia (1 kasus), France (3 kasus), Egypt (1 kasus), Finland (1 kasus), Ecuador (1 kasus), Chile (3 kasus), Belgium (2 kasus), Austria (1 kasus), Hungary (1 kasus), Argentine (1 kasus) dan Australia (2 kasus). Dengan total kasus 28 korban.
- **Iib C.1** teridentifikasi di 5 negara, menunjukkan penyebaran yang cukup signifikan. Diantara ialah Vietnam (2 kasus), Swis (1 kasus), Korea selatan (1 kasus), jepang (1 kasus) dan indonesia (2 kasus). Dengan total kasus 7 korban.
- **Iib 1.B.20** teridentifikasi di 3 negara, diantaranya ialah. Ecuador (3 kasus), Brazil (2 kasus) dan australia (1 kasus) dengan total 6 kasus.
- **IIB C.1.1** teridentifikasi di 5 negara, diantaranya ialah. Portugal (2 kasus), Netherland (1 kasus), Germany (1 kasus), China (2 kasus) dan Brazil (3 kasus) dengan total 9 kasus dengan sampel yang sama.
- **Iib A** Garis keturunan ini teridentifikasi di 2 negara yaitu Turkey 1 kasus dan Kongo 4 kasus dengan total 5 kasus dengan sampel yang sama.
- **Ia** Garis keturunan ini teridentifikasi di 3 negara yaitu Sudan (1 kasus), Gabon (1 kasus) dan Kongo (3 kasus) dengan total 5 kasus dengan sampel yang sama.
- **Ib** Garis keturunan ini teridentifikasi di 3 negara yaitu Uganda (2 kasus), Sweden (1 kasus) dan Kongo (2 kasus) dengan total 5 kasus, dengan sampel yang sama.
- **Iib A.2** Garis keturunan ini teridentifikasi di 2 negara yaitu Nigeria (2 kasus) dan India (1 kasus) dengan total 3 kasus sampel yang sama.
- **Iib B.1.6** Garis keturunan ini teridentifikasi di 3 negara yaitu United Kingdom (1 kasus), Russia (1 kasus) dan peru (1 kasus) dengan total 3 kasus sampel yang sama.

- **IIb B.1.7** Garis keturunan ini terindetifikasi di 2 negara yaitu United Kingdom (1 kasus) dan South Africa (2 kasus) dengan total 3 kasus sampel yang sama.
- **I** Garis keturunan ini terindetifikasi 3 negara yaitu Kongo (1 kasus), Siera Lion (1 kasus) dan Kenya (1 kasus) dengan total 3 kasus sampel yang sama.
- **IIb B.1.2** keturunan ini terindetifikasi 3 negara Slovenia, Singapore dan Chile dengan masing-masing memiliki 1 kasus dengan sampel yang sama.
- **IIb C.1** keturunan ini terindetifikasi 3 negara Japan, Indonesia dan Germany dengan masing-masing memiliki 1 kasus dengan sampel yang sama.
- **IIb B.1.22** keturunan ini terindetifikasi 1 negara yaitu chile dengan 2 kasus terjadinya sampel virus adalah sama.
- **IIb A2** keturunan ini terindetifikasi di 2 negara yaitu India dan Thailand dengan masing masing mendapatkan 1 sampel kasus yang sama.
- **IIb A.1** Garis keturunan ini ditemukan di singapura dengan terindifikasinya 1 sampel.
- **IIb A.2.1** Garis keturunan ini ditemukan di pakistan dengan terindifikasinya 1 sampel.
- **IIb B.1.1** Garis keturunan ini ditemukan di germany dengan terindifikasinya 1 sampel
- **IIb B.1.10** Garis keturunan ini ditemukan di Finland dengan terindifikasinya 1 sampel
- **IIb B.1.17** Garis keturunan ini ditemukan di Poland dengan terindifikasinya 1 sampel
- **IIa** Garis keturunan ini ditemukan di Liberia dengan terindifikasinya 1 sampel
- **IIb B.1.3** Garis keturunan ini ditemukan di Papua New Guinie dengan terindifikasinya 1 sampel
- **IIb** Garis keturunan ini ditemukan di Cameroon dengan terindifikasinya 1 sampel

- **Iib A.2.2** Garis keturunan ini ditemukan di Nigeria dengan terindifikasinya 1 sampel
- **Iib B.1.1** Garis keturunan ini ditemukan di Austria dengan terindifikasinya 1 sampel
- **Iib B.1.10** Garis keturunan ini ditemukan di Spain dengan terindifikasinya 1 sampel
- **Iib B.1.11** Garis keturunan ini ditemukan di Argentine dengan terindifikasinya 1 sampel
- **Iib B.1.20** Garis keturunan ini ditemukan di Ecuador dengan terindifikasinya 1 sampel
- **Iib B.1.3** Garis keturunan ini ditemukan di Czech Republic dengan terindifikasinya 1 sampel
- **Iib B.1.6** Garis keturunan ini ditemukan di Peru dengan terindifikasinya 1 sampel
- **Iib B.1.7** Garis keturunan ini ditemukan di Hongkong dengan terindifikasinya 1 sampel

Grafik ini juga memberikan wawasan penting tentang keragaman garis keturunan yang ditemukan di setiap negara, ditandai dengan variasi warna pada setiap batang. Semakin banyak warna dalam batang tertentu, semakin beragam negara yang memiliki garis keturunan tersebut. Misalnya, Iib B.1 tidak hanya memiliki jumlah negara terbanyak (16 negara), tetapi juga memperlihatkan keragaman geografis terbesar, yang mencerminkan penyebarannya yang luas. Sementara itu, garis keturunan lainnya, meskipun jumlah negara tempat ditemukan lebih sedikit, tetapi memberikan informasi penting tentang pola penyebaran lokal virus tersebut. garis keturunan Iib B.1 menonjol sebagai yang paling banyak terdistribusi secara global, diikuti oleh garis keturunan lain dengan distribusi yang lebih terbatas. Visualisasi ini memberikan gambaran jelas tentang bagaimana *lineage monkeypox* tersebar dan membantu memahami pola penyebarannya berdasarkan lokasi geografis. Berikut adalah gambar *Treemap* untuk melihat negara yang terindetikasi.

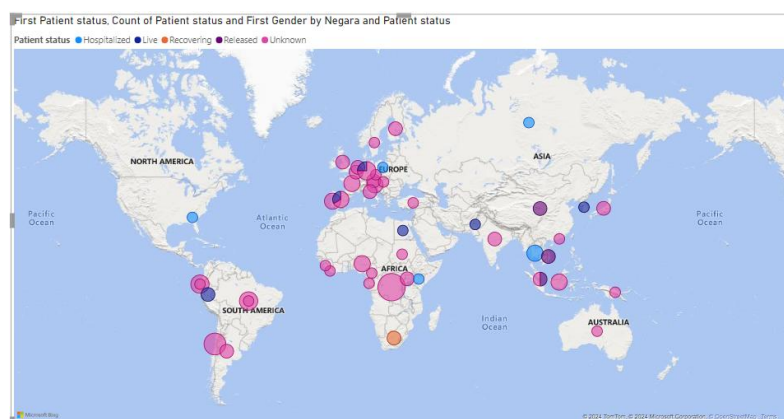


Gambar 5.3 Visualisasi *monkeypox* dengan *Treemap*

Pada visualisasi **gambar 5.3** menunjukkan bahwa data yang dianalisis berasal dari berbagai negara di seluruh dunia. Setiap persegi panjang mewakili satu negara, dan ukurannya menunjukkan jumlah *Accession ID* yang berasal dari negara tersebut. Adapun warna yang berbeda pada setiap persegi panjang kemungkinan menunjukkan perbedaan *Clade/Lineage*. Dengan dominasi dari beberapa negara memiliki area yang lebih besar dibandingkan negara lain, yang mengindikasikan bahwa terdapat jumlah *Accession ID* yang lebih banyak dari negara tersebut. Hal ini bisa menunjukkan bahwa penelitian lebih banyak dilakukan di negara-negara tersebut atau bahwa keragaman genetik di negara-negara tersebut lebih tinggi.

Pada visualisasi **gambar 5.3** juga terdapat beberapa perbedaan seperti ukuran panjang pada setiap kotak dan ada juga yang berbentuk *landscape*, hal ini terjadi hal ini menandakan jumlah data lebih banyak disuatu negara dengan clade yang sama, yang berarti negara tersebut aktif dalam pelaporan kasus wabah yang lebih besar. Adapun warna yang sama tapi berbeda benua di karenakan virus *monkeypox* dengan clade yang sama terjadi juga di negara tersebut seperti contoh **gambar 5.3** terdapat negara Ekuador, Brazil dan Australia. Jika di lihat letak geografis benua Australia tidak termasuk benua Amerika, namun pada **gambar 5.3** diwarnai dengan warna coklat yang disatukan dengan negara Ekuador dan Brazil, yang dimana bisa artikan bahwasanya terjadi mutasi virus dari benua Amerika ke benua Australia, hal ini terjadi dikarenakan mobilitas yang tinggi.

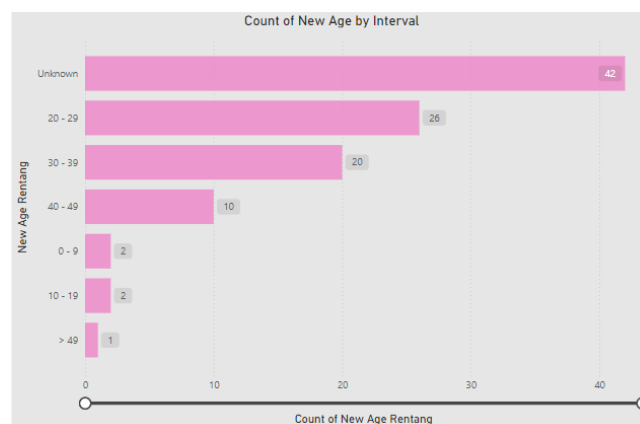
Adapun pada bagian distribusi *Clade/Lineage* warna-warna yang berbeda pada setiap persegi panjang memberikan gambaran kasar tentang distribusi *Clade/Lineage* dari *monkeypox* di setiap negara. Misalnya, jika warna merah dominan di suatu negara, maka kemungkinan besar *Clade/Lineage* tertentu lebih sering ditemukan di negara tersebut. Dengan melihat posisi relatif dari persegi panjang, dapat melihat hubungan antara negara-negara berdasarkan *Clade/Lineage* yang dimiliki. Misalnya, jika dua negara memiliki persegi panjang yang berdekatan dan berwarna serupa, maka kemungkinan besar mereka memiliki kemiripan genetik. Hasil treemap didapatkan gambaran yang komprehensif secara ringkas mengenai distribusi geografis dan genetik dari data *monkeypox*. Informasi ini dapat sangat berguna untuk penelitian lebih lanjut, misalnya untuk mengidentifikasi pusat keanekaragaman hayati, melacak asal-usul genetik suatu spesies, atau mempelajari pola migrasi.



Gambar 5.4 Peta penyebaran *Monkeypox* berdasarkan Negara dan Benua

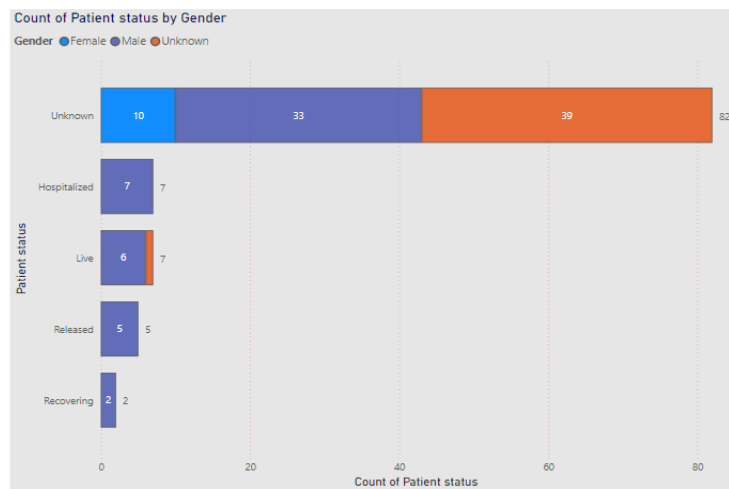
Berdasarkan visualisasi tersebut menunjukkan penyebaran global kasus *monkeypox* dengan distribusi kasus yang tersebar di berbagai wilayah dunia, mencakup Amerika, Afrika, Asia, Eropa, dan Australia, yang mengindikasikan bahwa penyakit ini telah meluas secara internasional dan tidak lagi terbatas pada wilayah endemiknya. Dalam visualisasi ini, setiap titik di peta merepresentasikan negara tertentu, dengan warna-warna berbeda yang menggambarkan status pasien, seperti kasus aktif (*Live*) yang ditandai dengan warna biru, pasien dalam pemulihan (*recovering*) di tandai dengan warna *orange*, pasien yang telah dinyatakan sembuh (*Released*) ditandai dengan warna ungu, adapun pasien dengan status *hopsitized* yang berarti sedang dalam perawatan di tandai dengan warna biru muda. Dan yang terakhir adalah status tidak di ketahui (*unknown*) yang ditandai

dengan warna pink, status ini cukup banyak, mengindikasikan ada keterbatasan data atau pelaporan yang belum lengkap di beberapa wilayah. Selain itu pada visualisasi yang disajikan oleh **gambar 5.4** terdapat perbedaan ukuran titik (*buble size*) yang berarti semakin besar ukuran titik maka semakin besar pula adanya jumlah kasus atau entri data pasien pertama dengan status tertentu di negara tersebut. Sebagai contoh terdapat ukuran titik yang besar yang tersebar di benua afrika dan benua amerika, yang bisa di artikan bahwasanya kasus monkeypox relative tinggi di wilayah tersebut. Distribusi pada warna memberikan gambaran yang jelas mengenai jumlah kasus dalam setiap kategori di berbagai negara, serta memperlihatkan bahwa di beberapa wilayah, terutama di Afrika, terdapat konsentrasi kasus yang lebih tinggi, sesuai dengan statusnya sebagai wilayah endemik. Sebaliknya, di wilayah non-endemik seperti Eropa atau Amerika Utara, terdapat titik-titik yang menunjukkan penyebaran lintas negara akibat mobilitas global. Selain itu, informasi tambahan mengenai gender pasien pertama yang tercatat di masing-masing wilayah memberikan konteks epidemiologis awal tentang pola infeksi yang dapat membantu dalam memahami dinamika penyebaran penyakit. Secara keseluruhan, visualisasi tidak hanya mencerminkan pola distribusi geografis *monkeypox* tetapi juga menyoroti pentingnya *surveilans* kesehatan yang efektif untuk mendeteksi dan melaporkan kasus, seperti yang terlihat dari negara-negara dengan banyak titik. Penyebaran yang terilustrasi dalam peta ini juga menggarisbawahi perlunya kerja sama global dalam pengendalian penyakit, baik melalui identifikasi kasus, penanganan pasien, maupun pencegahan penyebaran lebih lanjut.



Gambar 5.5 *Monkeypox* Berdasarkan kategori umur

Berdasarkan hasil visualisasi rata-rata umur yang terkena virus *monkeypox* dibagi menjadi 7 kategori usia, diantara ialah kategori dengan *Unknown* dengan (42 sampel korban) , kategori 0-9 tahun dengan (2 sampel korban), kategori 10-19 tahun terdapat (2 sampel korban), adapun untuk kategori umur 20-29 terdapat (26 sampel korban), lalu untuk kategori 30-39 terdapat (20 sampel korban), Selain itu kategori dengan umur 40-49 terdapat (1 sampel) dan yang terakhir adalah kategori >49 dengan hanya (1 sampel korban). Berdasarkan kesimpulan hasil maka kategori dengan *Unknown* adalah jumlah terbanyak.



Gambar 5.6 *Monkeypox* berdasarkan *Gender* dan Kondisi Pasien.

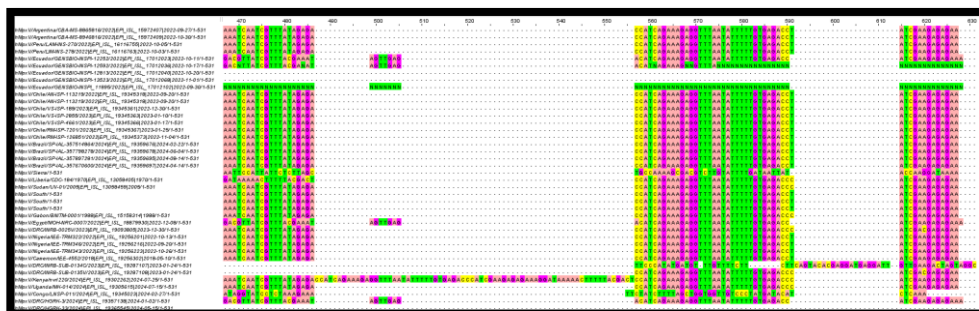
Berdasarkan hasil visualisasi penyebaran *monkeypox* terjadi di beberapa negara hingga benua, yang mana dari data korban tersebut banyak dijadikan sampel penelitian, salah satunya adalah dikategorikan berdasarkan riwayat pasien yang menjadi 5 kategori, yaitu *Unknown*, *Hospitalized*, *Live*, *Released* dan *Recovering*. Selain kategori riwayat pasien, terdapat pula kategori gender yang dibagi menjadi 3 kategori jenis diantara ialah *Female*, *Male* dan *Unknown*. Kesimpulan berdasarkan hasil visualisasi diatas ialah sebagai berikut :

1. ***Unknown (Tidak diketahui)***: Terdapat 10 orang wanita dengan kategori *sample* yang tidak diketahui, lalu ada 33 laki-laki yang riwayat nya tidak diketahui, dan ada kategori jenis kelamin yang tidak di ketahui (*unknown*) yang bahkan riwayat pasiennya pun tidak di ketahui (*unknown*) dengan jumlah terbanyak 39 sampel kasus.

2. **Hospitalized (Dalam perawatan Rumah Sakit)** : Di sisi lain untuk kategori *hospitalized* hanya terdapat 7 orang wanita yang terkena *Monkeypox* dari semua data yang di analisis.
3. **Live (Aktif Hidup)** : Adapun untuk kategori *Live* hanya terdapat 6 orang pria yang selamat dan 1 orang dengan jenis kelamin *Unknown*. Namun perlu dalam pengawasan.
4. **Released (Telah Sembuh/ Keluar dari Perawatan)** : Untuk kategori *released* hanya terdapat 5 orang pria yang terkena *Monkeypox*.
5. **Recovering (Sedang dalam Proses Sembuh)** : Adapun untuk kategori *Recovering* hanya terdapat 2 pria yang terkena *Monkeypox*.

5.2. Analisis Phylogenetics

Penelitian ini menggunakan data dari GISAID, yang kemudian dibaca dan di kombinasikan dengan menggunakan *Jalview* (*Waterhouse et al., 2009*). Untuk melakukan proses analisis *phylogenetics* dengan menggunakan metode *maximum likelihood* dibutuhkan proses penyelarasan urutan (*multiple sequence alignment, MSA*) pada semua sampel DNA menggunakan *Multiple Alignment using Fast Fourier Transform* (MAFFT) versi 7.490 (*Katoh & Standley, 2013*) melalui perintah *command-line* atau perintah *PowerShell*. Berikut adalah tampilan urutan yang telah diselaraskan:



Gambar 5.7 (*Multiple Sequence Alignment, MSA by MAFFT*)

Gambar 5.7 menunjukkan hasil penyelarasan urutan (*multiple sequence alignment, MSA*) dari berbagai sampel genom *Monkeypox*. Setelah proses penyelarasan sekuens (*multiple sequence alignment, MSA*) dilakukan menggunakan MAFFT, langkah *trimming* dilakukan untuk menyaring bagian-bagian alignment yang tidak sejajar secara optimal. Beberapa wilayah, khususnya

pada bagian terminal atau segmen yang tidak konservatif, menunjukkan tingkat keterisian (*occupancy*) yang rendah serta mengandung banyak celah (*gap*), yang dapat menyebabkan artefak pada analisis lanjutan. Oleh karena itu, trimming diterapkan untuk mempertahankan hanya posisi nukleotida yang sejajar secara konsisten dan konservatif di antara sekuens.

Proses ini penting untuk meningkatkan keandalan interpretasi, terutama dalam analisis filogenetika dan pencarian motif, karena wilayah yang tidak sejajar dengan baik berpotensi menghasilkan hasil yang bias atau tidak representatif. Dengan memfokuskan hanya pada daerah yang secara evolusioner stabil, trimming tidak hanya menyederhanakan kompleksitas data, tetapi juga meningkatkan akurasi inferensi filogenetika, seperti estimasi hubungan kekerabatan, perhitungan jarak genetik, dan penentuan posisi konsensus nukleotida. Selain itu, trimming membantu menghilangkan noise yang dapat mengganggu pemodelan evolusi, menjadikan dataset yang dihasilkan lebih representatif dan bermakna secara biologis.

Di sisi kiri **gambar 5.7**, terdapat daftar nama atau ID sampel yang dianalisis, yang kemungkinan berasal dari database seperti GISAID dan mewakili berbagai isolat genom. Bagian tengah gambar menampilkan urutan yang telah diselaraskan, di mana setiap kolom vertikal merepresentasikan posisi spesifik dalam genom yang diselaraskan untuk semua sampel. Warna pada kolom penyelarasan memberikan informasi penting. Warna hijau menunjukkan bahwa basa atau residu pada posisi tertentu sesuai dengan genom referensi. Sebaliknya, warna kuning, pink, atau ungu mengindikasikan adanya variasi atau mutasi pada posisi tersebut dibandingkan dengan genom referensi. Celah atau *gap* dalam urutan ditampilkan sebagai hitam atau titik-titik, yang menandakan adanya kekosongan dalam urutan akibat proses penyelarasan. Pada kolom pertama, yang menggunakan adenin (A) sebagai referensi, hanya sedikit baris yang menunjukkan kesesuaian dengan genom referensi. Sebagian besar baris lain menunjukkan adanya variasi atau basa yang berbeda. Secara keseluruhan, pola warna dalam gambar ini memberikan informasi tentang variasi dan mutasi pada posisi genom tertentu. Area dengan warna yang seragam menandakan tingkat konservasi tinggi, sedangkan area dengan variasi warna menunjukkan posisi dengan tingkat mutasi yang lebih besar di antara sampel.

Gambar ini berguna untuk memahami pola mutasi dan konservasi pada genom yang dianalisis.



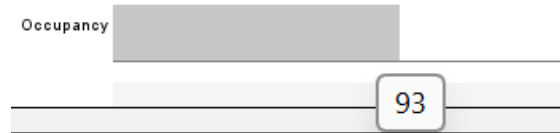
Gambar 5.8 *Consensus Adenin*

Berdasarkan **gambar 5.8** menunjukkan hasil dari penyelarasan urutan (*multiple sequence alignment*) dengan fokus pada bagian konsensus dari urutan yang diselaraskan. Di bagian atas gambar, terdapat urutan konsensus yang ditandai dengan warna-warni, yang mewakili basa DNA yang muncul paling sering di setiap posisi dari urutan yang diselaraskan. Presentase 69.8% menunjukkan bahwa 69.8% dari pohon filogenetik yang dihasilkan menunjukkan kesamaan dalam struktur atau hubungan antar spesies. Ini bisa mencakup cabang yang sama atau kelompok taksonomi yang konsisten, meskipun ada perbedaan pada sisa 30.2%. Angka ini memberikan gambaran tentang seberapa stabil atau dapat diandalkannya pohon filogenetik yang dihasilkan berdasarkan data yang dianalisis. Semakin tinggi persentase konsensus, semakin kuat dukungan terhadap hipotesis filogenetik yang dihasilkan. Adapun warna-warna ini biasanya menunjukkan tingkat keberagaman atau kesamaan basa pada posisi tertentu:

- Hijau biasanya mewakili basa yang sering muncul di posisi tersebut, yang menunjukkan konsensus.
- Warna lain seperti ungu, merah menunjukkan variasi atau mutasi yang terjadi pada posisi tertentu dibandingkan dengan konsensus yang ada.

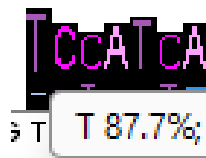
Di bawah urutan konsensus, ada beberapa urutan yang diselaraskan yang menunjukkan variasi atau perbedaan dari konsensus tersebut. Urutan yang ada di bagian bawah menunjukkan urutan asli dari sampel genom yang diselaraskan, yang mana beberapa posisi tidak sesuai dengan urutan konsensus dan berwarna berbeda. Gambar tersebut menggambarkan bagaimana variasi dan mutasi terlihat dalam

analisis penyelarasan urutan DNA, di mana konsensus menggambarkan pola umum sementara perbedaan di bawahnya menunjukkan mutasi atau *polimorfisme* genetik.



Gambar 5.9 *Occupancy Adenin*

Berdasarkan visualisasi pada **gambar 5.9** di atas menunjukkan indikator *occupancy* dalam konteks penyelarasan urutan (*multiple sequence alignment* atau MSA). Nilai *occupancy* sebesar 93, yang berarti bahwa 93% dari posisi dalam kolom tertentu telah diisi dengan basa yang sesuai dengan urutan referensi. Indikator ini menggambarkan tingkat kesesuaian atau konservasi di antara sampel yang diselaraskan pada posisi tersebut. Dalam hal ini, nilai 93% menunjukkan bahwa sebagian besar urutan pada posisi tertentu hampir sepenuhnya sesuai dengan genom referensi, dengan sedikit variasi atau mutasi. Seiring meningkatnya nilai *occupancy*, semakin tinggi pula tingkat kesesuaian antara posisi tersebut dengan referensi.



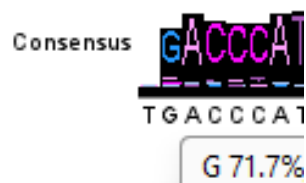
Gambar 5.10 *Concensus (Timin)*

Berdasarkan **gambar 5.10** Concensus pada *timin* menunjukkan hasil dari penyelarasan urutan (*multiple sequence alignment*) dengan fokus pada bagian konsensus dari urutan yang diselaraskan. Presentase menunjukkan 87.7% bahwa dari pohon filogenetik yang dihasilkan menunjukkan kesamaan dalam struktur atau hubungan antar spesies. Ini bisa mencakup cabang yang sama atau kelompok taksonomi yang konsisten, meskipun ada perbedaan pada sisa 13.3%.



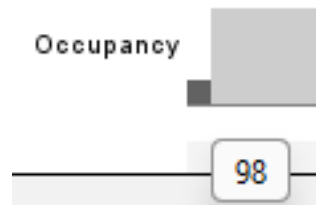
Gambar 5.11 *Occupancy (Timin)*

Berdasarkan **gambar 5.11** menunjukkan indikator *occupancy* dalam konteks penyelarasan urutan (*multiple sequence alignment* atau MSA). Nilai *occupancy* sebesar 96, yang diartikan bahwa 96% dari posisi dalam kolom tertentu telah diisi dengan basa yang sesuai dengan urutan referensi. Indikator ini menggambarkan tingkat kesesuaian atau konservasi di antara sampel yang diselaraskan pada posisi tersebut. Dalam hal ini, nilai 96% menunjukkan bahwa sebagian besar urutan pada posisi tertentu hampir sepenuhnya sesuai dengan genom referensi, dengan sedikit variasi atau mutasi.



Gambar 5.12 *Concensus (Guanine)*

Pada **Gambar 5.12** menampilkan referensi *genome sequence* yang menggambarkan urutan konsensus DNA dari sekelompok genom virus *monkeypox* hasil MSA (*multiple sequence alignment*). Pada bagian atas, urutan konsensus yang ditampilkan adalah “GACCCAT”, menunjukkan basa-basa nukleotida yang paling sering muncul pada setiap posisi di antara seluruh sekuens yang dianalisis. Tinggi setiap huruf merepresentasikan frekuensi kemunculannya, di mana huruf yang lebih tinggi (seperti G pada posisi pertama dengan frekuensi 71,7%) menunjukkan tingkat konservasi yang tinggi di posisi tersebut. Adapun pada baris “TGACCCAT” di bawahnya merepresentasikan urutan referensi dari genom *monkeypox* yang dibandingkan terhadap urutan konsensus. Adanya kesesuaian antara urutan konsensus dan urutan referensi pada sebagian besar posisi menandakan bahwa bagian tersebut merupakan wilayah genom yang relatif konservatif, kemungkinan memiliki peran penting secara biologis, seperti dalam pengkodean protein esensial atau elemen regulator. Dengan demikian, visualisasi ini memberikan informasi penting mengenai pola konservasi dan variasi nukleotida dalam genom virus Monkeypox, yang dapat digunakan untuk mengidentifikasi lokasi-lokasi penting bagi fungsi virus maupun potensi target diagnostik dan terapi.



Gambar 5.13 *Occupancy (Guanine)*

Berdasarkan visualisasi **gambar 3.13** *occupancy (Guanine)*, posisi ini memiliki nilai keterisian sebesar 98%, menunjukkan bahwa hampir seluruh sekuens dalam alignment memiliki nukleotida di posisi tersebut tanpa adanya gap. Hal ini mengindikasikan bahwa posisi tersebut bersifat konservatif dan penting secara evolusioner dalam struktur genom *monkeypox* yang dianalisis.



Gambar 5.14 *Concensus (Cytosyn)*

Pada **gambar 5.14** visualisasi sequence logo pada segmen genom *monkeypox* menunjukkan motif konsensus "AACGTGAGC", dengan dominasi nukleotida C pada posisi terakhir yang mencapai 84,0% frekuensi kemunculan. Tingginya frekuensi huruf tersebut mencerminkan tingkat konservasi yang tinggi antar-isolat virus pada posisi tersebut. Selain itu, pola urutan yang relatif stabil di sepanjang motif ini mengindikasikan bahwa wilayah tersebut kemungkinan merupakan bagian dari elemen fungsional penting, seperti promotor, regulator, atau situs pengikatan protein dalam genom virus. Konsensus yang kuat dalam urutan ini menegaskan potensi penggunaannya sebagai target dalam desain primer diagnostik maupun intervensi molekuler yang presisi.



Gambar 5.15 *Occupancy (Cytosyn)*

Berdasarkan **gambar 5.15** analisis *multiple sequence alignment*, Genome referensi nilai *occupancy* pada posisi nukleotida ini mencapai 97%, yang berarti bahwa sebanyak 97% dari total sekuens yang dianalisis memiliki basa nukleotida

pada posisi tersebut tanpa celah (gap). Tingginya tingkat keterisian ini mencerminkan bahwa posisi tersebut sangat konservatif dan jarang mengalami delesi di antara berbagai isolat virus Monkeypox. Hal ini memperkuat dugaan bahwa segmen ini merupakan bagian yang secara struktural atau fungsional penting, berpotensi terlibat dalam elemen genetik esensial seperti wilayah pengkode protein atau domain regulatoris yang dilestarikan selama evolusi virus. Posisi dengan nilai *occupancy* tinggi seperti ini sangat ideal untuk dijadikan fokus dalam studi komparatif evolusioner maupun pengembangan target molekuler untuk deteksi atau terapi.

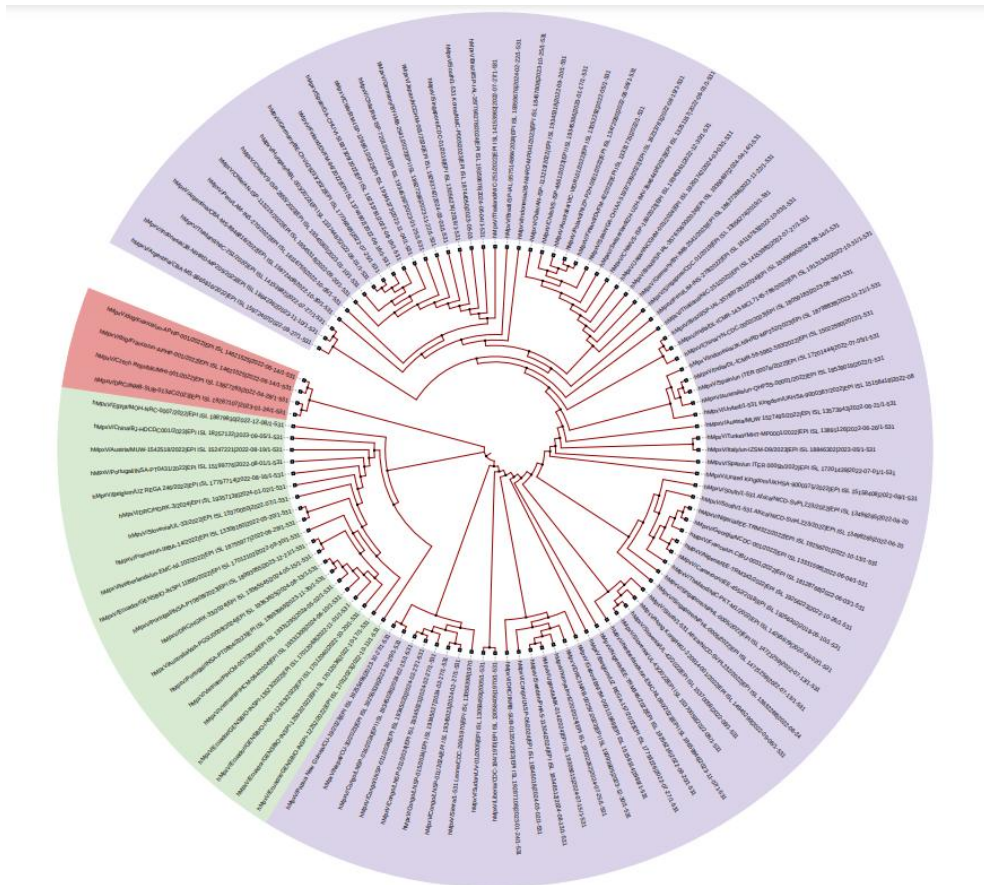
Name	Value
Type	Unrooted
# of Taxa	106
SBL	5.24487388
Log Likelihood	-6962.42
Model	GTR
Ts/Tv	1.4444

Gambar 5.16 *MX Information*

Analisis filogenetik yang dilakukan menghasilkan pohon *unrooted*, yang berarti pohon tersebut menggambarkan hubungan evolusi antar taksa tanpa menentukan nenek moyang atau arah evolusi secara spesifik. Pohon ini mencakup 106 taksa, yang merupakan jumlah unit analisis (organisme, spesies, atau sampel) yang digunakan dalam penelitian. Panjang total cabang dalam pohon (*SBL - Sum of Branch Lengths*) adalah 5.24487388. Nilai ini mencerminkan jumlah perubahan evolusi yang diasumsikan telah terjadi untuk menghasilkan hubungan antar taksa tersebut. Nilai panjang cabang yang relatif rendah menunjukkan bahwa pohon ini memiliki tingkat kompleksitas perubahan evolusi yang dapat diterima.

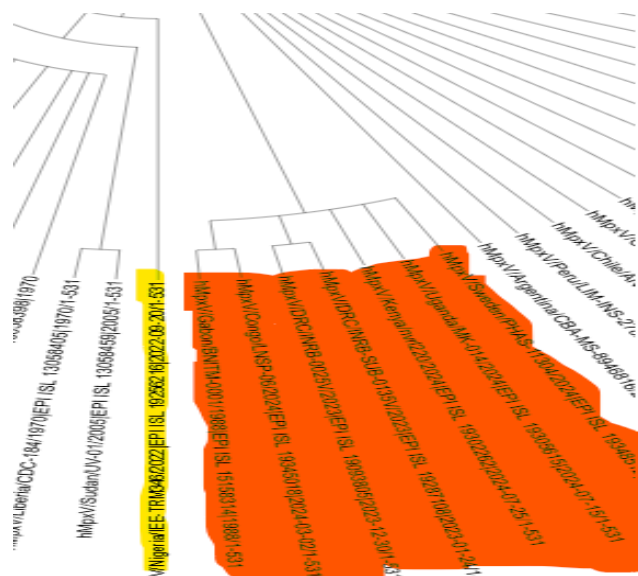
Model evolusi yang digunakan dalam analisis ini adalah GTR (*General Time Reversible*), yang merupakan model substitusi nukleotida yang fleksibel dan memungkinkan tingkat perubahan yang berbeda antar pasangan basa (A, T, C, G). Model GTR sangat cocok untuk data molekuler karena kemampuannya menangkap keragaman pola substitusi dalam evolusi DNA. Nilai *Log Likelihood* dari pohon ini adalah -6962.42, yang menunjukkan tingkat kecocokan pohon terhadap data sekuens yang digunakan. Meskipun nilai ini cukup besar dalam konteks negatif, hal

ini menunjukkan bahwa pohon yang dihasilkan masih memiliki kemampuan menjelaskan pola perubahan evolusi dalam data, meskipun dengan kompleksitas tertentu. Selain itu, rasio Ts/Tv (*Transition/Transversion*) adalah 1.4444. Nilai ini menunjukkan bahwa transisi (pergantian nukleotida antara purin ↔ purin atau pirimidin ↔ pirimidin) lebih sering terjadi dibandingkan transversi (pergantian nukleotida antara purin ↔ pirimidin). Hal ini sejalan dengan pola evolusi molekuler umum, di mana transisi cenderung lebih sering muncul karena perubahan tersebut lebih sedikit mengganggu struktur DNA. Secara keseluruhan, pohon filogenetik yang dihasilkan menggambarkan hubungan evolusi antara 106 taksa dengan menggunakan pendekatan model GTR, di mana transisi lebih dominan dibandingkan transversi. Pohon ini memiliki panjang total cabang yang moderat dan log likelihood yang menunjukkan tingkat kecocokan yang layak terhadap data evolusi molekuler yang dianalisis. Berikut adalah hasil dari *phylogenetics Tree*.



Gambar 5.17 *Phylogenetics Tree By Maximum Likelihood*

Pada **gambar 5.17** pohon menggunakan metode *maximum-likelihood* dengan 1000 bootstrapping. Proses perhitungan dan pemodelan *maximum-likelihood* dilakukan menggunakan Mega X. Berdasarkan hasil pohon pada **gambar 5.17**, beberapa klad ditemukan sesuai dengan studi kasus, yaitu *I*, *Ia* dan *Ib*. Pada kasus ini harus mengembangkan sistem nomenklatur untuk *clade* utama melalui GISAID. Pengenalan sistem nomenklatur ini disebabkan oleh variasi genetik virus *Monkeypox* yang berkembang secara alami. Berdasarkan mutasi penanda pada kelompok filogenetika tingkat tinggi akan dilakukan penentuan garis keturunan (*lineage*) virus yang ditemukan dalam beberapa sampel dari berbagai negara. Dapat dilihat melalui **Gambar 5.17** bahwa pohon filogenetika di visualisasikan dengan bentuk circular, yang gambarkan setiap cabang memiliki tingkat kemiripan yang tinggi, yang berarti Identik bahwa genom virus mengalami evolusi.



Gambar 5.18 Phylogenetics Tree circular sample

Penjelasan pada **gambar 5.18** peneliti mengambil 6 sampel dari hasil analisis filogenetika yang sudah berhasil menampilkan hasil resmi dari pohon filogenetika yang ada pada **gambar 5.17**. dengan mengambil 6 sampel *orange* dari *clade* I yang berasal dari kongo, gabon dan DRC, negara tersebut termasuk negara yang mayoritas terkenal asal muasal wabah virus *monkeypox*, Namun setelah dilakukannya analisis filogenetika, bahwa negara tersebut satu akar dan satu *clade*, tentu hal ini dikarenakan taksomi yang sama dan posisi evolusioner yang mirip, serta terjadinya evolusi kompleks yang menyebabkan kelompok tersebut membentuk satu

clade murni. Jika dibandingkan dengan sampel dari Nigeria dengan *clade* IIB A.2 yang ditandai dengan warna kuning tentu berbeda yang artinya dua spesies dari genus yang sama mungkin tidak terlalu mirip secara genetik karena salah satu sampel mengalami evolusi cepat atau karena salah satunya memiliki urutan yang lebih mirip dengan kelompok lain. Atau juga bisa diartikan sekelompok sekuens berwarna *orange* dan kuning membentuk *clade* tersendiri yang berasal dari percabangan terpisah dari kelompok lainnya, mengindikasikan bahwa sampel memiliki nenek moyang yang berbeda dan menyimpang lebih awal dalam evolusi. Hal ini menunjukkan adanya divergensi evolusioner yang signifikan dan potensi adanya struktur populasi yang berbeda.

Selain itu berdasarkan struktur filogenetika utama, Pohon di atas terbagi menjadi tiga kelompok utama yang ditandai dengan warna berbeda:

- Kelompok hijau mewakili *strain* yang memiliki hubungan genetik yang lebih dekat, menunjukkan kemungkinan asal geografis yang sama atau penyebaran yang terbatas. Adapun lineage kelompok utama yang ditandai dengan warna hijau yang tergolong menjadi beberapa *clade* yaitu *clade* I, dan IIB, adapun untuk *sub-clade* diantaranya adalah IIB B.1 , IIB B.1.20 , IIB C.1.1, IIB B.1, IIB B.1, IIB B.1, IIB B.1, IIB B.1, IIB B.1, IIB B.1, IIB B.1.20, IIB B.1.20, IIB C.1.1, IIB B.1.20, IIB B.1, IIB C.1, IIB C.1, IIB B.1.20 dan IIB C.1.1. Yang bisa disimpulkan bahwa *clade* I dan II adalah induk *clade* dan bisa dijadikan sebagai referensi asal muasal virus tersebut.
- Kelompok merah merupakan kluster terpisah yang mencerminkan kemungkinan strain unik dengan pola divergensi yang signifikan. Hal ini dapat mengindikasikan adaptasi lokal atau mutasi spesifik, selain itu *clade* dan *sub-clade* yang termasuk kategori golongan yang ditandai dengan warna merah adalah rata-rata induk *clade* berikut ialah I dan II yang merupakan referensi asal muasal virus tersebut muncul, kemudian *sub-clade* IIB B.1, IIB B.1.3, IIB 1.3 dan Ia yang merupakan evolusi dari *clade* I dan IIB.
- Kelompok ungu mencakup sebagian besar strain, menandakan keragaman genetik yang lebih tinggi di antara kelompok ini. Hal ini dapat menunjukkan

penyebaran geografis yang luas atau evolusi yang beragam seperti IIB A , IIB A, I, IIB A, IIB A, IIB A, Ia , IIA, Ia, Ia, I, I, Ib, Ib, IIB B.1 , IIB A.2, IIB C.1.1, IIB B.1.7, IIB B.1, IIB B.1.2 , IIB B.1.2 , IIB B.1.7 , IIB A.2 , IIB , IIB A.2 , IIB B.1 , IIB B.1 , IIB A.2.2, IIB C.1 , IIB C.1.1, IIB B.1, IIB B.1.6, IIB C.1.1, IIB C.1, IIB B.1, IIB B.1, IIB B.1, IIB B.1, IIB B.1.10, IIB B.1 dan IIB B.1.

Cabang-cabang pendek dalam kelompok menunjukkan hubungan evolusi yang lebih erat dengan sedikit variasi genetik. Hal ini dapat dikaitkan dengan penyebaran baru atau populasi virus yang masih homogen. Selain itu juga hubungan evolusi dan divergensi yang merupakan bagian-bagian cabang-cabang panjang menunjukkan divergensi genetik yang lebih besar, mencerminkan akumulasi mutasi dari strain nenek moyang. Disisi lain juga terdapat implikasi epidemiologis yang dimana ditandai dengan kluster hijau dan merah dapat mengindikasikan asal-usul lokal atau wabah terbatas, sementara keragaman dalam kluster ungu mencerminkan kemampuan adaptasi virus terhadap berbagai lingkungan atau populasi inang. Dari hal tersebut struktur pohon ini dapat membantu dalam pelacakan asal-usul wabah dan penyebaran strain virus *monkeypox* di berbagai wilayah geografis. Adapun kluster bisa berpotensi mengadaptasi dan evolusi, seperti halnya kluster merah yang jelas terpisah dari kelompok lain menunjukkan strain yang berpotensi memiliki mutasi unik. Hal ini dapat dikaitkan dengan perbedaan patogenisitas, transmisi, atau respon terhadap tekanan seleksi lingkungan, seperti imunitas inang. Berikut adalah beberapa penjelasan dari setiap *clade* dan *sub-clade*.

CLADE I

Clade I dari virus *Monkeypox* pertama kali diidentifikasi di wilayah endemik Afrika Tengah, terutama di Republik Demokratik Kongo (DRC), yang dikenal sebagai salah satu daerah dengan tingkat keendeman tinggi untuk virus ini. Virus *Monkeypox* memiliki reservoir utama pada hewan liar, khususnya hewan pengerat dan primata, yang berperan penting dalam mempertahankan siklus virus di alam. Transmisi kepada manusia umumnya terjadi melalui kontak langsung dengan cairan tubuh, lesi kulit, atau droplet dari individu yang terinfeksi, serta melalui

paparan terhadap hewan liar yang terinfeksi. *Clade I* dikenal dengan tingkat virulensi yang lebih tinggi dibandingkan *Clade II*, sering kali dikaitkan dengan tingkat mortalitas yang lebih tinggi dan gejala klinis yang lebih parah. Data epidemiologis menunjukkan bahwa *Clade I* mendominasi laporan kasus di Afrika Tengah, dengan lebih dari 19.667 kasus yang dicurigai atau terkonfirmasi dan 575 kematian hingga saat ini. Kasus-kasus ini mewakili lebih dari 90% laporan kasus terkait *Monkeypox* di Afrika ([European Centre for Disease Prevention and Control, 2024](#)). DRC tetap menjadi episentrum *Clade I*, di mana infeksi endemik terus terjadi, terutama di komunitas yang memiliki interaksi tinggi dengan satwa liar. Selain itu, negara-negara Afrika lainnya, termasuk Nigeria, Kenya, Republik Afrika Tengah, Rwanda, dan Uganda, juga melaporkan kasus *Clade I*. Sebagian besar kasus di negara-negara ini bersifat lokal dan ditransmisikan melalui kontak domestik atau lingkungan yang padat.

Distribusi global *Clade I* menunjukkan bahwa lebih dari 90% kasus tetap terkonsentrasi di Afrika, dengan negara-negara berikut mencatat sebagian besar kasus diantaranya ialah Republik Demokratik Kongo (DRC) 42%, Nigeria (23%) , Amerika Serikat (15%) (kasus impor), Swedia (5%) (kasus impor), Negara-negara Afrika lainnya seperti (Kenya, Republik Afrika Tengah, Rwanda, Uganda) sisanya merupakan Potensi transmisi lintas benua dari *Clade I* seperti (sierra lion), meskipun masih rendah, menjadi perhatian penting, terutama karena kasus impor dapat memperkenalkan virus ke wilayah yang sebelumnya tidak memiliki riwayat infeksi lokal. Upaya pengawasan global terhadap *Clade I*, termasuk analisis filogenetika dan pemetaan distribusi, sangat penting untuk memahami evolusi virus, penyebarannya, dan risiko terhadap kesehatan masyarakat global. Selain itu virus *monkeypox* (*Monkeypox*) *Clade I* juga memiliki dua sub-clade utama, yaitu *Clade Ia* dan *Clade Ib*. Kedua sub-clade ini pertama kali teridentifikasi di wilayah endemik *Monkeypox* di Afrika Tengah, dengan Republik Demokratik Kongo (DRC) sebagai pusat penyebaran utama. Sub-clade *Ia* lebih dominan ditemukan di DRC, sementara Sub-clade *Ib* tersebar lebih luas di negara-negara lain, termasuk Nigeria, Republik Afrika Tengah, dan Uganda. Berdasarkan kasus dan pengambilan data *Clade I* juga memiliki *sub-clade* yaitu *Ia* dan *Ib*, berikut keteranganya :

Ia : *Sub-clade* Ia memiliki karakteristik genetik yang sangat serupa dengan *Clade I* secara keseluruhan dan menunjukkan tingkat virulensi yang lebih tinggi dibandingkan dengan *clade* lainnya. Sebagian besar kasus *Clade* Ia bersifat lokal dengan prevalensi tinggi di wilayah Afrika Tengah yang memiliki intensitas interaksi manusia-hewan yang tinggi, seperti di Kenya dan Uganda.

Ib : Sebaliknya, *Sub-clade* Ib menunjukkan distribusi yang lebih luas, ditemukan di negara-negara seperti Nigeria dan Republik Afrika Tengah, dengan penyebaran yang didukung oleh tingkat mobilitas manusia yang tinggi, seperti perdagangan lintas negara dan perjalanan internasional. Selain beberapa faktor yang disebutkan virus *Monkeypox* juga bisa tersebar karena adanya penyebaran geografis dan faktor lingkungan distribusi geografis kedua *sub-clade* Ia dan Ib, sebagian besar terkonsentrasi di wilayah endemik Afrika Tengah dan Afrika Barat. Faktor geografis berperan penting dalam penyebaran lintas negara, dengan kasus impor *Monkeypox* dari *Clade* Ia dan Ib tercatat di negara non-endemik seperti Amerika Serikat, Inggris, dan Swedia. Sebagai contoh, kasus pertama *Clade Ib* di Amerika Serikat dikaitkan dengan riwayat perjalanan dari wilayah endemik di Afrika. Selain itu, penyebaran virus dipengaruhi oleh faktor lingkungan seperti kepadatan penduduk, sanitasi yang buruk, dan keberadaan reservoir hewan, termasuk hewan pengerat dan primata, yang menjadi vektor utama dalam siklus alami *Monkeypox*. Perubahan iklim turut memengaruhi distribusi hewan reservoir, memperluas kemungkinan penyebaran virus. keberagaman genetik berdasarkan geografi juga memengaruhi keberagaman genetik virus *Monkeypox*. *Sub-clade* Ia lebih sering ditemukan di DRC, yang menjadi pusat sirkulasi virus, sedangkan *Sub-clade* Ib memiliki prevalensi lebih tinggi di Nigeria dan Republik Afrika Tengah. Variasi distribusi ini mencerminkan dinamika evolusi virus yang dipengaruhi oleh adaptasi terhadap kondisi lingkungan dan interaksi manusia, sehingga mempertegas pentingnya analisis geografis dalam memahami penyebaran dan evolusi virus *Monkeypox*.

CLADE II

Clade II virus *monkeypox* (*Monkeypox*) pertama kali ditemukan di wilayah Afrika Barat dan menunjukkan peningkatan prevalensi dalam beberapa tahun

terakhir akibat mobilitas manusia dan perdagangan satwa liar. Negara-negara dengan distribusi signifikan Clade II mencakup Nigeria Menjadi negara dengan prevalensi tertinggi *Clade II*, menyumbang sekitar 60% dari total kasus Clade II yang dilaporkan di Afrika Barat sejak 2017. Liberia, Pantai Gading, dan Sierra Leone Bersama-sama menyumbang sekitar 30% kasus Clade II di wilayah tersebut. Kasus di negara non-endemik dalam wabah global Monkeypox pada 2022–2023, Clade II, khususnya sub-clade IIB, dilaporkan di negara-negara non-endemik seperti Amerika Serikat (15%), Spanyol (12%), Inggris (10%), dan Brasil (8%). Dalam kasus sebelumnya *clade II Monkeypox* terutama ditularkan melalui diantaranya kontak langsung dengan hewan reservoir seperti tikus Afrika dan primata, yang menjadi sumber utama infeksi atau ada hal lain juga seperti penularan antarmanusia meskipun lebih jarang dibandingkan *Clade I*, *Clade II* dapat menyebar melalui kontak erat, termasuk hubungan seksual, kontak dengan cairan tubuh, dan lesi kulit pasien dan yang terakhir adalah paparan lingkungan yang bisa terjadi melalui sehingga terkontaminasi virus, seperti pakaian atau tempat tidur pasien. Dalam studi kasus, clade II memiliki beberapa *sub-clade*, seperti IIA, IIB, IIB B.1, IIB C.1, IIB 1.B.20, IIB C.1.1, IIB A, IIB A.2, IIB B.1.6, IIB B.1.7, IIB B.1.2, IIB C.1, IIB B.1.22, IIB A2, IIB A.1, IIB A.2.1, IIB B.1.1, IIB B.1.10, IIB B.1.17, IIB B.1.3, IIB A.2.2, IIB B.1.1, IIB B.1.10, IIB B.1.11, IIB B.1.20, IIB B.1.3, IIB B.1.6 dan IIB B.1.7. Berikut adalah deskripsi tiap *sub-clade*:

- IIA : *Clade IIA* virus *Monkeypox* terutama ditemukan di wilayah Afrika Barat, yang merupakan daerah endemik utama untuk clade ini. Distribusi spesifik Clade IIA mencakup Nigeria: Negara dengan prevalensi tertinggi untuk Clade IIA, menyumbang sekitar 40% dari total kasus yang teridentifikasi. Pantai Gading Berkontribusi sekitar 25%, dengan kasus-kasus sporadis yang dilaporkan selama beberapa dekade terakhir Liberia menyumbang sekitar 20% dari total kasus *Clade IIA*, Sierra Leone menyumbang sekitar 10%, dan Liberia dengan beberapa laporan wabah kecil. dari kasus yang dijabarkan diatas hal ini bisa menyebar ke berbagai negara hal ini terjadi karena melalui berbagai faktor yang diantaranya adalah seperti, kontak hewan virus ini ditemukan pada hewan liar seperti tikus Afrika dan primata non-manusia yang menjadi *reservoir* utama.

Kontak langsung atau paparan cairan tubuh hewan-hewan ini adalah rute utama infeksi. Kontak Antarmanusia, meskipun hal ini lebih jarang, kasus penularan langsung antar individu telah dilaporkan, terutama di lingkungan rumah tangga atau fasilitas perawatan kesehatan tanpa protokol pencegahan infeksi yang memadai. Dan yang terakhir adalah paparan lingkungan, Penularan juga dapat terjadi melalui bahan yang terkontaminasi, seperti pakaian, tempat tidur, atau barang-barang pribadi pasien yang terinfeksi. Selain faktor yang telah disebutkan geografis dan mobilitas juga menjadi faktor sebagian besar kasus *Clade IIa* terbatas pada wilayah endemik Afrika Barat, mobilitas manusia, seperti migrasi atau perdagangan satwa liar, telah menyebabkan penyebaran ke wilayah non-endemik. Namun, Clade IIa lebih jarang terdeteksi di luar Afrika dibandingkan dengan sub-clade IIB yang terkait dengan wabah global.

- IIB : *Clade IIB* telah berkembang menjadi *sub-clade* dominan dalam wabah global Monkeypox, dengan pola penyebaran yang melibatkan penularan langsung antarmanusia dan pergerakan internasional. Wabah ini menekankan pentingnya pengawasan global, pengendalian wabah lintas negara, dan respons kesehatan masyarakat yang terkoordinasi. *Clade IIB* adalah *sub-clade* dari *Clade II* virus Monkeypox yang muncul sebagai varian dominan dalam wabah global tahun 2022-2023. Tidak seperti Clade IIa yang terkonsentrasi di Afrika Barat, Clade IIB menunjukkan distribusi internasional dengan kasus signifikan di negara-negara non-endemik. Berdasarkan WHO (*World Health Organization*) berikut adalah distribusinya berdasarkan data global Afrika Barat (Wilayah Endemik) Clade IIB pertama kali diidentifikasi di wilayah ini, menyumbang sekitar 20%-30% kasus awal sebelum menyebar secara global. Amerika Serikat, Kasus *Clade IIB* melonjak menjadi sekitar 25%-30% dari total kasus global selama wabah 2022-2023, sebagian besar melalui transmisi lokal di Eropa Negara-negara seperti Inggris, Spanyol, dan Jerman mencatatkan sekitar 25%-35% kasus Clade IIB, terkait dengan interaksi sosial dan perjalanan internasional. Asia dan Timur Tengah, Kasus sporadis dilaporkan di beberapa negara Asia, termasuk India dan Singapura, dengan kontribusi

sekitar 5%-10% dari total kasus. Berdasarkan Karakteristik Geografis, Amerika Utara dan Eropa , Menjadi pusat utama penyebaran Clade IIB selama wabah global. Afrika Barat, Tetap menjadi sumber utama Clade IIB, tetapi dengan pola penyebaran lokal yang lebih terkendali dibandingkan dengan penyebaran global. Asia dan Timur Tengah: Meskipun kontribusi kasus lebih kecil, wabah Clade IIB menunjukkan bahwa wilayah ini tidak sepenuhnya terlindungi dari penyebaran global.

- IIB B.1 : *Subclade Monkeypox IIB B.1* adalah salah satu garis keturunan (*lineage*) dari virus *Monkeypox* (*Monkeypox*) yang ditemukan dalam filogenetika. *Subclade* ini biasanya diidentifikasi melalui analisis sekuens genetik yang menunjukkan perbedaan evolusi dari subclade atau clade lainnya. *Subclade* IIB B.1 kemungkinan muncul melalui mutasi bertahap dalam genom virus *Monkeypox*. Mutasi ini mungkin terjadi akibat seleksi alam, adaptasi terhadap inang manusia, atau penyebaran geografis yang berulang kali. Kemunculan subclade ini sering kali terkait dengan perubahan tertentu pada wilayah genom virus yang relevan terhadap replikasi, transmisi, atau patogenisitas. Berdasarkan data penelitian *sub clade* IIB B.1 telah muncul di berbagai negara hingga benua seperti pada benua asia yang terjadi di negara georgia dan thailand, adapun untuk benua africa teridentifikasi adanya 1 pasien yaitu negara mesir, kemudia di susul oleh benua amerika yaitu chile, argentina dan ecuador, dan yang terakhir yang paling banyak terdapat pada benua europa seperti negara *spain, austria, france , german, finland, italy, netherland, slovenia , belgium* dan *portugal*. Berdasarkan laporan WHO dan analisis filogenetik wabah global 2022, *clade* IIB B.1 mendominasi lebih dari 80% hingga 90% kasus yang teridentifikasi di Eropa. Hal ini terjadi karena jaringan sosial yang kompleks seperti transmisi lokal di Eropa, transmisi *Monkeypox* sering kali terkait dengan kontak dekat dalam jaringan sosial tertentu, termasuk komunitas dengan interaksi intensif. Selain itu keterkaitan antar-negara konektivitas di Eropa mempermudah penyebaran virus lintas batas, terutama melalui transportasi darat, udara, dan kereta api yang sangat efisien.

- IIB C.1 : *Sub-clade* IIB C.1 adalah salah satu garis keturunan dalam clade IIB dari *Monkeypox*. Pada wabah 2022, *sub-clade* IIB C.1 menunjukkan pola penyebaran yang lebih luas dan lebih cepat dibandingkan dengan strain sebelumnya yang hanya ditemukan di wilayah endemik di Afrika Tengah dan Barat. Wabah ini menunjukkan bahwa penyebaran virus ini tidak hanya terbatas pada populasi yang berisiko tinggi, seperti mereka yang memiliki kontak langsung dengan hewan yang terinfeksi, tetapi juga menyebar lebih banyak di populasi umum, terutama melalui kontak manusia-ke-manusia. Namun pada kasus yang sekarang banyak negara yang terkena dengan varian ini, salah satunya terjadi di benua Asia, berdasarkan penelitian sebelumnya yang dilakukan Xingfei Ji et.al (2024), yaitu negara yang terdampak adalah Indonesia, Vietnam, Korea Selatan dan Jepang. Hal ini terjadi karena peningkatan kesadaran tentang *Monkeypox* dan penguatan sistem pelaporan di banyak negara, termasuk Indonesia, dapat berkontribusi pada lebih banyaknya kasus yang terdeteksi dan dilaporkan. Negara-negara yang sebelumnya tidak melihat banyak kasus *Monkeypox* mungkin baru mulai mendeteksi *sub-clade* IIB C.1 lebih sering setelah wabah 2022, karena sistem kesehatan mereka sekarang lebih peka terhadap jenis infeksi ini. Adapun WHO menunjukkan bahwa virus *Monkeypox*, termasuk *sub-clade* IIB C.1, dapat mengalami mutasi yang memungkinkan mereka beradaptasi lebih baik pada populasi manusia. Ini dapat meningkatkan penularannya antar individu. Variasi genetik tertentu pada subklade IIB C.1 mungkin memengaruhi kemampuannya untuk bertahan lebih lama di tubuh manusia atau menjadikannya lebih mudah menular melalui kontak dekat
- IIB 1.B.20 : *Sub-clade* IIB B.20 adalah salah satu garis keturunan dalam clade IIB dari *Monkeypox*. Seiring dengan penyebaran global *Monkeypox* selama wabah yang meningkat pada 2022, *sub-clade* IIB 1.B.20 juga mungkin ditemukan di berbagai wilayah, baik di negara endemik (seperti Afrika) maupun non-endemik (seperti Australia dan Amerika Utara). Namun berdasarkan data penelitian varian ini terjadi lebih banyak di benua Amerika seperti Brazil, Ecuador, Chile hal ini ditunjukkan karena letak

geografis yang sama, namun ada juga kasus yang terjadi australia yang mana hal tersebut dikarena impor dari benua amerika.

- IIB C.1.1 : Sub-clade ini merupakan turunan dari clade IIB, *Clade IIB* mencakup garis keturunan (*lineage*) yang bertanggung jawab atas wabah global pada tahun 2022. Garis keturunan C.1.1 adalah klasifikasi yang lebih spesifik di bawah Clade IIB. Garis keturunan ini telah dipelajari terkait dinamika penularan dan mutasi genetik yang mungkin memengaruhi tingkat keganasan, penularan, atau respons imun tubuh inang, adapun negara yang terdampak dari sub-clade ini seperti brazil dan china hal tersebut bisa terjadi karena, pengembangan mutasi yang meningkatkan kemampuan virus untuk menyebar di komunitas manusia, terutama melalui kontak dekat faktor lain juga terjadi karena ekologi dan lingkungan yang mana Amerika Latin memiliki keanekaragaman hayati yang tinggi, termasuk hewan yang berpotensi menjadi perantara penyebaran virus. Walaupun manusia adalah vektor utama dalam penyebaran wabah global, lingkungan lokal juga bisa memengaruhi pola infeksi. sedangkan di China, kasus cenderung lebih terkait dengan transmisi antar-manusia daripada faktor ekologis.
- IIB A : Merupakan *sub-clade* dari IIB yang mana *sub-clade* ini cenderung banyak ditemukan di kongo yang terinditifikasi langsung dengan induk cabang, adapun negara yang dilaporkan terdapat di africa dan europa serta asia seperti kongo, turkey dan india yang mana hal ini terjadi adanya transmisi.
- IIB A.2 : *Sub-clade* A.2 adalah salah satu cabang dalam Clade IIB, dengan mutasi genetik yang membedakannya dari subclade lain seperti IIB C atau A.1. *Sub-clade* A.2 menunjukkan mutasi tertentu pada genom *Monkeypox virus*, yang dapat memengaruhi kemampuan virus untuk menyebar, beradaptasi, atau menginfeksi manusia. Wilayah utama, Subclade A.2 terdeteksi di beberapa wilayah selama wabah global 2022, tetapi mungkin lebih dominan di wilayah tertentu dibandingkan subclade lainnya. Seperti yang terjadi di thaliand dan india. Penyebaran global terjadi melalui jaringan kontak erat, perjalanan internasional, dan acara sosial yang mempertemukan banyak orang.

- B.1.6 dan B.1.7 : Adalah cabang spesifik dalam garis keturunan B.1, yang muncul selama penyebaran virus ke berbagai wilayah di luar Afrika. Seperti terjadi di benua europa (rusia) dan amerika latin (peru) secara mutasi genetik , Kedua *subclade* menunjukkan mutasi genetik spesifik yang membedakannya dari subclade lain dalam garis keturunan B.1. Mutasi ini dapat memengaruhi berbagai aspek virus, termasuk, Efisiensi penularan antar manusia, Adaptasi terhadap inang manusia, Kemungkinan respons imun inang.
- IIB B.1.2 : IIB B.1.2 Garis keturunan bagian dari *sublineage* B.1 dalam Clade IIB. Yang memiliki karakteristik subclade ini muncul selama penyebaran global dan terkait dengan kluster penularan tertentu dan menunjukkan mutasi genetik yang membantu pelacakan rute epidemiologis. Adapun Penyebaran, Teridentifikasi di beberapa wilayah, terutama di negara-negara dengan aktivitas perjalanan internasional tinggi. Signifikansi Penelitian tentang B.1.2 membantu memahami dinamika penyebaran awal wabah global. Adapun garis keturunan, Subclade ini adalah cabang dari lineage B.1, yang merupakan bagian penting dari Clade IIB. Memiliki ciri-ciri genetik yang membedakannya dari sublineage B.1 lainnya, seperti B.1.2. Subclade ini muncul sebagai bagian dari evolusi virus saat menyebar di luar Afrika. Adapun Distribusinya ditemukan di wilayah tertentu dengan transmisi lokal yang signifikan. Disisi lain Sublineage dari A.2, bagian dari Clade IIB. Memiliki karakteristik mutasi yang membedakannya dari sublineage lain dalam Clade IIB. A.2 mungkin memiliki adaptasi genetik yang meningkatkan kemampuan virus untuk menyebar melalui kontak manusia. Penyebaran terdeteksi di berbagai negara selama wabah global. Penyebarannya sering dikaitkan dengan aktivitas sosial atau kontak erat di komunitas tertentu.
- IIB C.1 : Merupakan garis keturunan dari *sublineage* dalam Clade IIB yang cukup penting. Adapun karakteristik dari subclade ini menunjukkan mutasi unik yang membedakannya dari sublineage lain dalam Clade IIB. Berkaitan dengan peningkatan kemampuan virus untuk menyebar di antara manusia. Serta penyebaran Terlibat dalam wabah di berbagai negara selama tahun

2022 yang mungkin terkait dengan rute penyebaran yang berbeda dari lineage B.1. dengan memiliki Mutasi genetik, Mutasi pada C.1 memengaruhi penularan antar manusia, meskipun keparahan penyakit tetap rendah.

- Iib A.1: Merupakan *sub clade* turunan dari clade Iib , yang diperkirakan menjadi salah satu subklade awal yang muncul dalam wabah 2022. Varian ini diidentifikasi pada kasus-kasus pertama yang terdeteksi di Eropa pada awal tahun 2022. Seperti yang terjadi pada Mei 2022 kasus pertama wabah MPXV global dilaporkan di Inggris, melibatkan individu yang memiliki riwayat perjalanan ke Nigeria. Dalam waktu singkat, virus menyebar ke banyak negara di Eropa, Amerika Utara, dan belahan dunia lainnya.
- Iib A.2.1 : Merupakan hasil divergensi lebih lanjut dari sub-clade A.1. Subklade ini ditemukan pada kasus-kasus yang dilaporkan di berbagai wilayah pada pertengahan hingga akhir 2022, seperti yang terjadi di Pakistan seorang pasien berumur 22 tahun dinyatakan diagnosis MPXV dari *sub-clade Iib A.2.2*, namun pasien berhasil selamat dan dinyatakan sembuh.
- Iib B.1.1 : *Sub clade* ini juga diidentifikasi selama tahun 2022 dan menunjukkan penyebaran luas di berbagai negara, sering kali terkait dengan transmisi lokal salah satunya di benua Eropa yang berasal dari *Germany*.
- Iib B.1.10 : *Sub-clade* ini merupakan cabang lebih spesifik dari garis keturunan B.1. Penamaan seperti B.1.10 menunjukkan mutasi genetik yang unik yang membedakannya dari sub-clade lain dalam garis keturunan B.1. *Sub-clade* ini mungkin diidentifikasi melalui analisis genom lengkap dari virus yang diisolasi pada kasus-kasus tertentu. Munculnya *sub-clade* ini dapat dikaitkan dengan evolusi virus melalui akumulasi mutasi saat menyebar di berbagai populasi manusia, seperti yang terjadi di benua Amerika, Eropa dan Asia yang terdampak varian ini yaitu Kanada, Amerika, Inggris, Germany, Spain dan Singapura.
- Iib B.1.17 : *Sub-clade* ini merupakan cabang dalam garis keturunan B.1 yang teridentifikasi selama wabah global MPXV tahun 2022. Garis keturunan B.1 mendominasi kasus-kasus global di luar wilayah endemik.

Sub-clade ini diidentifikasi melalui analisis genomik virus yang diisolasi dari pasien selama wabah tahun 2022. B.1.17 kemungkinan muncul di tengah-tengah penyebaran global MPXV. Distribusi spesifik B.1.17 dapat melibatkan negara-negara dengan aktivitas genomik tinggi untuk melacak penyebaran MPXV, seperti Amerika Serikat, Inggris, dan negara-negara Eropa lainnya. Informasi lebih rinci membutuhkan data genomik terbaru. Penularan utama terjadi melalui kontak langsung dengan lesi kulit yang terinfeksi, cairan tubuh, atau krusta dari pasien yang memiliki gejala aktif. Permukaan atau benda yang terkontaminasi, seperti pakaian, tempat tidur, atau barang pribadi lainnya, dapat menjadi media transmisi. Kasus pertama di luar wilayah endemik sering dikaitkan dengan perjalanan internasional dari atau ke negara endemik di Afrika. Wabah 2022 melihat penyebaran virus secara cepat melalui jaringan perjalanan global, termasuk transportasi udara.

- Iib B.1.3 : *Sub-clade* ini dikenali melalui mutasi spesifik dalam genom virus *Monkeypox*, *Sub-clade* ini kemungkinan muncul selama fase awal wabah global 2022. Kasus-kasus terkait subklade ini terdeteksi di beberapa wilayah di luar daerah endemik, seperti Amerika Utara, Eropa, dan Asia. Distribusi geografis Seperti subklade lain dalam Clade Iib, B.1.3 menyebar ke negara-negara dengan jaringan perjalanan internasional yang aktif.
- Iib A.2.2 : Merupakan *sub-clade* yang bercabang dari garis keturunan A.2 dalam Clade Iib. *Sub-clade* ini ditentukan berdasarkan mutasi genomik spesifik yang membedakannya dari subklade lain dalam Clade Iib. Adapun identifikasi melalui genomik *sub-clade* ini diidentifikasi melalui analisis filogenetik virus yang diisolasi dari pasien selama wabah global. Selain hal itu awal penyebarannya juga kemungkinan berasal dari kasus-kasus awal dalam wabah 2022 di luar Afrika yang menyebar ke beberapa wilayah melalui perjalanan internasional dan kontak antar manusia. Kasus-kasus yang terkait dengan A.2.2 telah dilaporkan di berbagai negara, terutama di wilayah dengan koneksi perjalanan global. Adapun Signifikansi epidemiologi A.2.2 menunjukkan pola penyebaran serupa dengan *sub-clade* lain dari Clade Iib, melalui kontak langsung, droplet, dan permukaan yang

terkontaminasi, *sub-clade* ini berperan dalam wabah global karena melibatkan kasus-kasus yang menyebar melalui jaringan sosial dan mobilitas tinggi.

- Iib B.1.22 : *Sub-clade* Iib B.1.22 adalah salah satu garis keturunan dalam Clade Iib dari virus Monkeypox (MPXV). Clade Iib mencakup virus yang ditularkan antarmanusia dan telah terdeteksi dalam wabah global pada tahun 2022. Iib B.1.22 mempunyai karakteristik *seperti* posisi filogenetik B.1.22 merupakan subklade dari garis keturunan B.1 dalam Clade Iib. Analisis filogenetik menunjukkan bahwa subklade ini memiliki mutasi spesifik yang membedakannya dari subklade lain dalam Clade Iib. Adapun penyebaran berdasarkan geografis informasi spesifik mengenai distribusi geografis subklade B.1.22 saat ini terbatas. Namun, mengingat *Clade Iib* telah terdeteksi di berbagai negara selama wabah 2022, kemungkinan *sub-clade* ini juga memiliki distribusi yang luas. [Nextstrain.com](https://nextstrain.com).
- Iib B.1.1: *Sub-clade* B.1.1 bercabang dari garis keturunan B.1 yang mendominasi wabah global 2022. Dikenali melalui mutasi genetik unik yang membedakannya dari subklade lain dalam garis keturunan B.1. jika dilihat dari sisi analisis genomik identifikasi *sub-clade* B.1.1 dilakukan melalui analisis urutan genom virus yang diperoleh dari pasien selama wabah. *Sub-clade* ini mencerminkan adaptasi virus selama penyebarannya di populasi manusia. *Sub-clade* ini menyebar secara geografis yang telah dilaporkan di beberapa wilayah di luar daerah endemik, terutama di negara-negara dengan aktivitas perjalanan internasional tinggi. Negara-negara yang teridentifikasi memiliki kasus B.1.1 termasuk Amerika Utara dan beberapa negara di Eropa. Kemunculannya kemungkinan besar berkaitan dengan rantai transmisi tertentu dalam wabah tersebut.
- Iib B.1.10 dan Iib B.1.11 : *Sub-clade* Iib B.1.10 dan Iib B.1.11 pertama kali terdeteksi selama wabah global *monkeypox* pada tahun 2022. Keduanya berasal dari garis keturunan Clade Iib B.1 dan menyebar melalui kontak langsung dengan lesi atau cairan tubuh, droplet, serta permukaan yang terkontaminasi. Penyebaran pertama kali terjadi di negara-negara non-endemik, seperti Amerika Serikat dan Eropa, yang memiliki perjalanan

internasional aktif. Proses penyebaran virus ini melalui mobilitas manusia dan interaksi sosial yang intens. Disisi lain terdapat perbedaan utama antara kedua *sub-clade* ini terletak pada mutasi genetik yang membedakan mereka yang mana terjadi perubahan kecil dalam urutan DNA virus yang dapat mempengaruhi sifat-sifat virus, seperti kemampuannya untuk menginfeksi sel, penyebarannya antar individu, atau respons tubuh terhadap infeksi. Meskipun mutasi ini tidak selalu berarti virus menjadi lebih berbahaya, perubahan tersebut dapat mempengaruhi cara virus menyebar atau beradaptasi dengan lingkungan baru. *Sub-clade* B.1.10 dan B.1.11 menunjukkan pola penularan yang serupa, namun analisis terhadap mutasi ini penting untuk memantau potensi perubahan dalam karakteristik virus.

- IIB B.1.20 : *Sub-clade* IIB B.1.20 pertama kali terdeteksi selama wabah global *Monkeypox* pada tahun 2022. Subklade ini berasal dari garis keturunan Clade IIB B.1 dan menyebar melalui kontak langsung dengan lesi atau cairan tubuh, droplet, serta permukaan yang terkontaminasi. Penyebaran pertama kali terjadi di negara-negara non-endemik, seperti Amerika latin yaitu ekuador dan brazil yang mana seorang pasien wanita berumur 48 dan 46 tahun yang telah melakukan perjalanan internasional aktif dari australia. Proses penyebaran virus ini melalui mobilitas manusia dan interaksi sosial yang intens.
- IIB B.1.3 : *Sub-clade* IIB B.1.3 pertama kali terdeteksi pada April 2023 di Republik Korea. Kasus pertama melibatkan individu yang terinfeksi *monkeypox* setelah perjalanan internasional. Analisis genomik terhadap isolat virus dari pasien menunjukkan bahwa virus yang menyebabkan infeksi tersebut termasuk dalam garis keturunan B.1.3 dari *clade* IIB B.1. Meskipun *sub-clade* ini terdeteksi pada tahun 2023, ia berasal dari wabah global *monkeypox* yang dimulai pada tahun 2022 dan menyebar ke berbagai negara non-endemik. Proses penyebaran subklade B.1.3 mirip dengan subklade lainnya, yaitu melalui kontak langsung dengan lesi atau cairan tubuh dari individu yang terinfeksi, melalui droplet besar yang dikeluarkan ketika seseorang yang terinfeksi batuk atau bersin dan kontak dengan permukaan atau objek yang terkontaminasi dengan virus, yang kemudian

menyentuh bagian tubuh yang terluka atau tergores. Penyebaran ini terjadi dengan cepat, terutama di negara-negara dengan mobilitas internasional yang tinggi, yang memungkinkan virus untuk menyebar lebih luas ke negara-negara non-endemik. Proses ini didorong oleh interaksi sosial yang intens, yang meningkatkan kemungkinan kontak antara individu yang terinfeksi dan yang belum terinfeksi.

- IIB B.1.6 dan IIB B.1.7 : *Sub-clade* IIB B.1.6 dan IIB B.1.7 berasal dari *Clade* IIB B.1 dan terdeteksi selama wabah global *monkeypox* pada tahun 2022, dengan penyebaran yang sebagian besar terkait dengan perjalanan internasional. Kasus pertama dari B.1.6 dan B.1.7 ditemukan di beberapa negara non-endemik, seperti Eropa dan Amerika Utara, setelah individu yang terinfeksi melakukan perjalanan internasional ke wilayah yang lebih luas. Penyebaran kedua sub-clade ini terjadi melalui kontak langsung dengan lesi atau cairan tubuh, droplet besar, dan permukaan yang terkontaminasi, dengan faktor utama penyebarannya adalah mobilitas internasional yang tinggi dan interaksi sosial yang intens di negara-negara non-endemik. Meskipun keduanya memiliki mutasi genetik yang membedakannya, belum ada bukti bahwa mutasi tersebut mempengaruhi virulensi atau transmisibilitas secara signifikan. Faktor-faktor seperti kurangnya tindakan pencegahan atau kesadaran akan penularan dapat mempercepat penyebarannya. Pencegahan dilakukan melalui vaksinasi *smallpox*, isolasi pasien, dan kebersihan pribadi untuk mengurangi penyebaran.

BAB VI

KESIMPULAN

6.1 Kesimpulan

1. Persentase pasien *monkeypox* berdasarkan gender: laki-laki 51,46%, perempuan 9,71%, dan unknown 38,83%. Kesimpulannya, kategori laki-laki memiliki persentase tertinggi. Selain itu, rata rata penyebaran MPXV terbanyak berdasarkan persebaran peta ada di lokasi endemik yaitu benua afrika, lalu disusul oleh eropa yang mana hal ini terjadi dikarenakan tingkat mobility yang tinggi dan geografis yang sama pada satu wilayah tertentu. Adapun rata-rata umur pasien dibagi menjadi 7 kategori, dengan kategori *Unknown* mencatatkan korban terbanyak (42 sampel). Kategori usia lainnya adalah 0-9 tahun (2), 10-19 tahun (2), 20-29 tahun (26), 30-39 tahun (20), 40-49 tahun (1), dan > 49 tahun (1). Penyebaran monkeypox terjadi di beberapa negara, dengan data korban terbagi dalam kategori riwayat pasien: *Unknown, Hospitalized, Live, Released, dan Recovering*. Yang mana masing- masing di kategorikan berikut :
 - a. *Unknown*: 10 wanita, 33 laki-laki, 39 kasus dengan riwayat dan jenis kelamin tidak diketahui.
 - b. *Hospitalized*: 7 wanita.
 - c. *Live*: 6 pria, 1 unknown.
 - d. *Released*: 5 pria.
 - e. *Recovering*: 2 pria.
2. Dari studi kasus, terdapat 3 clade, yaitu: I , II dan IIb. Untuk clade I terdapat (3 sampel) yang terjadi di negara Sierra Lion, Kenya dan Kongo, hal ini terjadi karena penularan Clade I ke berbagai Afrika Tengah biasanya terjadi melalui kasus impor atau penularan lintas negara. Sedangkan untuk clade II (0 sampel) dan IIb (6 sampel) yang terjadi di negara uganda, swedia, kongo dan cameroon, hal tersebut bisa terjadi dikarenakan letak geografis yang sama , adapun untuk yang di luar africa seperti swedia bisa terjadi dikarena kasus impor dari africa. Selain itu, terdapat 24 jenis garis keturunan (lineages), yaitu: dari clade I terdapat sub-clade Ia (5 sampel) yang terjadi di kongo, gabon dan sudan. Adapun untuk jenis garis keturunan *Subclade*

dari clade II/Iib adalah Iib B.1 (25 sampel), Iib C.1 (7 Sampel), Iib 1.B.20 (6 sampel yang rata-rata terdapat di benua amerika), IIB C.1.1 (8 sampel) terjadi di amerika latin dan europa tengah, Iib A (4 sampel yang terjadi di africa dan impor ke europa, Iib A.2 (5 sampel yang terjadi di asia dan africa), B.1.6 (4 sampel yang terjadi di Asia, europa dan amerika), B.1.7 (4 sampel yang terjadi di africa selatan dan asia) , Iib B.1.2 (3 sampel yang terjadi di Asia, Amerika dan Europa, Iib C.1 (10 sampel yang rata-rata terjadi di asia namun ada beberapa kasus impor asia ke eropa), Iib A.1 (1 sampel yang terjadi di singapore yang melakukan perjalanan travel ke nigeria, Iib A.2.1 (1 sampel yang terjadi di pakistan) , Iib B.1.1 (1 sampel terjadi di germany), Iib B.1.10 (2 sampel terjadi di Europa), Iib B.1.17 (1 sampel terjadi di poland), Iib B.1.3 (2 sampel terdapat di benua Australia dan Europa, Iib A.2.2 (1 sampel di Nigeria), Iib B.1.22 (2 sampel terjadi di amerika selatan), Iib B.1.1 (5 sampel terjadi di Europa), Iib B.1.10 (2 sampel di Europa) , Iib B.1.11 (1 sampel terjadi di Argentina), Iib B.1.20 (7 sampel pada umumnya terjadi di Amerika selatan namun ada beberapa impor ke Australia hal ini terjadi karena perjalanan travel dan mobilitas), Iib B.1.3 (2 sampel terjadi di Europa dan Australia), Iib B.1.6 (4 sampel terjadi di benua Europa dan Africa) dan Iib B.1.7 (4 sampel terjadi di 3 benua Asia, Europa dan Amerika). Beberapa sampel merupakan kasus impor dan kemungkinan kasus lokal serta terdapat beberapa sampel dengan riwayat perjalanan ke negara sebelumnya yang memiliki garis keturunan yang sama dengan sampel dari negara tersebut. *Clade* Iib digunakan sebagai referensi untuk *clade* lain, yang memang dari analisis yang dilakukan terdapat banyak *sub-clade* yang tercipta. Untuk melihat perkembangan *Monkeypox virus* (MPXV) di Indonesia, terdapat beberapa genom dengan garis keturunan yang berasal dari Indonesia, yaitu garis keturunan Iib C.1. Garis keturunan ini muncul karena kombinasi dari perjalanan internasional, mutasi adaptif pada virus, kurangnya deteksi awal, dan kondisi sosial yang mendukung penularan. Penanganan cepat seperti peningkatan deteksi, isolasi, dan vaksinasi adalah langkah penting untuk mencegah penyebaran lebih lanjut.

6.2 Saran

1. Metode *maximum likelihood* membutuhkan proses komputasi yang berat dan memakan waktu lebih lama, namun hasil lebih optimal dan akurat dan sangat cocok untuk mencari model evolusi dengan data set besar. Dibandingkan metode berbasis jarak seperti *Neighbor-Joining* dan *parsimony* yang tidak mempertimbangkan hasil evolusi. Jika menggunakan banyak sampel penelitian ini dapat dikembangkan lebih lanjut menggunakan metode canggih, seperti UGMA yang selaras bisa digunakan di *software Mega-X* dengan *bootstrapping* yang lebih cepat. Selain itu, penelitian ini juga dapat dikembangkan sebagai perbandingan dengan beberapa metode dan model analisis filogenetika lainnya, baik berbasis jarak maupun berbasis karakter, selain itu untuk lebih dikembangkan bisa menggunakan metode *Bayesian Markov Chain Monte Carlo* (BMCMC) dengan perangkat lunak BEAST dengan pengambilan sampel yang lebih banyak.
2. Fokus pada *clade* spesifik Jika ditemukan kelompok yang menarik, lakukan analisis lebih mendalam pada kelompok tersebut. Hubungkan dengan ekologi atau morfologi: Gunakan hasil untuk menjelaskan pola adaptasi spesies atau hubungan antara struktur genetik dan fungsi ekologi. Bagi perusahaan farmasi, pembuat antivirus dapat memanfaatkan penelitian garis keturunan dari penelitian ini sebagai acuan referensi untuk membuat antivirus yang sesuai dengan garis keturunan tertentu.

DAFTAR PUSTAKA

- Aksamentov, I., Roemer, C., Hodcroft, E. & Neher, R. Nextclade: clade assignment, mutation calling and quality control for viral genomes. *J. Open Source Softw.* 6, 3773 (2021).
- Alexandros Stamatakis. 2014. RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. https://www.jcsantosresearch.org/Class_2014_Spring_Comparative/pdf/week_4/RAxML_v8_1_X.pdf?utm_source=chatgpt.com.
- Allman ES, Baños H, Rhodes JA (2019) NANUQ: a method for inferring species networks from gene trees under the coalescent model. *Algorithms Mol Biol* 14:1–25
- Amiroch Siti. 2020. *Filogenetika dari Perspektif Matematika*.
- Andrew Rambaut, Alexei J Drummond, Dong Xie, Guy Baele, Marc A Suchard. 2018. Posterior Summarization in Bayesian Phylogenetics Using Tracer 1.7.
- Casanellas, M., Homs Pons, R., & Torres, A. (2023). *A novel algebraic approach to time-reversible evolutionary models*. arXiv:2309.02276
- Darupamenang, A., et al. 2022. "Analisis Filogenetik *Alocasia Jacklyn* Berdasarkan Gen *rbcL*." *Jurnal Bios Logos*, 12(2): 157-163. (ejournal.unsrat.ac.id)
- Dennis A. Benson, Ilene Karsch-Mizrachi, David J. Lipman et al. 2015. GenBank. https://academic.oup.com/nar/article/37/suppl_1/D26/1007536?login=false
- Diep Thi Hoang., Le Sy Vinh., Alexandros Stamatakis. 2018. MPBoot: fast phylogenetic maximum parsimony tree inference and bootstrap approximation. https://bmcecolvol.biomedcentral.com/articles/10.1186/s12862-018-1131-3?utm_source=chatgpt.com.
- Epidemiological update: Monkeypox due to monkeypox virus clade I. 2024. Accessed October 10, 2024. <https://www.ecdc.europa.eu/en/news-events/epidemiologicalupdate-Monkeypox-due-monkeypox-virus-clade-i>.
- Fabio Scarpa., Ilenia Azzena., Alessandra Ciccozzi., et al. 2024 Update of the Genetic Variability of Monkeypox Virus Clade I Ib Lineage B.1. *Microorganisms* 2024, 12, 1874.

<https://doi.org/10.3390/microorganisms12091874>

<https://www.mdpi.com/journal/microorganisms>.

- Felsenstein J (2004) *Inferring phylogenies*. Sinauer Associates, New York
- Felsenstein. 2004. *Inferring phylogenies: an instant classic*.
<https://www.nature.com/hdy/>.
- Fischer M, Kelk S (2016) On the maximum parsimony distance between phylogenetic trees. *Ann Comb* 20:87–113
- Forster et al. 2020. Analisis jaringan filogenetika MonkeyPox genom.
- GISAID Initiative. GISAID. Retrieved June 30, 2022, from <http://www.gisaid.org/>
- Maldives Immigration Statistics. Maldives Immigration. (2021). Retrieved April 3, 2022, from <https://immigration.gov.mv/statistics/>.
- Gunadi, Hendra Wibawa., Marcellus., et al. 2020. Full-length genome characterization and phylogenetic analysis of SARS-CoV-2 virus strains from Yogyakarta and Central Java, Indonesia.
- Gunadi, Wibawa, H., Marcellus, Hakim, M. S., Daniwijaya, E. W., Rizki, L. P., Supriyati, E., Nugrahaningsih, D., Afiahayati, Siswanto, Iskandar, K., Anggorowati, N., Kalim, A. S., Puspitarani, D. A., Athollah, K., Arguni, E., Nuryastuti, T., & Wibawa, T. (2020). Full-length genome characterization and phylogenetic analysis of SARS-CoV-2 virus strains from Yogyakarta and Central Java, Indonesia. *PeerJ*, 8, e10575. <https://doi.org/10.7717/peerj.10575>
- Happi, C.; Adetifa, I.; Mbala, P.; Njouom, R.; Nakoune, E.; Happi, A.; Ndodo, N.; Ayansola, O.; Mboowa, G.; Bedford, T.; et al. Urgent need for a non-discriminatory and non-stigmatizing nomenclature for monkeypox virus. *PLoS Biol.* 2022, 23, 20–80
- Huerta-Cepas, J., Serra, F. and Bork, P. (2016) ETE 3: reconstruction, analysis, and visualization of phylogenomic data. *Mol. Biol. Evol.*, 33, 1635–1638
- Hoang, D. T., Chernomor, O., von Haeseler, A., Minh, B. Q., & Vinh, L. S. (2018). *UFBoot2: Improving the ultrafast bootstrap approximation*. *Molecular Biology and Evolution*, 35(2), 518–522. <https://doi.org/10.1093/molbev/msx281>
- Isidro et al. (2022). Phylogenomic characterization and signs of microevolution in the 2022 multi-country outbreak of monkeypox virus.

https://www.nature.com/articles/s41591-022-01907-y.pdf?utm_source=chatgpt.com.

Ivica Letunic, Peer Bork, Interactive Tree Of Life (iTOL) v5: an online tool for phylogenetic tree display and annotation, *Nucleic Acids Research*, Volume 49, Issue W1, 2 July 2021, Pages W293– W296, <https://doi.org/10.1093/nar/gkab301>

Ivica Letunic, Peer Bork. 2021. Interactive Tree Of Life (iTOL) v5: an online tool for phylogenetic tree display and annotation Vol. 49

Ivica Letunic., Peer Bork. 2021. Interactive Tree Of Life (iTOL) v5: an online tool for phylogenetic tree display and annotation. *Nucleic Acids Research*, Volume 49, Issue W1, 2 July 2021, Pages W293– W296, <https://doi.org/10.1093/nar/gkab301>.

Julie D. Thompson, Desmond G. Higgins, Toby J. Gibson.1994. CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. <https://doi.org/10.1093/nar/22.22.4673>.

Katoh, K., Kuma, K., Toh, H., & Miyata, T. (2005). MAFFT version 5: improvement in accuracy of multiple sequence alignment. *Nucleic acids research*, 33(2), 511–518. <https://doi.org/10.1093/nar/gki198>

Kemenkes RI. 2021. MONKEYPOX Frequently Asked Questions (FAQ) <https://infeksiemerging.kemkes.go.id/document/download/wQz>.

Koichiro Tamura, Daniel Peterson, Nicholas Peterson, Glen Stecher, Masatoshi Nei, Sudhir Kumar. 2011. MEGA5: Molecular Evolutionary Genetics Analysis Using Maximum Likelihood, Evolutionary Distance, and Maximum Parsimony Methods.

Kozlov, M. Monkeypox Declared a Global Emergency: Will It Help Contain the Outbreak? Available online: <https://www.nature.com/articles/d41586-022-02054-7> (accessed on 25 July 2022).

Kreft,L., Botzki,A., Coppens,F., Vandepoele,K. and Van Bel,M. (2017) PhyD3: a phylogenetic tree viewer with extended phyloXML support for functional genomics data visualization. *Bioinformatics*, 33, 2946–2947.

- L. L. Cavalli-Sforza, A. W. F. Edwards. 1967. Phylogenetics Analysis: Models And Estimation procedures. *Evolution*, Volume 21, Issue 3, 1 September 1967, Pages 550–570, <https://doi.org/10.1111/j.1558-5646.1967.tb03411.x>.
- Letunic,I. and Bork,P. (2019) Interactive Tree Of Life (iTOL) v4: recent updates and new developments. *Nucleic Acids Res.*, 47, W256–W259.
- Liu et al. 2021. Analisis Filogenetik Virus Flu Burung H5N1 Menggunakan Metode Maksimum likelihood.
- Lumongga Yustisia,. 2022. The Implementation of Computational Statistics in Moleculer Phylogenetics to Identify Lineage Diversity Through Viral Genome Sequences.
- Luo, Y.H.; Zhang, T.; Cao, J.L.; Hou, W.S.; Wang, A.Q.; Jin, C.H. Monkeypox: An outbreak of a rare viral disease. *J. Microbiol. Immunol. Infect.* 2024, 57, 1–10.
- Mande, G. et al. Enhanced surveillance of monkeypox in Bas-Uélé, Democratic Republic of Congo: the limitations of symptom-based case definitions. *Int. J. Infect. Dis.* 122, 647–655 (2022).
- Minchin, S., & Lodge, J. (2019). Understanding biochemistry: structure and function of nucleic acids. *Essays in biochemistry*, 63(4), 433–456. <https://doi.org/10.1042/EBC20180038>.
- Minh, B. Q. et al. IQ-TREE 2: new models and efficient methods for phylogenetic inference in the genomic era. *Mol. Biol. Evol.* 37, 1530–1534 (2020).
- Minh, B. Q. et al. IQ-TREE 2: New models and efficient methods for phylogenetic inference in the genomic era. *Mol. Biol. Evol.* 37, 1530–1534 (2020).
- Moritz Dirnberger , Pablo Peláez-Campomanes , Raquel López-Antoñanzas. 2024. Phylogenetic relationships of Neogene hamsters (Mammalia, Rodentia, Cricetinae) revealed under Bayesian inference and maximum parsimony.
- Parham JF, Irmis RB. 2008. Caveats on the use of fossil calibrations for molecular dating: a comment on near et al. *The American Naturalist* 171(1):132–136 DOI 10.1086/524198.
- Patil A, Kurlle S. Monkeypox virus – an evolutionary perspective from India. *J Infect* 2024;89(6):106323.

- Rambaut, A., Lam, T. T., Max Carvalho, L. & Pybus, O. G. Exploring the temporal structure of heterochronous sequences using TempEst (formerly Path-O-Gen). *Virus Evol.* 2, vew007 (2016).
- Robert Lanfear, Paul B. Frandsen, April M. Wright. 2016. PartitionFinder 2: New Methods for Selecting Partitioned Models of Evolution for Molecular and Morphological Phylogenetic Analyses.
- Shu, Y., & McCauley, J. (2017). GISAID: Global initiative on sharing all influenza data - from vision to reality. *Euro surveillance : bulletin European sur les maladies transmissibles = European communicable disease bulletin*, 22(13), 30494. <https://doi.org/10.2807/1560-7917.ES.2017.22.13.30494>
- Sinitsa MV, Delinschi A. 2016. The earliest member of Neocricetodon (Rodentia: Cricetidae): a redescription of *N. moldavicus* from Eastern Europe, and its bearing on the evolution of the genus. *Journal of Paleontology* 90(4):771–784 DOI 10.1017/jpa.2016.72.
- Smith MR. 2019. Bayesian and parsimony approaches reconstruct informative trees from simulated morphological datasets. *Biology Letters* 15(2):20180632 DOI 10.1098/rsbl.2018.0632.
- Smith MR. 2022. Using information theory to detect rogue taxa and improve consensus trees. *Systematic Biology* 71(5):1088–1094 DOI 10.1093/sysbio/syab099.
- Smith MR. 2023. Rogue: Identify Rogue Taxa in Sets of Phylogenetic Trees. Available at <https://zenodo.org/doi/10.5281/zenodo.5037327> (accessed 19 February 2024).
- Strumfels, 2015. Sequence Alignment. www.bio.cambridge.uk.
- Subramanian, B., Gao, S., Lercher, M.J., Hu, S. and Chen, W.H. (2019) Evolview v3: a webserver for visualization, annotation, and management of phylogenetic trees. *Nucleic Acids Res.*, 47, W270–W275.
- Turista, Dora & Islamy, Aesthetica & Kharisma, Viol & Ansori, Arif. (2020). Distribution of COVID-19 and Phylogenetic Tree Construction of SARS-2019-CoV-2 in Indonesia. *Journal of Pure and Applied Microbiology*. 14. 1035- 1042. 10.22207/JPAM.14.SPL1.42. \

- Ulfah, M., & Helianti, I. (2021). Bioinformatic analysis of the whole genome sequences of SARS-CoV-2 from Indonesia. *Iranian journal of microbiology*, 13(2), 145–155. <https://doi.org/10.18502/ijm.v13i2.5973>
- Vakaniaki EH, Kacita C, Kinganda-Lusamaki E, et al. Sustained human outbreak of a new MPXV clade I lineage in eastern Democratic Republic of the Congo. *Nat Med* 2024;30(10):2791–5.
- WHO Monkeypox, 2024. Accessed October 10, 2024. (<https://www.who.int/newsroom/fact-sheets/detail/Monkeypox>).
- World Health Organization. “2022-23 Monkeypox (Monkeypox) Outbreak: Global Trends”. Available online: https://worldhealthorg.shinyapps.io/mpx_global/ (accessed on 23 February 2024).
- World Health Organization. Who coronavirus (COVID-19) dashboard. World Health Organization. Retrieved April 1, 2022, from <https://covid19.who.int/>
- Xingpei Ji, Ruyue Liang Chao Bao et.al. 2024. Evolutionary variation of the monkeypox virus detected for the first time in Nantong, Jiangsu
- Xu, S. et al. Ggtree: a serialized data object for visualization of a phylogenetic tree and annotation data. *iMeta* 1, e56 (2022).
- Yu,G. (2020) Using ggtree to visualize data on tree-like structures. *Curr. Protoc. Bioinform.*, 69, e96.
- Yuelong Shu1 , John McCauley. 2017. GISAID: Global initiative on sharing all influenza data – from vision to reality. https://www.gisaid.org/fileadmin/c/gisaid/files/pdfs/Eurosurveillance_From_Vision-to-Reality.pdf?utm_source=chatgpt.com.
- Zahmatyar, M.; Fazlollahi, A.; Motamedi, A.; Zolfi, M.; Seyedi, F.; Nejadghaderi, S.A.; Sullman, M.J.M.; Mohammadinasab, R.; Kolahi, A.-A.; Arshi, S.; et al. Human monkeypox: History, presentations, transmission, epidemiology, diagnosis, treatment, and prevention. *Front. Med.* 2023, 10, 1157670.
- Zhang C, Huelsenbeck JP, Ronquist F (2020) Using parsimony-guided tree proposals to accelerate convergence in Bayesian phylogenetic inference. *Syst Biol* 69(5):1016–1032
- Zhang, S.-Y., & Liu, S.-L. (2013). *Bioinformatics. Brenner's Encyclopedia of Genetics*, 338–340. <https://doi.org/10.1016/b978-0-12-374984-0.00155-8>

- Zhang, S.-Y., & Liu, S.-L. (2013). Bioinformatics. *Brenner's Encyclopedia of Genetics*, 338–340. <https://doi.org/10.1016/b978-0-12-374984-0.00155-8>.
- Zhou, P., Yang, XL., Wang, XG. et al. A pneumonia outbreak associated with a new coronavirus of probable bat origin. *Nature* 579, 270–273 (2020). <https://doi.org/10.1038/s41586-020-2012-7>

LAMPIRAN

Lampiran 1 Data pasien

Accession ID	Clade / Lineage	Collection date	Negara	Host	Gender	Age	Patient status
EPI_ISL_13058398	I	1970	Sierra Lion	Human	Unknown	Unknown	Unknown
EPI_ISL_13251157	IIb C.1	6/1/2022	Swis	Human	Male	35	Unknown
EPI_ISL_13331715	IIb B.1	2022	Finland	Human	Unknown	Unknown	Unknown
EPI_ISL_13374487	IIb B.1	6/1/2022	Hungary	Human	Male	24	Unknown
EPI_ISL_13472080	IIb B.1.17	6/6/2022	Poland	Human	Male	40	Hospitalized
EPI_ISL_13744902	IIb B.1.10	2022-06-16	Finland	Human	Unknown	Unknown	Unknown
EPI_ISL_13827283	IIb B.1.3	4/28/2022	Czech Republic	Human	Unknown	Unknown	Unknown
EPI_ISL_13891126	IIb A	2022-06-26	Turkey	Human	Unknown	Unknown	Unknown
EPI_ISL_14153982	IIb B.1	2022-07-27	Thailand	Human	Male	47	Hospitalized
EPI_ISL_14295679	IIb A.2	2022-08-02	Thailand	Human	Male	25	Hospitalized
EPI_ISL_14721259	IIb B.1.2	7/13/2022	Singapore	Human	Male	48	Unknown
EPI_ISL_14945299	IIb B.1.7	2022-09-06	hongkong	Human	Unknown	Unknown	Unknown
EPI_ISL_15158314	Ia	1988	Gabon	Human	Female	9 months	Unknown
EPI_ISL_15158418	IIb B.1.6	2022-08	United Kingdom	Human	Unknown	Unknown	Unknown
EPI_ISL_15370050	IIb B.1	2022-07	Slovenia	Human	Unknown	Unknown	Unknown
EPI_ISL_15370058	IIb B.1.2	2022-08	Slovenia	Human	Unknown	Unknown	Unknown
EPI_ISL_15370059	IIb B.1	2022-08	Slovenia	Human	Unknown	Unknown	Unknown
EPI_ISL_15972407	IIb B.1.11	9/27/2022	Argentine	Human	Male	38	Unknown
EPI_ISL_16116755	IIb B.1.6	2022-10-05	Peru	Human	Male	31	Live
EPI_ISL_16233781	IIb B.1	2022-08-19	Spain	Human	Male	27	Live
EPI_ISL_17012036	IIb B.1.20	10/17/2022	Ecuador	Human	Male	48	Unknown
EPI_ISL_17012040	IIb B.1.20	10/20/2022	Ecuador	Human	Male	30	Unknown
EPI_ISL_17012068	IIb B.1	11/1/2022	Ecuador	Human	Male	34	Unknown
EPI_ISL_17012102	IIb B.1.20	9/30/2022	Ecuador	Human	Male	29	Unknown
EPI_ISL_17201444	IIb B.1	7/5/2022	Spain	Human	Unknown	Unknown	Unknown
EPI_ISL_17793226	IIb B.1	7/27/2022	Belgium	Human	Unknown	Unknown	Unknown
EPI_ISL_18257122	IIb C.1.1	2023-09-05	China	Human	Male	37	Released
EPI_ISL_18744050	IIb C.1	2023-05-03	Korea	Human	Unknown	Unknown	Live
EPI_ISL_18798838	IIb C.1	2023-11-21	Indonesia	Human	Male	24	Unknown
EPI_ISL_18846302	IIb B.1	2023-05	Italy	Human	Unknown	Unknown	Unknown
EPI_ISL_18879930	IIb B.1	2022-12-08	Egypt	Human	Male	39	Live
EPI_ISL_19093805	I	12/30/2023	Kongo	Human	Male	11	Unknown

Accession ID	Clade / Lineage	Collection date	Negara	Host	Gender	Age	Patient status
EPI_ISL_19255496	IIb B.1.3	2023-10-27	Papua Newguinea	Human	Unknown	Unknown	Unknown
EPI_ISL_19256216	IIb A.2	2022-09-20	Nigeria	Human	Unknown	Unknown	Unknown
EPI_ISL_19256223	IIb A.2	10/26/2022	Nigeria	Human	Unknown	Unknown	Unknown
EPI_ISL_19287108	Ia	2023-01-24	Kongo	Human	Female	28	Unknown
EPI_ISL_19293742	IIb C.1	2024-03-03	Japan	Human	Unknown	Unknown	Unknown
EPI_ISL_19331295	IIb C.1	5/2/2024	Vietnam	Human	Male	29	Released
EPI_ISL_19331300	IIb C.1	2024-06-10	Vietnam	Human	Male	27	Released
EPI_ISL_19345018	Ia	3/2/2024	Kongo	Human	Female	23	Unknown
EPI_ISL_19345028	IIb A	2/15/2024	Kongo	Human	Female	11	Unknown
EPI_ISL_19345318	IIb B.1.2	9/20/2022	Chile	Human	Male	35	Unknown
EPI_ISL_19345366	IIb B.1.22	2023-01-17	Chile	Human	Male	28	Unknown
EPI_ISL_19345373	IIb B.1	2022-11-04	Chile	Human	Male	61	Unknown
EPI_ISL_19348512	Ib	2024-08-13	Sweden	Human	Unknown	Unknown	Unknown
EPI_ISL_19357138	Ib	1/2/2024	Kongo	Human	Unknown	Unknown	Unknown
EPI_ISL_19359676	IIb C.1.1	2024-02-22	Brazil	Human	Male	25	Unknown
EPI_ISL_19359678	IIb C.1.1	2024-06-04	Brazil	Human	Male	26	Unknown
EPI_ISL_19359695	IIb B.1.20	2024-08-14	Brazil	Human	Male	46	Unknown
EPI_ISL_19359697	IIb B.1.20	2024-04-14	Brazil	Human	Male	31	Unknown
EPI_ISL_13052292	IIb B.1	2022-05	Aussie	Human	Male	Unknown	Unknown
EPI_ISL_13056274	IIb A.1	2019	Singapore	Human	Male	30	Released
EPI_ISL_13058405	Iia	1970	Liberia	Human	Unknown	Unknown	Unknown
EPI_ISL_13058459	Ia	2005	Sudan	Human	Unknown	Unknown	Unknown
EPI_ISL_13308160	IIb B.1	5/20/2022	France	Human	Unknown	Unknown	Unknown
EPI_ISL_13331598	IIb B.1	6/4/2022	Georgia	Human	Male	29	Hospitalized
EPI_ISL_13498265	IIb B.1.7	2022-06-20	South Africa	Human	Male	30	Recovering
EPI_ISL_13573943	IIb B.1.1	6/21/2022	Austria	Human	Male	44	Unknown
EPI_ISL_13632288	IIb B.1.7	2022-06-24	South Africa	Human	Male	32	Recovering
EPI_ISL_14153982	IIb B.1	7/27/2022	Thailand	Human	Male	47	Hospitalized
EPI_ISL_14621525	IIb B.1	6/14/2022	France	Canis lupus familiaris	Unknown	Unknown	Unknown

Accession ID	Clade / Lineage	Collection date	Negara	Host	Gender	Age	Patient status
EPI_ISL_15022590	IIb A.2	2022	India	Human	Male	35	Unknown
EPI_ISL_15104903	IIb B.1.1	2022-09-15	Germany	Human	Female	35	Unknown
EPI_ISL_15158408	IIb B.1.7	2022-08	United Kingdom	Human	Unknown	Unknown	Unknown
EPI_ISL_15199776	IIb B.1	2022-08-01	Portugal	Human	Unknown	Unknown	Unknown
EPI_ISL_15247221	IIb B.1	2022-08-19	Austria	Human	Male	35	Unknown
EPI_ISL_15747900	IIb B.1.6	2022-09-07	Russia	Human	Male	28	Hospitalized
EPI_ISL_15972409	IIb B.1	10/30/2022	Argentine	Human	Male	42	Unknown
EPI_ISL_16116763	IIb B.1.6	2022-10-03	Peru	Human	Male	26	Live
EPI_ISL_17012023	IIb B.1.20	2022-10-11	Ecuador	Human	Male	21	Unknown
EPI_ISL_17201439	IIb B.1.10	7/1/2022	Spain	Human	Unknown	Unknown	Unknown
EPI_ISL_17518107	IIb C.1	2023-03-17	Japan	Human	Male	Unknown	Unknown
EPI_ISL_17736867	IIb C.1	2022-06-11	Germany	Human	Unknown	Unknown	Unknown
EPI_ISL_17736868	IIb B.1	7/25/2022	Germany	Human	Unknown	Unknown	Unknown
EPI_ISL_17797714	IIb B.1	8/16/2022	Belgium	Human	Unknown	Unknown	Unknown
EPI_ISL_18059183	IIb C.1.1	2023-06-28	China	Human	Male	49	Released
EPI_ISL_18128768	IIb B.1	2022-06-03	France	Human	Male	35	Unknown
EPI_ISL_18467808	IIb C.1	2023-10-25	Indonesia	Human	Male	36	Unknown
EPI_ISL_18539019	IIb B.1	2022	Asussie	Human	Unknown	Unknown	Unknown
EPI_ISL_18627266	IIb C.1.1	2023-11-22	Germany	Human	Male	39	Live
EPI_ISL_18642362	IIb C.1	2023-11-10	Indonesia	Human	Male	24	Unknown
EPI_ISL_18659846	IIb C.1.1	2023-11-07	Netherland	Human	Unknown	Unknown	Unknown
EPI_ISL_18755977	IIb B.1	2022-06-28	Netherland	Human	Unknown	Unknown	Unknown
EPI_ISL_18846285	IIb B.1	2023-05	Italy	Human	Unknown	Unknown	Unknown
EPI_ISL_18993955	IIb C.1.1	2023-12-21	Portugal	Human	Unknown	Unknown	Unknown
EPI_ISL_18993969	IIb C.1.1	2023-11-30	Portugal	Human	Unknown	Unknown	Unknown
EPI_ISL_19131342	IIb A.2	10/31/2022	India	Human	Male	29	Unknown
EPI_ISL_19256201	IIb A.2.2	10/13/2022	Nigeria	Human	Unknown	Unknown	Unknown
EPI_ISL_19256302	IIb	5/10/2018	Cameroon	Human	Unknown	Unknown	Unknown
EPI_ISL_19287107	Ia	2023-01-24	Kongo	Human	Male	3	Unknown
EPI_ISL_19302262.1	I	7/25/2024	Kenya	Human	Male	45	Hospitalized
EPI_ISL_19305615	Ib	7/15/2024	Uganda	Human	Female	22	Unknown

Accession ID	Clade / Lineage	Collection date	Negara	Host	Gender	Age	Patient status
EPI_ISL_17012023	IIb B.1.20	2022-10-11	Ecuador	Human	Male	21	Unknown
EPI_ISL_17201439	IIb B.1.10	7/1/2022	Spain	Human	Unknown	Unknown	Unknown
EPI_ISL_17518107	IIb C.1	2023-03-17	Japan	Human	Male	Unknown	Unknown
EPI_ISL_17736867	IIb C.1	2022-06-11	Germany	Human	Unknown	Unknown	Unknown
EPI_ISL_17736868	IIb B.1	7/25/2022	Germany	Human	Unknown	Unknown	Unknown
EPI_ISL_17797714	IIb B.1	8/16/2022	Belgium	Human	Unknown	Unknown	Unknown
EPI_ISL_18059183	IIb C.1.1	2023-06-28	China	Human	Male	49	Released
EPI_ISL_18128768	IIb B.1	2022-06-03	France	Human	Male	35	Unknown
EPI_ISL_18467808	IIb C.1	2023-10-25	Indonesia	Human	Male	36	Unknown
EPI_ISL_18539019	IIb B.1	2022	Asussie	Human	Unknown	Unknown	Unknown
EPI_ISL_18627266	IIb C.1.1	2023-11-22	Germany	Human	Male	39	Live
EPI_ISL_18642362	IIb C.1	2023-11-10	Indonesia	Human	Male	24	Unknown
EPI_ISL_18659846	IIb C.1.1	2023-11-07	Netherland	Human	Unknown	Unknown	Unknown
EPI_ISL_18755977	IIb B.1	2022-06-28	Netherland	Human	Unknown	Unknown	Unknown
EPI_ISL_18846285	IIb B.1	2023-05	Italy	Human	Unknown	Unknown	Unknown
EPI_ISL_18993955	IIb C.1.1	2023-12-21	Portugal	Human	Unknown	Unknown	Unknown
EPI_ISL_18993969	IIb C.1.1	2023-11-30	Portugal	Human	Unknown	Unknown	Unknown
EPI_ISL_19131342	IIb A.2	10/31/2022	India	Human	Male	29	Unknown
EPI_ISL_19256201	IIb A.2.2	10/13/2022	Nigeria	Human	Unknown	Unknown	Unknown
EPI_ISL_19256302	IIb	5/10/2018	Cameroon	Human	Unknown	Unknown	Unknown
EPI_ISL_19287107	Ia	2023-01-24	Kongo	Human	Male	3	Unknown
EPI_ISL_19302262.1	I	7/25/2024	Kenya	Human	Male	45	Hospitalized
EPI_ISL_19305615	Ib	7/15/2024	Uganda	Human	Female	22	Unknown
EPI_ISL_19305615	Ib	7/15/2024	Uganda	Human	Female	22	Unknown
EPI_ISL_19344622	IIb A.2.1	2024-02-16	Pakistan	Human	Male	22	Live
EPI_ISL_19345023	IIb A	2/27/2024	Kongo	Human	Female	28	Unknown
EPI_ISL_19345027	IIb A	2/27/2024	Kongo	Human	Female	36	Unknown
EPI_ISL_19345361	IIb B.1.22	12/30/2022	Chile	Human	Male	28	Unknown
EPI_ISL_19345363	IIb B.1	1/10/2023	Chile	Human	Male	22	Unknown
EPI_ISL_19345367	IIb B.1	1/25/2023	Chile	Human	Male	34	Unknown
EPI_ISL_19359676	IIb C.1.1	2024-02-22	Brazil	Human	Male	25	Unknown
EPI_ISL_19363625	IIb B.1.20	8/13/2024	Australia	Human	Unknown	Unknown	Unknown
EPI_ISL_19365545	Ib	5/15/2024	Kongo	Human	Unknown	Unknown	Unknown

Lampiran 2 Data DNA MPXV (Monkeypox)

Lampiran 3 *Data Full Information DNA Patient By GISAID*

1	Accession ID	Clade / Lineage	Collection date	Benua	Negara	Provinsi/Kota	Host	Additional information	Gender	Age	Pat
2	EPI_ISL_19293742	IIb C.1	2024-03-03	Asia	Japan	Tokyo	Human	International Case	Unknown	Unknown	Un
3	EPI_ISL_14295679	IIb A.2	2022-08-02	Asia	Thailand	Phuket	Human	International Case	Male	25	Ho
4	EPI_ISL_15747900	IIb B.1.6	2022-09-07	Europe	Russia	Saint Petersburg	Human	Travelled from Spain (Bar	Male	28	Ho
5	EPI_ISL_13331598	IIb B.1	6/4/2022	Asia	Georgia	Tbilisi	Human	International Case	Male	29	Ho
6	EPI_ISL_13472080	IIb B.1.17	6/6/2022	Europe	Poland	Mazowieckie	Human	International Case	Male	40	Ho
7	EPI_ISL_19302262	IIb B.1	7/25/2024	Africa	Kenya	Taita Taveta	Human	Taveta	Male	45	Ho
8	EPI_ISL_14153982	IIb B.1	2022-07-27	Asia	Thailand	Bangkok	Human	International Case	Male	47	Ho
9	EPI_ISL_14153982	IIb B.1	7/27/2022	Asia	Thailand	Bangkok	Human	International Case	Male	47	Ho
10	EPI_ISL_19344622	IIb A.2.1	2024-02-16	Asia	Pakistan	Islamabad	Human	International Case	Male	22	Liv
11	EPI_ISL_16116763	IIb B.1.6	2022-10-03	South America	Peru	San Vicente	Human	International Case	Male	26	Liv
12	EPI_ISL_16233781	IIb B.1	2022-08-19	Europe	Spain	Galicia	Human	International Case	Male	27	Liv
13	EPI_ISL_16116755	IIb B.1.6	2022-10-05	South America	Peru	Chiclayo	Human	International Case	Male	31	Liv
14	EPI_ISL_18627266	IIb C.1.1	2023-11-22	Europe	Germany	Munich	Human	Travel history Cruise Ship	Male	39	Liv
15	EPI_ISL_18879930	IIb B.1	2022-12-08	Africa	Egypt	unknown	Human	International Case	Male	39	Liv
16	EPI_ISL_18744050	IIb C.1	2023-05-03	Asia	Korea	Seoul	Human	International Case	Unknown	Unknown	Liv
17	EPI_ISL_13056274	IIb A.1	2019	Asia	Singapore	Singapore	Human	Travel History: Nigeria	Male	30	Re
18	EPI_ISL_13498265	IIb B.1.7	2022-06-20	Africa	South Africa	Gauteng	Human	International Case	Male	30	Re
19	EPI_ISL_13632288	IIb B.1.7	2022-06-24	Africa	South Africa	Western Cape	Human	International Case	Male	32	Re
20	EPI_ISL_19331300	IIb C.1	2024-06-10	Asia	Vietnam	Dong Nai	Human	International Case	Male	27	Re