Edelweiss Applied Science and Technology
ISSN: 2576-8484
Vol. 8, No. 6, 9450-9460
2024
Publisher: Learning Gate
DOI: 10.55214/25768484.v8i6.4024

Constructing phylogenetic tree for viral hemorrhagic fever (VHF) by RAXML

Meaad Hussein Al-Abadi[1*], Bashar Talib Al-Nuaimi[2]
[1,2]Department of Computer Science, College of Science, University of Diyala, Iraq; scicomphd2210@uodiyala.edu.iq (M.H.A.A.).

Abstract: Various viruses can produce severe and frequently fatal infections known as viral hemorrhagic fevers (VHF). Comprehending the evolutionary connections among these viruses is essential for developing vaccines, treating illnesses, and conducting epidemiological surveillance. The evolutionary history and genetic relatedness of viruses that cause viral heart failure are better understood by phylogenetic research. Thus, this work aimed to undertake multiple sequence alignment and then use the Maximum Likelihood Estimate approach to produce a phylogenetic tree architecture of VHF. This study utilized 27 isolates of VHF (Ebola & Marburg) retrieved from the GenBank database (National Center for Biotechnology Information) for this investigation. The recent findings demonstrated that the Maximum Likelihood Estimate approach yielded a phylogenetic tree that was highly accurate and reliable for establishing the evolutionary relationships between VHF.
Keywords: *Maximum likelihood, Phylogenetic analysis, Phylogenetic tree, Viral hemorrhagic fever viruses.*

1. Introduction

Viral hemorrhagic fever continues to pose a concern to global public health. VHF persists today; according to the Iraqi Ministry of Health, the "hemorrhagic fever" has killed around 50 Iraqis and caused 273 injuries since the beginning of 2024 [1][2].

It is distinguished by fever, tiredness, and a propensity to bleed, frequently resulting in severe sickness and even death. The viruses that cause VHF are categorized into six families: (Arenaviridae, Hantaviridae, Nairoviridae, Phenuiviridae, Flaviviridae, and Filoviridae)[3]. The viruses that cause VHF are Lassa, Dengue, Marburg, and Ebola.

Marburg virus illness is a highly infectious that causes hemorrhagic fever [4]. It belongs to the same viral family as the Ebola virus [5]. Understanding the evolutionary relationships between these viruses is necessary to establish their origins, transmission methods, and potential for outbreaks [6]. Genetic sequence data, phylogenetic analysis, and a solid molecular biology approach may rebuild evolutionary relationships [7]. We advocate utilizing VHF in phylogenetic tree reconstruction to estimate and demonstrate the evolutionary relationships between those genomes [8]. The evolutionary analysis of viruses is a valuable tool in epidemiology. Phylogenetic trees may represent different viruses' evolutionary history and relatedness by comparing their genomic sequences [9].

By looking at genomic sequences, scientists may identify evolutionary relationships between viral isolates or strains and evaluate how the virus has changed over time by looking for mutations, genetic recombination events, and the emergence of new viral varieties[10][11]. It can reveal the mechanics of the virus's spread, such as how it spreads among different populations. Its inference is based on a maximum likelihood estimate due to its precision and efficiency. The maximum likelihood approach is one of the most used statistical estimating techniques. Applying the maximum probability strategy is more challenging and requires a deeper comprehension of the evolutionary models that these approaches are based on [12]. The maximum likelihood approach is limited to a small number of sequences due to its greater complexity, necessitating many computational steps that rise quickly with

* Correspondence: scicomphd2210@uodiyala.edu.iq

the number of sequences [13]. A supercomputer can be used to carry them out to assess numerous sequences at once [14].

To predict the best trees, illustrate the evolutionary relationships between the sequences, and create a well-supported phylogenetic tree using the maximum likelihood method, this paper thus attempts to investigate the possibility of finding relationships among the nucleotide sequences of VHF using the genes shared by these species.

For example, scientists can generate vaccines or antivirals that target multiple viral variants by identifying conserved regions of a virus' genome that need to be genetic gun sights linked with any virulence mean Temecula properties [15]. The confidence (corresponding to the support for individual clades in a consensus tree) may be inferred bootstrap values are placed on nodes [16].

2. Material and Methods

Description of Figure 1: A complete diagram detailing the many steps conducted (like a flow chart) shows how the maximum likelihood methods were used to build a phylogenetic VHF tree.

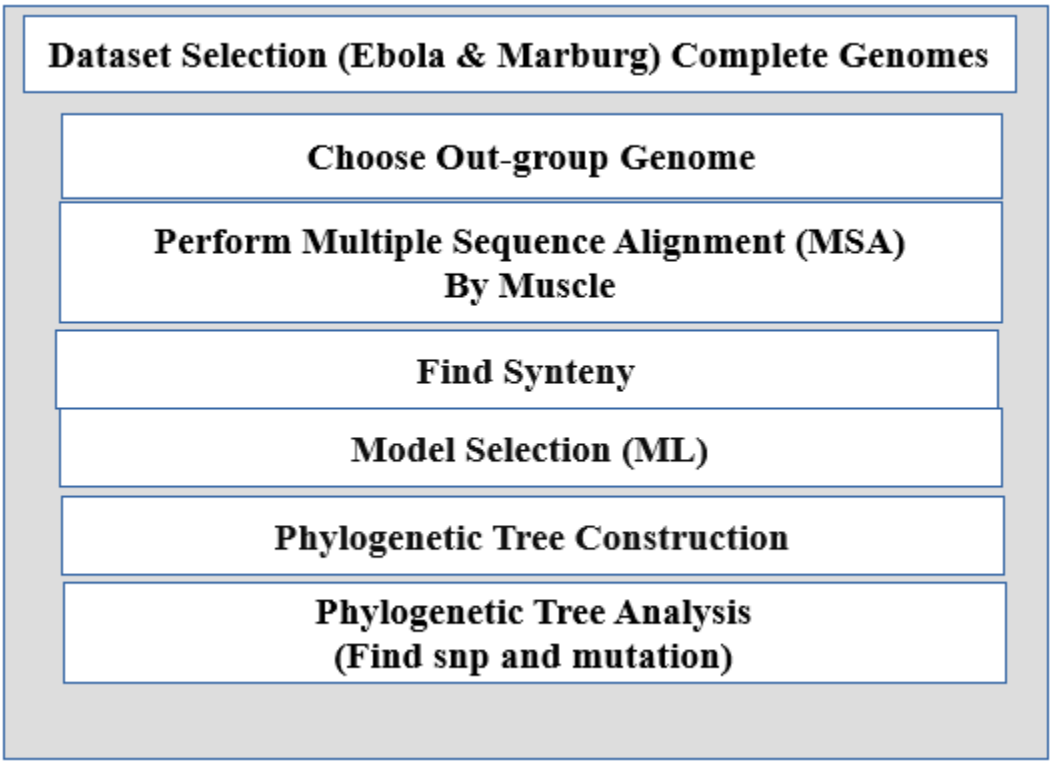

Figure 1.
General steps in phylogenetic tree construction.

2.1. Dataset Selection

To build a phylogenetic tree for VHF viruses using ML, we collected the genetic sequences of representative viruses (27) complete genomes, (10) Ebola, and (17) Marburg from public databases (NCBI) [17]. Some sequences covering the genomic regions of relevance for phylogenetic analysis were selected to ensure comprehensive coverage of the variation among the viral families that cause VHF [18]. Data about an organism's DNA or genetic makeup is known as genomic information. Listed Accession numbers, definitions, genome size in base pairs (bp), publication titles, host information, and place of origin are a few examples of the data that can be tabled in Table 1.

Edelweiss Applied Science and Technology
ISSN: 2576-8484
Vol. 8, No. 6: 9450-9460, 2024
DOI: 10.55214/25768484.v8i6.4024

2.1.1. Choose Out-group

Select a suitable out-group to establish the phylogenetic tree's root. Although it should be geographically far away to serve as an evolutionary benchmark, the out-group should share phylogenetically with the VHF viruses [19].

Table 1.
Information about some complete genomes (Ebola and Marburg) [19].

Accession	Organism name	Genome size(bp)	Host	Geolocation	Collection Data	Species
KY785985	Ebola virus	18871	Macaca fascicularis	**Gabon**	2001	Orthoebolavirus zairense
DQ217792	Marburg virus	19112	human| vertebrates	Kenya,	11-JAN-2024	Orthomarburgvirus marb
KU182901	Ebola virus	18959	Homo sapiens	Zaire	1995-05-04	Orthoebolavirus zairense
KT345616	Ebola virus	18794	Homo sapiens	Sierra Leone	2015-02-19	Orthoebolavirus zairense
KU143834	Ebola virus	18959	Homo sapiens	Sierra Leone	2014	Orthoebolavirus zairense
MK731986	Ebola virus	18912	Homo sapiens	Democratic Republic of the Congo	2018-08-16	Orthoebolavirus zairense
KR063670	Sudan ebolavirus	18875	Homo sapiens	Uganda	2000-10-01	Orthoebolavirus sudanense
KY471123	Zaire ebolavirus	18871	Macaca fascicularis	Gabon	2015	Orthoebolavirus zairense
MH121169	Sudan ebolavirus	18831	Homo sapiens	Sudan: Yambio	2004	Orthoebolavirus sudanense

2.2. Sequence Alignment

Perform multiple sequence alignments, including gap and ambiguity removal of the sequences for all VHF viruses, along with those selected from out-group viruses. Use MUSCLE (Multiple Sequence Comparison by Log-Expectation) on the combined set of sequences [20]. Like any analysis, phylogenetic analyses require all sequences to be in the proper reading frame. Multiple tools Genomics, or multiple sequence alignment, is an essential domain bioinformatics tool used in proteomics, evolutionary biology, and genomics, among other domains, that helps to perform numerous types of analyses. [21] The most relevant shape of the tree in biology is a phylogenetic tree a diagram that illustrates the evolutionary relations between organisms or genes, for example, and they are made by aligning sequences, so if you cannot create an MSA, then there is no phylogeny. It is an approach to identify genetic variants such as insertions and deletions (idles) or single nucleotide polymorphisms (SNPs) by comparing sequences from different individuals and populations.

Using multiple sequence alignment (MSA) helps us quickly identify similarities and differences in related sequences, which is essential for grasping how genes and proteins function and how they relate evolutionarily [22]. Figure 2 presents the MSA for complete Ebola genomes, while Figure 3 presents the MSA for Marburg genomes.

Edelweiss Applied Science and Technology
ISSN: 2576-8484
Vol. 8, No. 6: 9450-9460, 2024
DOI: 10.55214/25768484.v8i6.4024

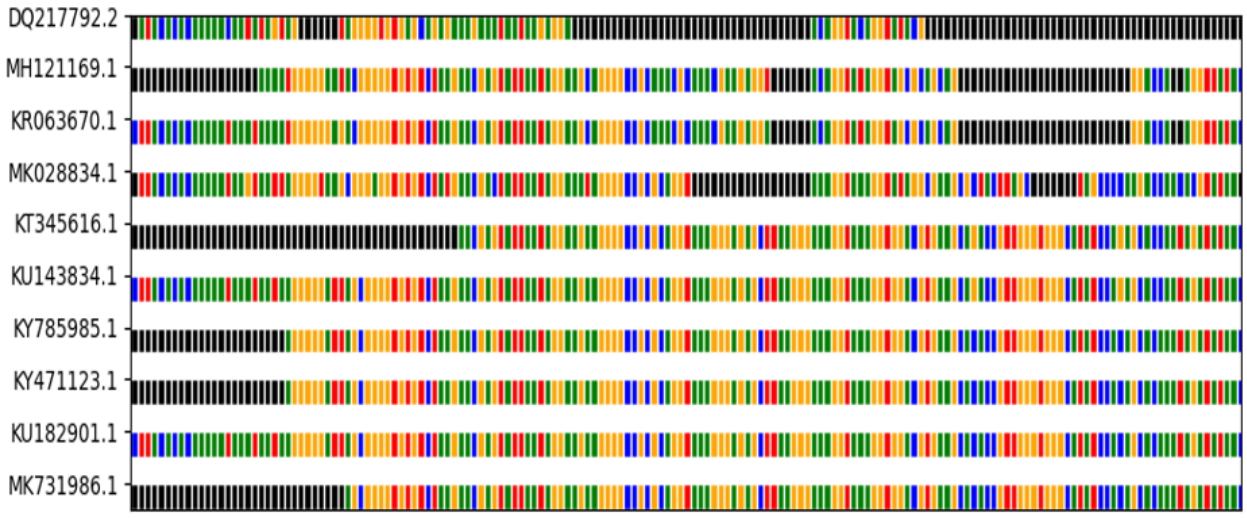

Figure 2.
Show alignment for complete genome ebola with the outgroup (DQ217792.2).

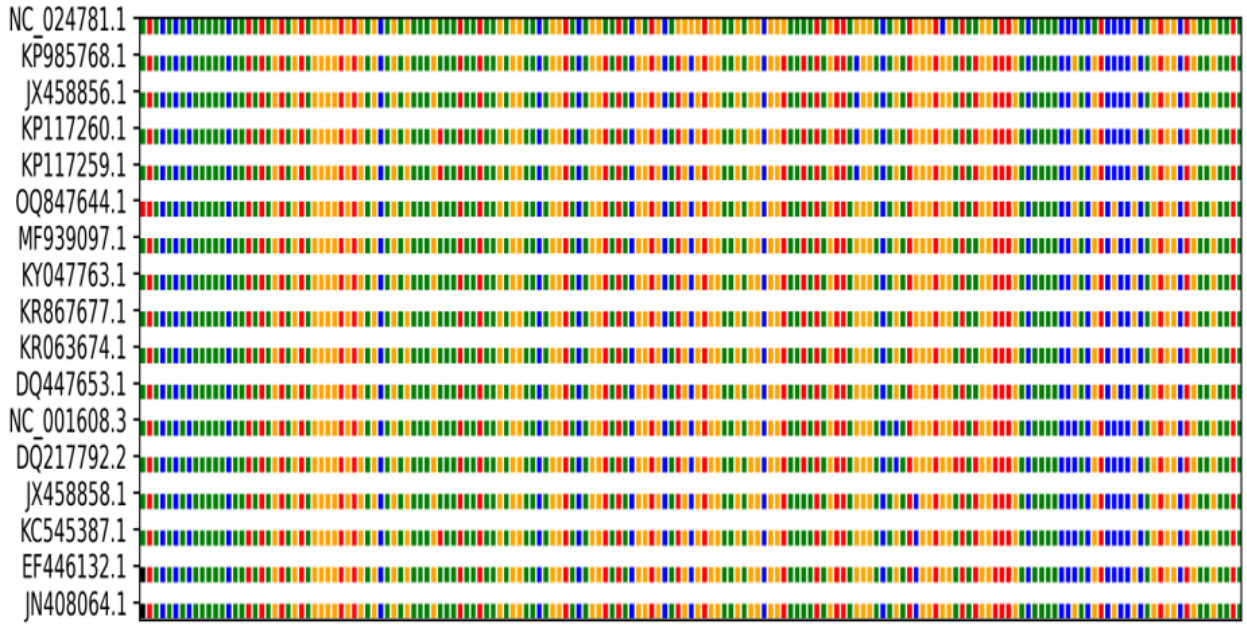

Figure 3.
Show alignment for complete genomes Marburg with the outgroup (NC_001608.3).

2.3. Find Synteny

Synteny analysis is a powerful method that offers insights into genomes' structure, function, and evolution, making it an essential part of genomics research. After aligning sequences, researchers can compare the genomic contexts of different species [23]. Typically, it involves looking for conserved gene arrangements, identifying syntenic blocks, and exploring chromosomal rearrangements or gene duplications [24].

While synteny analysis and MSA are valuable methods in genomics research, their applications and stages of study are usually distinct. While synteny analysis evaluates how genes or genomic areas are

Edelweiss Applied Science and Technology
ISSN: 2576-8484
Vol. 8, No. 6: 9450-9460, 2024
DOI: 10.55214/25768484.v8i6.4024

organized across species, MSA is used to compare sequences. Figures (4) and (5) show synteny for Ebola and Marburg.

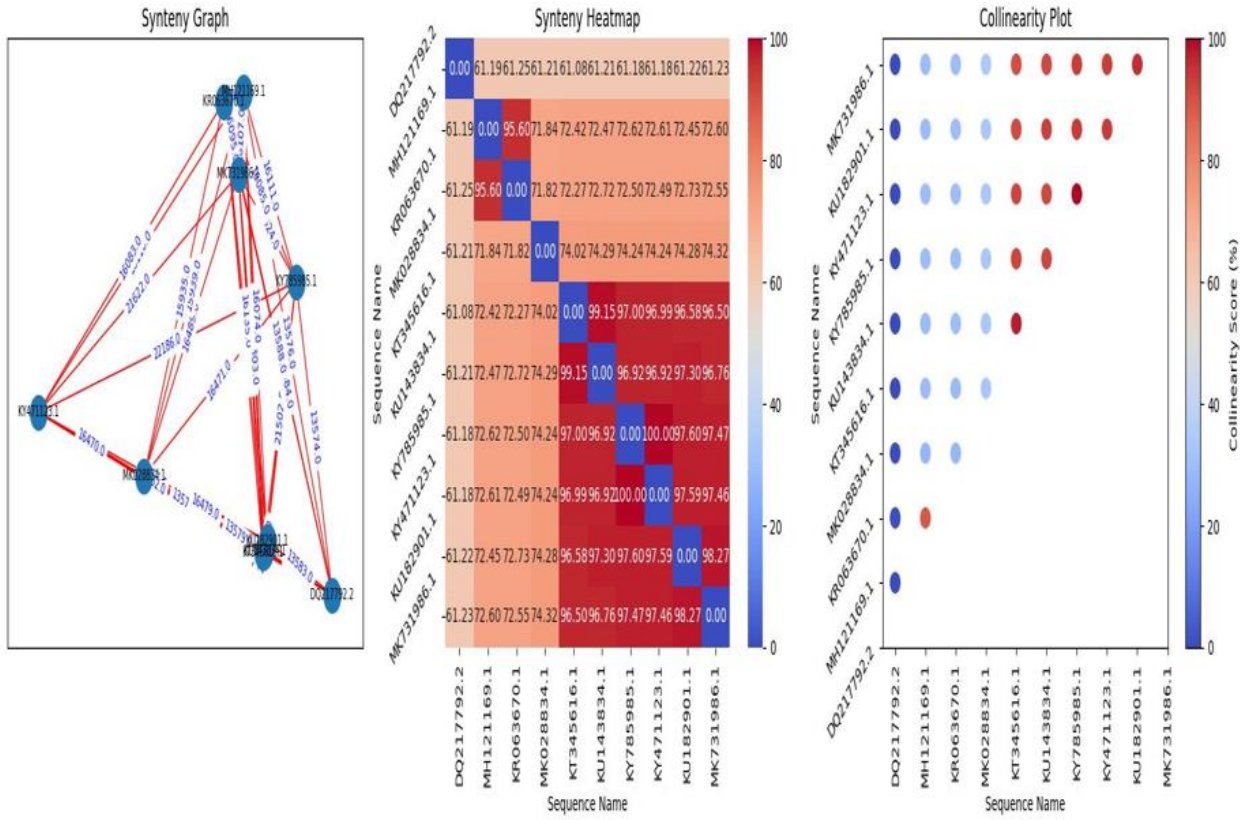

Figure 4.
Shows synteny for Ebola, where the dark blue color of the small circle shows similar genomes with average values, the light blue color shows good values, and the red color shows high values, meaning the similarity between the genomes is excellent.

Edelweiss Applied Science and Technology
ISSN: 2576-8484
Vol. 8, No. 6: 9450-9460, 2024
DOI: 10.55214/25768484.v8i6.4024

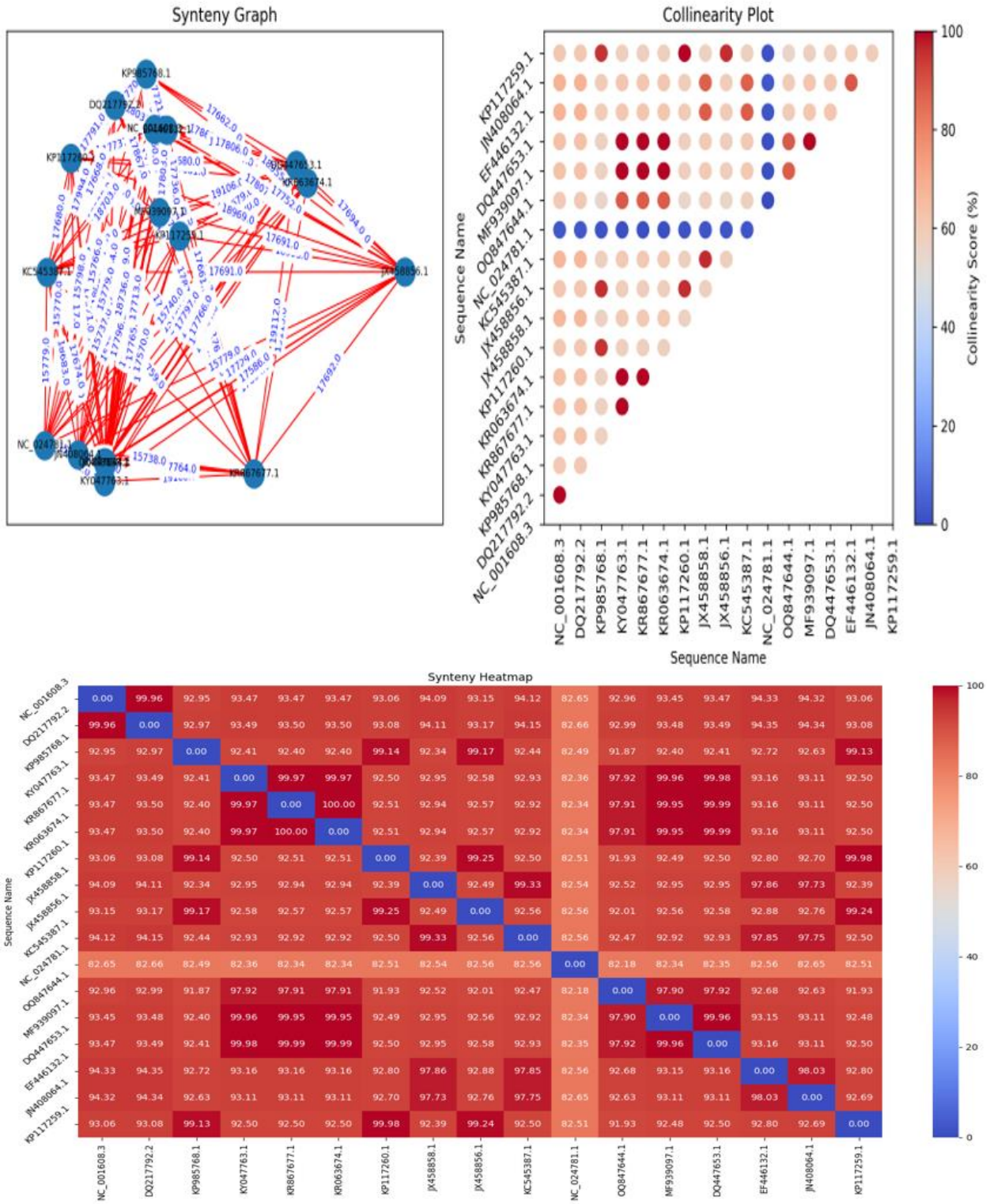

Figure 5.
Show synteny for Marburg, where the dark blue color of the small circle shows similar genomes with average values, the light blue color shows good values, and the red color shows high values, meaning the similarity between the genomes is excellent.

Edelweiss Applied Science and Technology
ISSN: 2576-8484
Vol. 8, No. 6: 9450-9460, 2024
DOI: 10.55214/25768484.v8i6.4024

2.3. Model Selection

This step entails deciding which evolutionary model best matches the available data. Using this model, a hypothesis regarding the organism's or gene family's evolutionary history will be produced. Choosing evolutionary tree models is a crucial stage in building phylogenetic trees for VHF. Using the proper evolutionary tree model appropriately depicts the evolutionary relationships between the various EBOLAs and Marburg.

The alignment of the various VHF nucleotide or amino acid sequences was the basis for this model's selection. The features of the sequence data, such as the quantity, location, and rate of evolution of sequence differences, should be taken into consideration while selecting the evolutionary tree model. A range of accessible software tools can be used to generate the GTR-GAMMA model, which is the proper evolutionary tree model after it has been selected. The evolutionary links between the various species may then be determined using the resulting phylogenetic tree, which can reveal crucial information about the virus's evolution and distribution [25]. The Maximum Likelihood Estimate (MLE) computes the likelihood of the observed data given the model and represents the evolutionary relationships among the species using a structure resembling a tree. Based on the tree topology, the model that maximizes the likelihood of the observed data is chosen. Using this method, phylogenetic trees between Marburg and Ebola have been created. Reconstructing the evolutionary history of any group of organisms, including Ebola and Marburg, is a powerful application of MLE [26].

2.4. Phylogenetic Tree

The aligned sequences were then utilized to create phylogenetic trees using ML [27]. Given the input sequencing data, it employs a maximum likelihood method to find the most likely evolutionary tree. In phylogenetic analysis, an out-group is a reference sequence or taxon distinct from the group of interest yet phylogenetically connected to it.

The phylogenetic tree aims to determine the direction of evolutionary change within the in-group or group of interest.

The out-group sequences usually occupy a basal position in the evolutionary tree concerning the VHF viruses, making the interactions directional. Frequently employ related filoviruses like the Marburg virus as an out-group for phylogenetic trees of the Ebola virus. The Marburg virus, which belongs to the Filoviridae family, shares phylogenetic similarities with Ebola viruses. Knowledge of the evolutionary relationships within the Ebola virus genus can be gained by using the Marburg virus as an out-group, which also aids in rooting the tree. The species with similar genetic sequences are grouped to form the tree, with the most nearly. Figures (6) and (7) show the phylogenetic tree for Ebola and Marburg.

Edelweiss Applied Science and Technology
ISSN: 2576-8484
Vol. 8, No. 6: 9450-9460, 2024
DOI: 10.55214/25768484.v8i6.4024

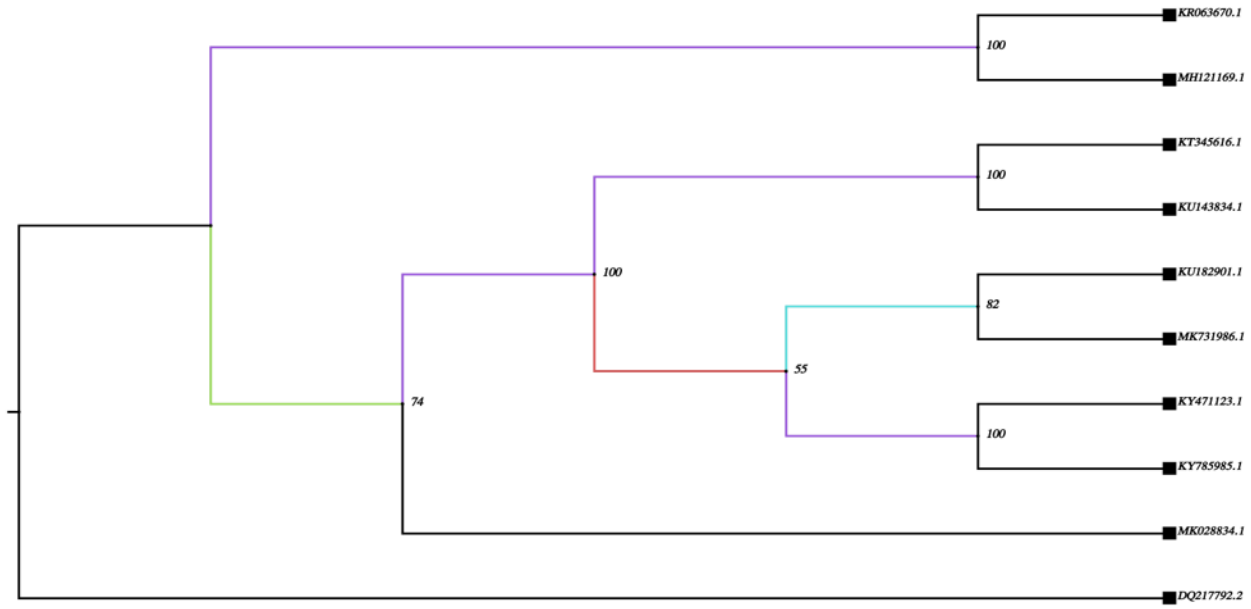

Figure 6.
The phylogenetic tree of (10) Ebola by the Likelihood approach with 100 bootstraps

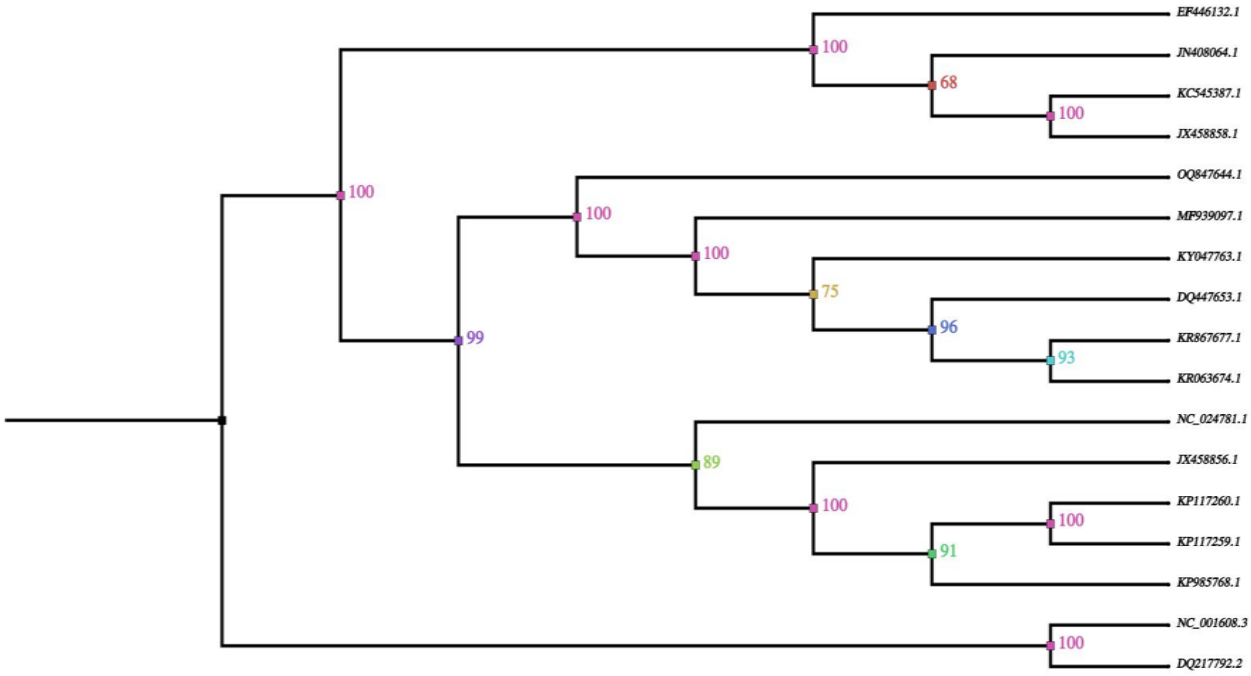

Figure 7.
The phylogenetic tree of (17) Marburg by the maximum likelihood approach with 100 bootstraps.

The correctness of the data used to construct a phylogenetic tree determines how well the tree is evaluated. A few examples of the variables that affect data correctness are the caliber of the data sources, the dependability of the analysis procedures, and the precision of the inference methods. A higher quality of data will result in a more accurate phylogenetic tree. One way to assess the correctness of the branch lengths is to see how closely the tree matches the species' natural evolutionary history. By

Edelweiss Applied Science and Technology
ISSN: 2576-8484
Vol. 8, No. 6: 9450-9460, 2024
DOI: 10.55214/25768484.v8i6.4024

bootstrapping the quality of the data supporting the tree, one can determine how certain the conclusions derived from it are.

2.5. Phylogenetic Tree Analysis

The proposed method extracts single nucleotide polymorphisms (SNPs) in the last stage, and genomes mutate. These are the steps that make up a phylogenetic tree study. SNPs aid in predicting the proper medication or immunization for a specific illness, the risk of infection and transmission, and environmental factors influencing the sickness. SNPs can help monitor family members' traits. Most substitution variations are associated with disease.

Table 2.
Information about SNPs and Mutation

SNPs for Ebola	Mutations for Ebola	SNPs for Marburg	Mutations for Marburg
C -> C: 4569 SNPs	-> A: 1452	A -> G: 12488 SNPs	A -> G: 20411
A -> -: 6897 SNPs	C -> G: 4385	A > -: 27 SNPs	- -> G: 2
A -> G: 8266 SNPs	- -> G: 642	A -> T: 2644 SNPs	G -> A: 21469
G -> -: 4258 SNPs	A -> G: 12177	T -> G: 1568 SNPs	A -> C: 19138
C -> -: 4365 SNPs	G -> A: 12705	T -> C: 13947 SNPs	C -> A: 24984
C -> G: 2788 SNPs	A -> C: 11638	C -> T: 13440 SNPs	A -> T: 29777
G -> A: 7164 SNPs	C -> A: 15093	G -> A: 11701 SNPs	T -> G: 20702
G -> T: 3398 SNPs	A -> T: 15852	C -> A: 1908 SNPs	G -> T: 15059
T -> A: 6765 SNPs	T -> G: 11654	C -> G: 685 SNPs	T -> A: 22834
T -> G: 4210 SNPs	G -> T: 8092	G -> T: 1378 SNPs	T -> C: 20515
T -> -: 7284 SNPs	T -> -: 1080	T -> A: 2559 SNPs	C -> T: 19227
-> T: 5694 SNPs	T -> A: 11752	A -> C: 2000 SNPs	G -> C: 10652
-> G: 4074 SNPs	T -> C: 11492	G -> C: 766 SNPs	C -> G: 6081
-> A: 6448 SNPs	C -> T: 11027		G -> -: 17
A -> C: 5790 SNPs	G -> C: 7379	-> A: 2 SNPs	-> A: 53
T -> C: 8083 SNPs	A -> -: 1334	C -> -: 2 SNPs	- -> T: 1
A -> T: 5999 SNPs	G -> -: 682	-> G: 26 SNPs	A -> -: 28
C -> A: 4959 SNPs	C > -: 788	T -> -: 16 SNPs	T -> -: 12
C -> T: 6495 SNPs	- -> T: 1007	-> T: 6 SNPs	C -> -: 13
G -> C: 2980 SNPs	-> C: 783		total for Marburg: 230975
total for ebola: 16285	total for ebola: 141014	total for Marburg:5835	

Edelweiss Applied Science and Technology
ISSN: 2576-8484
Vol. 8, No. 6: 9450-9460, 2024
DOI: 10.55214/25768484.v8i6.4024

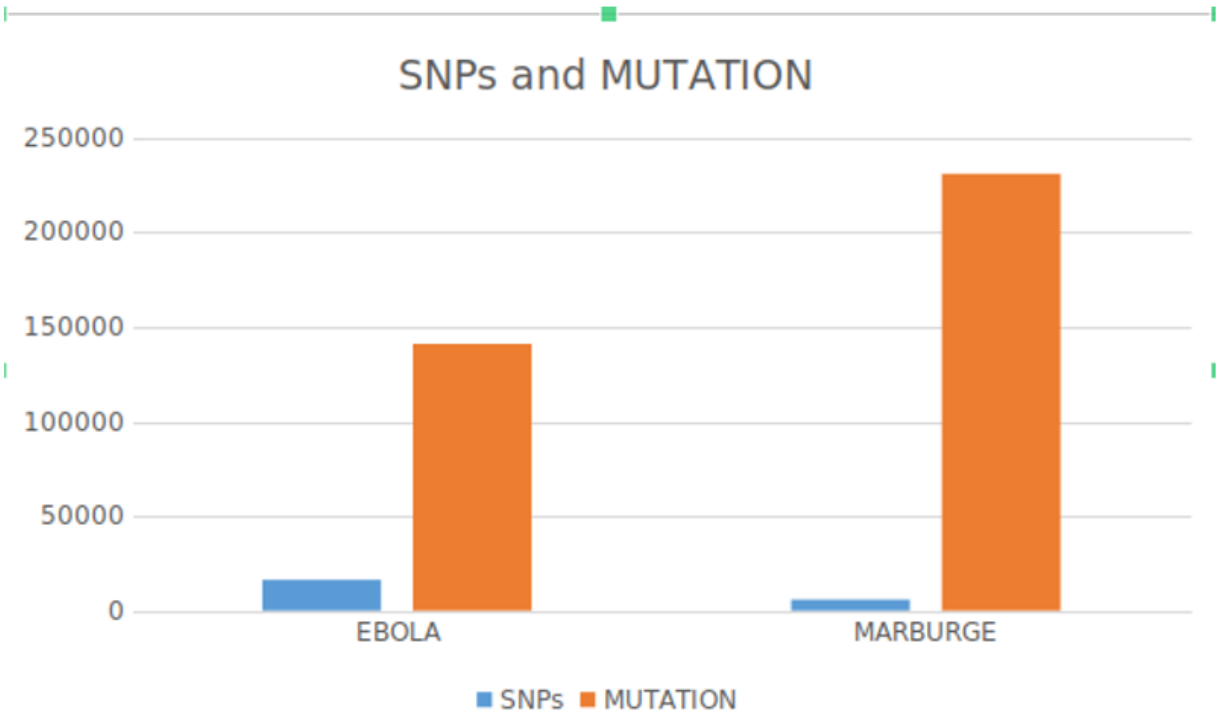

Figure 8.
Explain SNPs and Mutation for EBOLA and MARBURGE.

3. Results and Conclusion

The results of this research paper cover the study's results and the interpretation of these outcomes. We used an MSA program called Muscle because it is the best way to align the complete genome, use a function (find_synteny) to select a suitable complete genome from NCBI, and use the maximum likelihood estimate (MLE) approach to generate phylogenetic trees based on whole genome sequences. The first tree (Figure 6) was for Ebola, and the second (Figure 7) was for Marburg. Given the accuracy of the tree created, tracing the evolution of VHF with the phylogenetic tree and MLE approach is a successful and valuable strategy.

References

[1] Srivastav, Y., Mansoori, M. F., & Pandey, V. K. A Review of Epidemiology of Viral Hemorrhagic Fever. *Disease and Health Research-New Insights*, 117.

[2] Ahmad, A., & Dasril, Z. M. (2024). Virological and Clinical Studies of Viral Hemorrhagic Fever (VHF): A Systematic Literature Review. *Bioscientia Medicina: Journal of Biomedicine and Translational Research*,*8*(9), 4884-4893.

[3] Qi, R., Yu, H., & Yu, X. J. (2024). Hemorrhagic fever viruses. In *Molecular Medical Microbiology* (pp. 2479-2493). Academic Press.

[4] Düx, A., Lwitiho, S. E., Ayouba, A., Röthemeier, C., Merkel, K., Weiss, S., ... & Mangu, C. (2024). Detection of Bombali Virus in a Mops condylurus Bat in Kyela, Tanzania. *Viruses*, *16*(8), 1227.

[5] Malvy, D., & Baize, S. (2024). Ebola and Marburg viruses. In *Molecular Medical Microbiology* (pp. 2281-2308). Academic Press.

[6] Taylor, Derek J., and Max H. Barnhart. "Genomic transfers help to decipher the ancient evolution of filoviruses and interactions with vertebrate hosts." *PLoS pathogens* 20.9 (2024): e1011864.

[7] Düx, A., Lwitiho, S. E., Ayouba, A., Röthemeier, C., Merkel, K., Weiss, S., ... & Mangu, C. (2024). Detection of Bombali Virus in a Mops condylurus Bat in Kyela, Tanzania. *Viruses*, *16*(8), 1227.

[8] Makenov, M. T., Boumbaly, S., Tolno, F. R., Sacko, N., N'Fatoma, L. T., Mansare, O., ... & Karan, L. S. (2023). Marburg virus in Egyptian Rousettus bats in Guinea: investigation of Marburg virus outbreak origin in 2021. *PLos Neglected Tropical Diseases*, *17*(4), e0011279.

[9] Srivastava, S., Sharma, D., Kumar, S., Sharma, A., Rijal, R., Asija, A., ... & Sah, R. (2023). Emergence of Marburg virus: a global perspective on fatal outbreaks and clinical challenges. *Frontiers in Microbiology*, *14*, 1239079.

[10] Khrustalev, V. V., Barkovsky, E. V., & Khrustaleva, T. A. (2015). Local mutational pressures in genomes of Zaire ebolavirus and Marburg virus. *Advances in Bioinformatics*, *2015*(1), 678587.

[11] Espy, N., Nagle, E., Pfeffer, B., Garcia, K., Chitty, A. J., Wiley, M., ... & Palacios, G. (2019). T-705 induces lethal mutagenesis in Ebola and Marburg populations in macaques. *Antiviral Research*, *170*, 104529.

[12] Peterson, A. T., & Holder, M. T. (2012). Phylogenetic assessment of filoviruses: how many lineages of M arburg virus?. *Ecology and Evolution*, *2*(8), 1826-1833.

[13] Jun, S. R., Leuze, M. R., Nookaew, I., Uberbacher, E. C., Land, M., Zhang, Q., ... & Ussery, D. W. (2015). Ebolavirus comparative genomics. *FEMS microbiology reviews*, *39*(5), 764-778.

[14] y Piontti, A. P., Perra, N., Rossi, L., Samay, N., & Vespignani, A. (2018). *Charting the next pandemic: modeling infectious disease spreading in the data science age*. Springer.

[15] Scarpa, F., Bazzani, L., Giovanetti, M., Ciccozzi, A., Benedetti, F., Zella, D., ... & Ciccozzi, M. (2023). Update on the Phylodynamic and Genetic Variability of Marburg Virus. *Viruses*, *15*(8), 1721.

[16] Nyakarahuka, L., Ayebare, S., Mosomtai, G., Kankya, C., Lutwama, J., Mwiine, F. N., & Skjerve, E. (2017). Ecological niche modeling for filoviruses: a risk map for Ebola and Marburg virus disease outbreaks in Uganda. *PLoS currents*, *9*.

[17] "NCBI Virus.https://www.ncbi.nlm.nih.gov/labs/virus/vssi/#/ (accessed April 5, 2022).

[18] A.-N. Alaa Khudair Abbas Al-Khafaji, "Phylogenetic Tree Construction to Reveal the Detailed Evolution of SARS-CoV-2," JOURNAL OF ALGEBRAIC STATISTICS, vol. 13, no. 2, pp. 538-549, 2022.

[19] I. J. Myung, "Tutorial on maximum likelihood estimation," Journal of Mathematical Psychology, vol. 47, no. 1, pp. 90–100, 2003, doi: 10.1016/S0022-2496(02)00028-7.

[20] Nahhas, A. F., & Webster, T. J. (2022). A review of treating viral outbreaks with self-assembled nanomaterial-like peptides: From Ebola to the Marburg virus.*OpenNano,8*, 100094.

[21] Bruhn, J. F., Kirchdoerfer, R. N., Urata, S. M., Li, S., Tickle, I. J., Bricogne, G., & Saphire, E. O. (2017). Crystal structure of the Marburg virus VP35 oligomerization domain.*Journal of virology*, *91*(2), 10-1128.

[22] Babirye, P., Musubika, C., Kirimunda, S., Downing, R., Lutwama, J. J., Mbidde, E. K., ... & Wayengera, M. (2018). Identity and validity of conserved B cell epitopes of filovirus glycoprotein: Towards rapid diagnostic testing for Ebola and possibly Marburg virus disease. *BMC infectious diseases*, *18*, 1-19.

[23] Belyi, V. A., Levine, A. J., & Skalka, A. M. (2010). Unexpected inheritance: multiple integrations of ancient bornavirus and ebolavirus/marburgvirus sequences in vertebrate genomes. *PLoS pathogens*, *6*(7), e1001030.

[24] Taylor, D. J., Leach, R. W., & Bruenn, J. (2010). Filoviruses are ancient and integrated into mammalian genomes. *BMC evolutionary biology*,*10*, 1-10.

[25] Baize, S., Pannetier, D., Oestereich, L., Rieger, T., Koivogui, L., Magassouba, N. F., ... & Günther, S. (2014). Emergence of Zaire Ebola virus disease in Guinea. *New England Journal of Medicine*, *371*(15), 1418-1425.

[26] C. Guyeux, B. Al-Nuaimi, B. AlKindy, J. F. Couchot, and M. Salomon, "On the reconstruction of the ancestral bacterial genomes in genus Mycobacterium and Brucella," BMC Systems Biology, vol. 12, Nov. 2018, doi:10.1186/s12918-018-0618-2.

[27] B. Talib, H. Al-Nuaimi, and P. C. Guyeux, "Ancestral Reconstruction and Investigations of Genomics Recombination on Chloroplast Genomes," Journal of Integrative Bioinformatics, 2017.

Edelweiss Applied Science and Technology
ISSN: 2576-8484
Vol. 8, No. 6: 9450-9460, 2024
DOI: 10.55214/25768484.v8i6.4024