Genomic Variation and Phylogenetic Analysis of Rabies Lyssavirus: Online Database Mining and Computational Approaches

Wardah Shams ,[1] Farrukh Mehmood Alvi.[2]

Abstract: Rabies lyssavirus, a species within the Rhabdoviridae family. This neurotropic virus is primarily transmitted through the bite of infected animals, particularly dogs, bats, and other wildlife. Its genomic structure consists of a single-stranded RNA and exhibits a bullet-shaped appearance under electron microscopy. Rabies poses a significant public health threat globally due to its nearly 100% fatality rate once clinical symptoms appear. Clinical symptoms of rabies include neurological signs such as hydrophobia and aerophobia, which progress to severe encephalitis and ultimately result in death.

Material and Methods: The study, using a retrospective observational design over six months, precisely selected rabies lyssavirus sequences meeting stringent inclusion criteria and excluding irrelevant or low-quality data. GenBank and NCBI Nucleotide databases provided comprehensive genomic sequences for the N and G genes. Computational analysis, including Clustal Omega and BLAST, identified nucleotide variations, while MEGA software constructed phylogenetic trees. This approach allowed for dynamic insights into genomic diversity and evolutionary relationships among the studied rabies lyssavirus strains.

Results: The N gene demonstrates lower variability than the G gene, reflecting functional constraints essential for viral survival. Synonymous mutations prevail in the conserved N gene, contrasting the G gene's unique mutations, hinting at immune-driven adaptations aiding viral survival. Mutations in the G gene which are responsible for impact blood-brain barrier permeability, influencing RABV pathogenesis. A pattern of Adenine to Guanine substitutions, suggesting selective pressures or mutational biases. Phylogenetic analyses unveil interconnectedness among European and global RABV strains, hinting at shared transmission routes.

Conclusion: This study unravels contrasting mutation rates between Rabies lyssavirus' N and G genes, showcasing the N gene's conserved role and the G gene's diverse mutations in potential immune evasion. However, these findings offer a limited view of the virus's complex evolution, emphasizing the need for further research to better understand RABV's dynamics for improved vaccine development and control.

Keywords: Genomic variation, Phylogenetic Analysis, Rabies Lyssavirus, MEGA.

1. University of the Punjab.
2,,. Institute of Public Health, Lahore
.

Submission Date: 15-05-2025
1st Revision Date: 05-24-2025
Acceptance Date: 25-06-2025

Introduction

Lyssaviruses (RABVs) are a distinct group of RNA viruses characterized by their unique bullet-shaped morphology. With a single-stranded, negative-sense RNA genome, these viruses are primarily known for causing the persistent zoonotic disease, rabies.[1] Regional genetic diversity is evident among rabies viruses (RABVs) across different continents, with bats playing a key role in viral spread.[2]. Virus primarily targets neurons, it can also infect non-neuronal cells, albeit with reduced efficiency.[3] Notably, significant variations in infectivity have been observed between circulating and deposited strains, particularly in astrocytes.[4] The RABV genome is a non-segmented negative-sense

RNA of approximately 12 kb, encoding five viral proteins: nucleoprotein (N), phosphoprotein (P), matrix protein (M), glycoprotein (G), and large protein (L). These proteins are integral for viral replication, assembly, immune evasion, and pathogenesis. The N protein binds to viral RNA, forming a ribonucleoprotein (RNP) complex that serves as a template for viral RNA synthesis. The G protein, located on the viral envelope, facilitates receptor binding and viral entry into host cells. RABV infection occurs through the saliva of infected animals, primarily transmitted via bites. Upon entry into the host, RABV targets peripheral nerves and undergoes retrograde transport to the central nervous system (CNS). [5]

RABV can target Schwann cells (SCs), which are non-neuronal neuroglia that envelop axons within peripheral nerves.[6] Leading to the severe neurological consequences associated with the disease.[7] Individuals infected with rabies do not survive the disease, and the outcome is fatal.[8]

A comprehensive analysis of the genomic variation and phylogenetic relationships among different strains of dog rabies lyssavirus was conducted. The collected sequences were subjected to computational analysis using software tools to aid in comparing the obtained sequences with reference genomes to identify similar strains, constructing phylogenetic trees, and estimating genetic distances among different virus strains concerning the reference genomes.

Material and Methods

This study employed a retrospective observational design to investigate the genomic variation and phylogenetic relationships of the Rabies lyssavirus. The research was conducted over a period of six months following the finalization and approval of the research question. For sample selection, inclusion criteria prioritized high-quality, full-genome sequences that were accompanied by essential metadata such as host species, sampling location, and collection date. Sequences specifically targeting the nucleoprotein (N) and glycoprotein (G) genes of Rabies lyssavirus, with complete or nearly complete gene coverage, were particularly sought. Conversely, exclusion criteria were applied to omit sequences from other lyssavirus species, low-quality or partial sequences, those lacking critical metadata, and sequences derived from laboratory-adapted or vaccine strains. Data collection was conducted using well-established online genetic databases, specifically GenBank and NCBI Nucleotide, chosen for their extensive repository of complete rabies lyssavirus genomic sequences. The search process incorporated structured retrieval strategies focused on the N and G genes. During the data analysis phase, a series of computational tools and software were utilized. Initially, Clustal Omega was employed to identify the most dissimilar sequences among those released within the same year from a given country. This was followed by a BLAST (Basic Local Alignment Search Tool) analysis to align the selected N gene sequences with a reference sequence and perform pairwise comparisons to detect nucleotide variations. Finally, MEGA (Molecular Evolutionary Genetics Analysis) software was used to construct phylogenetic trees, facilitating a detailed understanding of the evolutionary relationships among the viral strains.

Results

Total 100 sequences were downloaded in FASTA. format. The retrieved data were all from host specie *Canis lupus familiaris*. All the sequences selected were released on online databases after year 2005. After downloading the sequences all of them were subject to rigorous filtering, removing any vaccine strain or laboratory adapted sequences. Sequences reported from a country in the same year were subjected to Clustal Omega in which the most dissimilar sequence was selected. In year 2015, five sequences were published from Germany in which only LN713661.1 was selected. In 2008, two sequences were published from which EF206718.1 was selected. From 2018 in Poland LT993233.1 was selected out of three downloaded sequences. From China in 2012 sequence JX276550.1, 2019 MT895968.1 and in 2021 MT895964.1 were selected.

After the analysis, each type of nucleotide variation is identified located on sequences. Most number of variations were located in Hungary 2019. Total 46 Locus at which mutation occurred. Least number of variations were identified in China 2012. (See Table 1)

Table 1: Genomic sequence selection

No. of Sequences Downloaded	Included	Excluded			Comments on Dissimilar Excluded
		Total	Similar	Dissimilar	
100	12	88	16	72	• Vaccine Strain: 13 • Incomplete Sequences: 45 • Laboratory Adapted Strains: 14

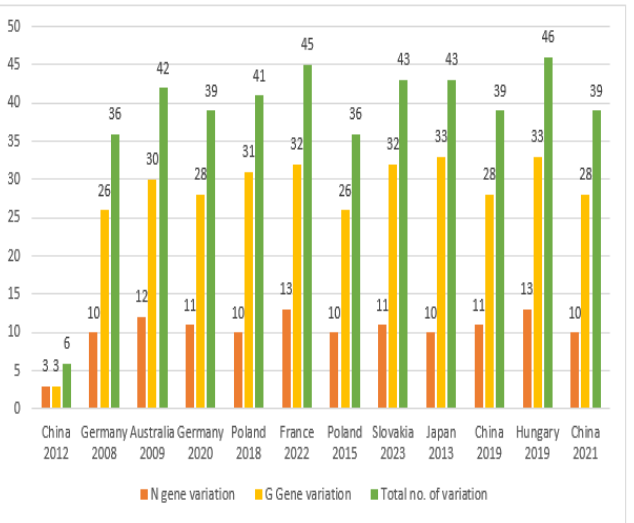
After filtering 100 sequences only 12 of the following were selected which fulfilled the inclusion criteria.

Table 2: Selected Sequences

Sr. No.	Accession Number	Country	Year
1	JX276550.1	China	2012
2	EF206718.1	Germany	2008
3	EU182346.1	Australia	2009
4	LR812028.1	Germany	2020
5	LT993233.1	Poland	2018
6	OV696601.1	France	2022
7	LN713661.1	Poland	2015
8	OP642459.1	Slovakia	2023
9	AB781935.1	Japan	2013
10	MT895968.1	China	2019
11	MK111080.1	Hungary	2019
12	MT895964.1	China	2021

Reference sequences of N gene (RABVgp1) and G gene (RABVgp4) were downloaded from Gene NCBI in FASTA. format. All 12 sequences were BLAST using NCBI nucleotide BLAST separately for N Gene and G Gene. Aligned sequences from BLAST were saved in txt. file for further analysis locating the nucleotide variation in sequences. Pairwise with dots for identity was used to easily identify the variation.

Figure 1: Nucleotide Variation based on each Seq



The sequences selected in Germany, China and Poland has been analyzed. As expected, the number of mutations in all of the stated countries increased over the years. Although it has been observed that number of mutations increase in N gene is less. The least number of mutations were observed in China 2012.

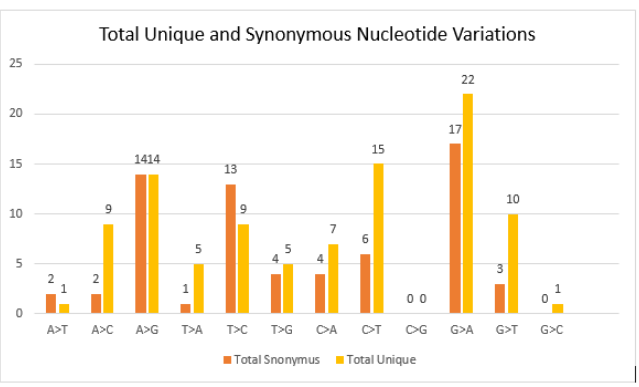
Table 3: Nucleotide Variation Analysis based on Countries over the years

Country	Year	Total Nucleotide Variation	Gene	No. of Nucleotide Variation
Germany	2008	36	N	10
			G	26
	2020	39	N	11
			G	28

China	2012	6	N	3
			G	3
	2019	39	N	11
			G	28
	2021	39	N	10
			G	29
Poland	2015	36	N	10
			G	26
	2018	40	N	10
			G	30

Most number of synonymous mutations in nucleotide were observed in variation of G with A (17).

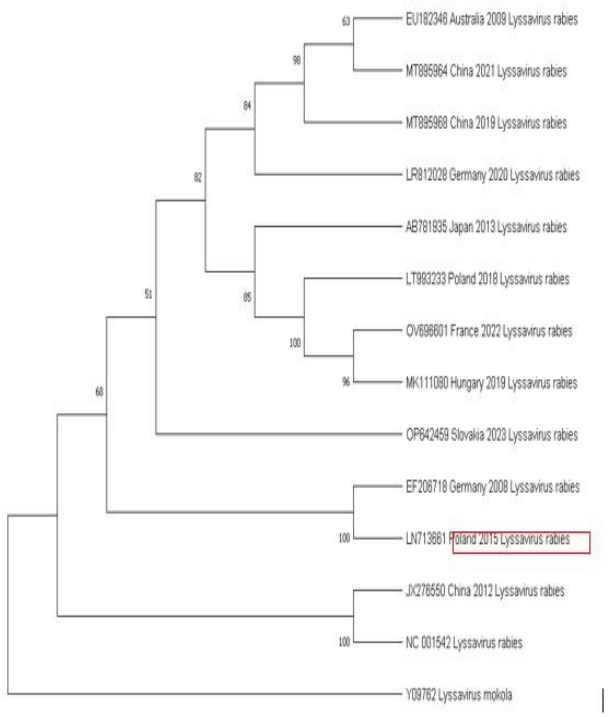
Figure 2: Comparison of Synonymous and Unique Nucleotide Variation



A phylogenetic tree is a graphical representation of the evolutionary relationships among a group of organisms, genes, or sequences.9 It illustrates the evolutionary divergence and relatedness between different species or entities based on their genetic, morphological, or behavioral characteristics. Lyssavirus rabies sequences collected from China in 2021 show a close relationship with the Reference Sequence NC_001542. Sequence Isolates from France and Hungary have common ancestor from Poland. Sequences isolated from Poland in 2015 and China in 2012 are closely related. All the isolates from Europe are from the same clade. The isolates from Europe, including Slovakia, Germany, and others, form a distinct clade, suggesting a shared evolutionary history and likely a common source. This clade might represent a regional distribution or possibly a prevalent strain circulating within Europe. The most recent isolate from the Slovakia has common ancestor with isolates from Germany and China. Phylogenetic analysis is performed using the MEGA11 software with maximum likelihood method having bootstraps 1000 with Kimura 2 parameter model.

The sequence obtained from Australia in 2009 notably exhibits the least similarity with the reference sequence. The clades observed in the Lyssavirus rabies reference sequence trace their origin back to the reference strain, which evolved from Lyssavirus mokola. Phylogenetic tree is as follows

Figure 2: Phylogenetic Analysis of Lyssavirus rabies sequences of interest



Discussion

The N and G genes of rabies viruses show varying rates of mutation, raising concerns about the evolutionary dynamics and potential functional effects of these variations. The N gene is less variable than the G gene. This sheds light on the virus's putative virulence factors, adaptive processes, and interactions with the host immune system. Because of functional restrictions that are critical for viral survival, the N gene—which is required for viral replication—tends to be more conserved than the G gene.

A study of Georgian canine RABV strains found that the nucleoprotein (N) gene was more conserved than the attachment glycoprotein (G) gene, with the majority falling within the Cosmopolitan clade, underlining the N gene's role in regional RABV evolution.[10]

In contrast, the G gene has more distinct mutations, indicating exposure to the host immune system and probable adaptation for immune evasion and survival. This reflects the G gene's functional significance in viral attachment and immune evasion, as compared with the more stable N gene.

The study conducted China on a wild-type RABV (CNIM1701) from a rabid bovine underscores the critical role of mutations in the G gene of the Rabies lyssavirus. G gene mutations impact blood-brain barrier (BBB) permeability These results demonstrate that the pathogenesis of RABV is partially dependent on G expression and BBB permeability.[11]

A continuous pattern of A to G substitutions, implying that the virus's genetic evolution is shaped by selective pressure or mutational bias. This could be due to environmental or host factors that favour RABV variants with these mutations.[12] The phylogenetic analysis of Lyssavirus strains, notably European RABV isolates, revealed interconnectivity and probable transmission paths across multiple geographic regions. The common ancestry of isolates from France, Hungary, Poland, China, Germany, and Slovakia supports global transmission channels or shared infection sources. The study's reach is limited by sequence diversity representation and gene emphasis. More research should look into broader genetic segments and functional implications to better comprehend.

Conclusion

This study on Rabies lyssavirus highlights a substantial difference in mutation rates between the nucleoprotein (N) and attachment glycoprotein (G) genes, portraying their distinct genomic variation. The observed lower variability in the N gene suggests a conserved role vital to the virus's life cycle, likely constrained by functional necessities. Conversely, the G gene displays a higher frequency of unique mutations, particularly in antigenic sites, indicating potential involvement in immune evasion and viral persistence. These findings present a limited understanding of the complex evolutionary dynamics and the underlying influence of selective pressures on the virus. Further investigation is necessary to unravel the intricate mechanisms governing RABV evolution and its implications for effective vaccine development and control strategies.

References

1. Rupprecht C KI, Meslin F. Lyssaviruses and rabies: current conundrums, concerns, contradictions and controversies. F1000Research [Internet]. 2017;6:184. lyssaviruses and rabies: current conundrums, concerns, contradictions and controversie. 2017.
2. Schöler L, Le-Trilling VTK, Dittmer U, Fiedler M, Trilling M. 10;16(5) Establishment and clinical validation of an in-cell-ELISA-based assay for the rapid quantification of Rabies lyssavirus neutralizing antibodies. PLoS Negl Trop Dis. 2022
3. Walker PJ, Blasdell KR, Calisher CH, Dietzgen RG, Kondo H, Kurath G, Longdon B, Stone DM, Tesh RB, Tordo N, Vasilakis N, Whitfield AE. ;99(4):447–448. ICTV Virus Taxonomy Profile: Rhabdoviridae. *J Gen Virol*. 2018.
4. Cai M, Liu, H., Jiang, F., Sun, Y., Wang, W., An, Y., Zhang, M., Li, X., Liu, D., Li, Y., Yu, Y., Huang, W., & Wang, Y. (2022). 11(1), 1474–1487. Analysis of the evolution, infectivity and antigenicity of circulating rabies virus strains. Emerging microbes & infections,. 2022.
5. Takahashi T, Inukai, M., Sasaki, M., Potratz, M., Jarusombuti, S., Fujii, Y., Nishiyama, S., Finke, S., Yamada, K., Sakai, H., Sawa, H., Nishizono, A., Sugiyama, M., & Ito, N. (2020). Genetic and Phenotypic . Viruses, 12(9), 914. Characterization of a Rabies Virus Strain Isolated from a Dog in Tokyo, Japan in the 1940s. 2020.
6. Nel LH. Lyssaviruses and the fatal encephalitic disease rabies. Front Immunol. 2021
6. Potratz M, Zaeck LM, Weigel C, Klein A, Freuling CM, Müller T, Finke S. Neuroglia infection by rabies virus after anterograde virus spread in peripheral neurons. Acta Neuropathol Commun. 2020
7. Jackson AC. Human disease. In: Wunner WH, Jackson AC, editors. Rabies: Scientific basis of the disease and its management. 3rd ed. p. 269–298. Academic Press; 2013.
8. Koraka P, Martina BM. Antivirals for human use against rabies and prospects for their future application. 37(2):673–80 Rev Sci Tech. 2018.
9. Olarinmoye AO, Kamara KV, Jomah ND, Olugasa BO, Ishola OO, Kamara A, Luka PD. Molecular detection of rabies virus strain with N-gene that clustered with China lineage 2 co-circulating with Africa lineages in Monrovia, Liberia: First reported case in Africa. Epidemiol Infect. 2019
10. Huaman C PA, Clouse C, Feasley A, Rader M, Rice GK, Luquette AE, Fitzpatrick MC, Drumm HM, Yan L, Cer RZ, Donduashvili M, Buchukuri T, Nanava A, Hulseberg CE, Washington MA, Laing ED, Malagon F, Broder CC, Bishop-Lilly KA, Schaefer BC. . Genomic Surveillance of Rabies Virus in Georgian Canines. Viruses. 2023 Aug 24;15(9):1797. 2023.
11. Li C WY, Liu H, Zhang X, Baolige D, Zhao S, Hu W, Yang Y. . Change in the Single Amino Acid Site 83 in Rabies Virus Glycoprotein Enhances the BBB Permeability and Reduces Viral Pathogenicity. Front Cell Dev Biol. 2021 Feb 9;8:632957. 2021.
12. Condori RE, Niezgoda M, Lopez G, Acosta Matos C, Diaz Mateo E, Gigante C, Hartloge C, Pereira Filpo A, Haim J, Satheshkumar PS, Petersen B, Wallace R, Olson V, Li Y. Using the LN34 Pan-Lyssavirus Real-Time RT-PCR Assay for Rabies Diagnosis and Rapid Genetic Typing from Formalin-Fixed Human Brain Tissue. Viruses. 2020 Jan 18;12(1)

13. Bonnaud EM, Clément T, Dacheux L, Holmes EC, Monchatre-Leroy E, Tanguy M, Bouchier C, Cliquet F, Barrat J, Bourhy H. Comparison of intra- and inter-host genetic diversity in rabies virus during experimental cross-species transmission. PLoS Pathog. 2019 Jun 20;15(6).
14. Bourhy H, Reynes JM, Dunham EJ, Dacheux L, Larrous F, Huong VTQ, Xu G, Yan J, Miranda MEG, Holmes EC. The origin and phylogeography of dog rabies virus. J Gen Virol. 2018 Nov;89(Pt 11):2673–81.
15. De Almeida GL, Cargnelutti JF, Ries AS, Ferreira JC, Rosa JCA, Batista HBC, Flores EF, Weiblen R. Sequence analysis of nucleoprotein gene reveals the co-circulation of lineages and sublineages of rabies virus in herbivorous in Rio Grande do Sul state, Brazil. Braz J Microbiol. 2020 Jun;51(2):837-46