Identification of clade-specific single nucleotide polymorphisms for improved Rabies virus surveillance in *Canis familiaris* host

Ankeet Kumar[1], M Kaustubh[1], Sujith S Nath[1] & Utpal Tatu[1]*

Affiliation:

[1]Department of Biochemistry, Division of Biological Science, Indian Institute of Science, Bangalore, 560012

Corresponding Author:

*Utpal Tatu, Department of Biochemistry, Division of Biological Science, Indian Institute of Science, Bangalore, 560012

Email: tatu@iisc.ac.in

Abstract

Rabies infection continues to be prevalent in many countries, causing a significant number of fatalities annually, predominantly in resource-poor countries. Multiple RABV variants have stabilised throughout evolution, leading to the formation of several clades which have been named after their geographic region of prevalence. Extensive studies have been conducted on the origin, occurrence and spread of RABV clades. However, a systematic understanding of the genetic diversity within and among these clades is still lacking. Previous studies on phylogeny, evolution, and diversity were mainly based on nucleotide sequences of one or two genes. In this study, we utilised whole genome sequences obtained from dog hosts belonging to four major clades to investigate the diversity and phylogeny of the RABV clades. We identify Single Nucleotide Polymorphisms (SNPs) of varying frequencies across the genomes of these clades. We categorise these SNPs into various classes based on their impact on the protein sequence and region of occurrence in the genome. Notably, we report 138 "universal" mutations, which occur with a frequency of more than 90% across all clades. We also report 23 "clade-specific mutations" in three clades: Asian, Arctic, and Africa-2, which could play a crucial role in the fitness of the virus and offer a potential for targeted clade surveillance. Our study also investigates the effects of these mutations on protein function and virus pathogenesis. Overall, our findings contribute to expanding knowledge about RABV diversity and evolution, with important implications for effectively tracking and combatting RABV transmission effectively.

Keywords Rabies virus (RABV), mutation analysis, whole-genome analysis, SNPs, canine rabies.

Introduction

Viruses of the genus *Lyssavirus* cause a disease called Rabies. It is a deadly zoonotic viral encephalitis that significantly threatens human and animal health. The *Lyssavirus* genus comprises 17 species, including the classical Rabies virus (RABV) (prototype of the genus) and other Rabies-related viruses. These species are classified into phylogroups based on genetic composition and antigenic differences (1). While most *Lyssavirus* species have a limited impact on human health due to their narrow host range (2,3), RABV displays broad

host tropism and is responsible for approximately 59,000 deaths annually across the globe (4). Bat and Dog serve as the primary reservoirs of RABV, forming distinct phylogenetic groups suggesting distinct host-dependent transmission trajectories and nominal mixing. Bat-associated RABV strains are prevalent in the New World (Americas), while carnivore-associated RABV strains are distributed globally, except in Antarctica and Australia (5). Almost 99% of human deaths caused by RABV globally are attributed to dog bites (6). In India and other developing nations of Asia and Africa, dogs are the only known transmitters of rabies to humans. India bears the highest burden of rabies deaths, accounting for around 20,000 fatalities annually (7), followed by China (8).

The RABV genome is around 12 kilobases long, consisting of non-segmented, single-stranded negative-sense RNA enclosed within an enveloped, bullet-shaped virion (9). It contains five protein-coding genes arranged in the order of Nucleoprotein (N), Phosphoprotein (P), Glycoprotein (G), Matrix (M), and Large (L) protein gene. The L gene is the largest, accounting for approximately 53% of the total nucleotide content, encoding a crucial multifunctional protein called RNA-dependent RNA polymerase (L protein) of 242 kDa (10). L protein plays a vital role in the replication process and mRNA capping with GDP polyribonucleotidyltransferase activity (11). During replication, the L protein collaborates with the P protein, which enhances the activity of the L protein (10). Notably, the RABV L protein lacks proofreading activity, resulting in variations and the emergence of new variants during evolution, as well as the formation of quasi-species during infection (12). The full-length P protein (297 amino acids) interacts with the L protein via its N-terminal region, while its shorter isoforms (P2-P5) with truncated N termini likely serve alternative functions (13,14). The N protein forms the core of the nucleocapsid, encapsulating the viral RNA in a helical structure (15). The M protein, encoded by the M gene, interacts with the N protein and the cytoplasmic domain of the G protein, facilitating viral envelopment (16,17). The extracellular region of the G protein interacts with host receptors, playing a crucial role in viral entry (18).

RABV exhibits a high mutation rate that varies across the genome, with the N protein exhibiting the lowest mutation rate (20). Previous studies on RABV phylogeny and evolution have mainly relied on sequences of N and G genes. These studies provide an incomplete picture of RABV genome evolution owing to varying selection pressure across genes. Now, with ample whole genome sequences of RABV, a comprehensive understanding of its evolutionary trajectory and selective mutational pressure across the entire genome can be obtained. The whole genome-based approach provides a holistic view by considering the mutational landscape across the genome, thus eliminating any bias arising from selection pressure differences on genes. The impact of mutations that have stabilised over the evolution is essential for applications such as drug and vaccine development (21). Also, predicting variants of clinical and epidemiological importance and identifying the biological mechanisms underlying viral genome evolution, immune evasion, and vaccine responses. Some variants of RABV are geographically restricted, while others are distributed across multiple continents. As these variants coexist, there is a critical need to investigate genomic differences and their distribution to gain deeper insights into viral evolution and transmission dynamics. However, to our knowledge, no systematic study has been conducted to discover the global occurrence of mutations in the RABV genome and their distribution among its different clades.

Therefore, in this study, we present a comprehensive analysis of SNPs across the entire genome of RABV and compare their occurrence across four clades: Arctic, Asian, Africa-2 and Cosmopolitan. The dataset comprised 417 whole genome sequences isolated exclusively from dog hosts, representing spatial information from 59 countries and a temporal range from 1931 to 2020. We find that the SNPs predominantly occur in the coding regions of the five genes, with a significant number also observed in non-coding regions as well. While most SNP variations result in synonymous changes (SNP-silent mutation), some lead to non-synonymous amino acid residue modifications (SNP mutation). We found that most of the SNP variations are shared among or between clades, but there exist clade-specific mutations. These clade-specific mutations may not only have functional implications for the virus's fitness but also hold potential as tools for clade identification and tracking. By unravelling the complex landscape of genetic diversity within the RABV, our work provides valuable insights into its evolutionary dynamics, clade-specific variants, and potential implications for transmission and control strategies.

Methods:

Collection and Curation of the Data

A dataset comprising 421 whole genome sequences of the RABV was obtained from the Bacterial and Viral Bioinformatics Resource Center (BV-BRC) database accessed on 13 March 2023. The dataset was further categorised into clades using RABV-GLUE (RABV-GLUE), a web-based tool from the University of Glasgow (22). This tool classified the dataset into twenty-seven subclades belonging to six major clades: Africa-2, Asia, Arctic, Cosmopolitan, Indian-subcontinent, and Bat. A metadata file was manually curated for the dataset, containing information regarding clades, country, collection year, and accession number of the sequence, and this metadata file is available as a Supplementary file 1.

Multiple Sequence Alignment and Phylogenetic Analysis

To determine the evolutionary relationship and origin of different clades in relation to each other, firstly, a Multiple Sequence Alignment (MSA) was performed using the MAFFT v7.505 software in the command line (bash) with default options (23). Subsequently, maximum likelihood tree was constructed for 421 RABV sequences from dogs and 17 reference sequences of lyssaviruses using the IQTREE software (24), incorporating evolutionary models and likelihood scores. Bootstrap iterations (1000 iterations) were performed to assess the reliability of the inferred relationships in the phylogenetic trees. The resulting tree was visualised using the iITOL tool (25), with sequences colour-coded based on the major clades.

Identification and Classification of SNPs Based on Occurrence in the Genome

The identification of SNPs involves several steps. The script utilises the NUCMER v3.1 (26) tool to generate pairwise alignments of query sequences with the reference (NC001542) and produces a table with nucleotide changes. To convert nucleotide changes into amino acid changes, an R script (27) was modified, which is available as Supplementary File 2. The modifications involved enhancements related to the consolidation of SNPs within the same codon. Additionally, a refinement was made by switching from the "round" function to the

"ceiling" function, thereby contributing to improved data processing and accuracy. The modified script was employed to categorise these mutations into different classes based on their impact on protein sequence. The script integrates various tools to translate nucleotide changes (SNPs) into amino acids and label these changes into different categories like "SNP-silent" (synonymous) mutation, "SNP" (non-synonymous) mutation, "insertion frameshift", "deletion frameshift" if the change lies within the coding region. If the change lies in the non-coding region, it is labelled as "5'UTR" and "3'UTR" in the "extragenic" category.

Classification of Mutations Based on Frequency of Occurrence in Clades

The identified mutations were manually classified into clade-specific and universal mutations based on their occurrence in different clades. "Clade-specific mutations" are variations that are observed in more than 90% of genomes of a particular clade and do not occur in any other clade. "Universal mutations" are variations that occur in more than 90% of genomes across all clades. Indian-subcontinent and Bats clades were excluded from further analysis due to limited genome sequences. The study focused only on four clades: Africa-2 (36 genome sequences), Asian (59 genome sequences), Arctic (46 genome sequences), and Cosmopolitan (276 genome sequences), comprising a total of 417 genomes. In total, 500,780 nucleotide changes were identified, which were subsequently translated into amino acid changes. An R script was created to merge mutations occurring in the same codon, resulting in 478,818 encompassing all mutations.

The methods employed in this study provided a rigorous dataset analysis, including phylogenetic analysis, SNP identification and classification into various classes based on amino acid change and occurrence in clades. These findings contribute to understanding the genetic diversity and evolutionary dynamics of the RABV across different clades.

Results

Phylogeny Reveals the Origins and Divergence of Various Subclades in Rabies Evolution.

A maximum-likelihood-based phylogenetic tree (Figure 1A) was constructed using whole genome sequences isolated from the dog host and reference sequences of other lyssaviruses. The phylogenetic analysis revealed the origin and divergence patterns of various RABV clades. The Asian clade exhibited two distinct branches, one branch represented mainly by the South East Asian 1b (SEA1b) subclade showing continuous presence within China over a range of 1992 to 2018, suggesting localised transmission of the subclade. All SEA1b sequences share a common ancestor with SEA1a (one sequence available). The other branch further leads to two sub-clusters, one leading to the formation of two subclades: SEA2a and SEA4. These clades share a common node but are present in different countries; SEA2a is found in China, and SEA4 in the Philippines, suggesting the formation of regional lineages. The other subcluster, represented solely by the SEA3 clade, demonstrated circulation in four south-east Asian countries and China. The JQ730892 sequence collected in 2010 forms the root of SEA3, which then spread to Myanmar, Thailand, Laos and Cambodia.

In our dataset, the Africa-2 clade is represented by 36 sequences collected from 12 countries (Western African countries, Figure 1B) over the range of years 1987 to 2018. All of the

sequences originated from a single ancestor with the root formed by sequence KX148244 isolated from Guinea in 1990. The sequences from other countries show some mixing, but sequences from Senegal showed no mixing with neighbouring regions.

The Arctic clade, represented by ten countries, had its base formed by the Arctic-like 2 (AL2) subclade, and the sequences were isolated from Japan and China. The Arctic clade arises from the I754 node. Other subclades in the Arctic consist of isolates from various other countries, such as AL3 (represented from Nepal and Iran) and AL1 (Iran, India and Afghanistan). The AL1 clade further branches to AL1a, found in Iran and India, while AL1b is found in Iran and Afghanistan. Notably, the United States of America, Greenland and Canada harboured a distinct subclade called Arctic A. Some sequences isolated from Canada lie on the same branch containing sequences from Greenland, suggesting transboundary migration, while some sequences are exclusive to Canada and form a separate sub-cluster with one sequence from the USA isolated in 2018.

The Cosmopolitan clade is the largest group, representing around 64% of the dataset. It has multiple subclades, which mainly show a localised transmission, but there are some evidences of transboundary migration. The root of the clade is formed at node I542, and all the subclades originate from this point. The earliest branching subclade is AF4, isolated from Egypt and Israel forms a separate branch. AM2a (isolated from Mexico) and AM3a (Brazil) also show independent evolution. The original root is represented by the reference sequence of RABV (NC001542) and LT909527. From this node, two main branches arise, one represented by the group of 6 subclades (CA1, CA2, NEE, ME1a, ME1b, and ME2) isolated from Middle Eastern, Central Asian, Eastern European countries, while the other group of clades (AF1a, AF1b and AF1c) is represented mostly from African countries. The CA2 subclade (isolated from Iran, Russia and Turkey) forms the root, and all other subclades arise from it. There is a single sequence collected by Serbia, which forms an individual branch. The ME2 subclade original sequences are from Turkey, from where it has spread to Iran. The NEE subclade is represented by 5 Eastern European countries, and the CA1 subclade also lies on the same branch whose sequences are isolated from Kazakhstan, Russia and Iran. The largest subclade in this group is represented by ME1a, which majorly circulates in Iran and is closely related to the ME1b subclade. The other major branch, the largest subclade, is AF1b, consisting of 142 sequences, majorly represented from Tanzania. This clade is found to circulate in multiple countries of Central and Southern Africa. Also, sequences isolated from South Africa are closely related to sequences found in Tanzania, which supports previously reported events of transboundary migration (22). Another subclade is AF1c, which is represented by two sequences from Madagascar. Both AF1b and AF1c originate from AF1a (14 sequences), represented by sequences collected from two continents (Africa and Asia). The world map with clade distribution is presented in Figure 1B.

Asian Clade Sequences Show High Diversity at Nucleotide Level

The genetic landscape, based on sequence similarity among distinct clades, reveals intriguing patterns. Our objective was to unravel the nucleotide-level similarity of RABV clades, particularly those isolated from dogs. The degree of similarity spans a range of 87% to 99.9%. Notably, the Asian clade sequences exhibited the lowest similarity, indicating a significant nucleotide diversity within this clade. Specifically, the SEA4 (KX148263) and SEA3 (LC717426) subclade sequences shared a mere 87.6% similarity. For other clades, the

lowest similarity between the two sequences was 90.9% for Arctic clade sequences, 93.6% for Cosmopolitan sequences, and 96% for Africa-2 sequences (Table 1).

Major Clade	Sequence1	Sub-clade	Sequence2	Sub-clade	Percentage
Asian	LC717426	SEA3	KX148263	SEA4	87.60%
Arctic	KX148228	AL3	KM272192	AL2	90.90%
Cosmopolitan	KY210311	AF1b	KX148195	AF1a	93.60%
Africa-2	OK135148	Africa-2	MW057694	Africa-2	96%

Table 1: Nucleotide-Level Divergence in Major RABV Clades.
This table presents percentage sequence similarity between the two least similar genomes of major clades of RABV. These sequences belong to distinct sub-clades within the major clades. The Asian clade exhibits the highest nucleotide-level diversity, with 87.6% similarity between the sequences of SEA3 and SEA4 subclades. In the Arctic clade, the two sequences belonging to the sub-clades AL3 and AL2 showed the lowest sequence similarity. In the Cosmopolitan clade, the lowest sequence similarity is 93.6%. Both these genome sequences were isolated from the African continent, which shows that the African region sequences are more diverse. The lowest similarity between the two Africa-2 sequences is 96%, which is the highest among the clades described. The findings emphasise the genetic variability within and across RABV major clades, contributing to our understanding of their evolutionary dynamics and genetic relationships.

To investigate the variations in selection pressure at the protein level, we conducted an analysis of the occurrence of mutation frequency in clades and different proteins of RABV. The calculation involved normalising the number of mutations by sample number and protein length. Our findings unveiled significant disparities in average mutation frequency across the major clades. The Asian clade exhibited the highest average mutations, followed by Africa-2, Arctic, and Cosmopolitan in order. Importantly, a trend persisted across the individual proteins as well. The P protein consistently displayed the highest mutation frequency across all clades followed by the G, L, and M proteins. In concordance with previous studies, the N protein exhibited the lowest normalized mutation frequency, suggesting a notably higher purifying selection pressure acting on the N protein (28) (Table 2).

Major Clade	Clade	N protein	P protein	M protein	G protein	L protein
Asian	1609.2	0.384	0.511	0.409	0.485	0.443
Africa-2	1456.89	0.366	0.444	0.377	0.443	0.398
Arctic	1319.67	0.317	0.426	0.336	0.395	0.362
Cosmopolitan	849.623	0.225	0.284	0.221	0.254	0.227

Table 2: Normalised Mutation Events Across RABV Clades and Proteins.
This table presents an overview of the normalised mutation frequencies observed across the major clades of RABV (**total mutation events for the clade/number of samples**) and the proteins (**total mutation events for protein/protein length/number of samples**). The clades are arranged in decreasing order of mutation frequency observed. Notably, the Cosmopolitan clade displays the lowest mutation frequency among the four clades, while the Asian clade exhibits the highest frequency of mutations. When examining the individual proteins, we find that the P protein demonstrates the highest mutation frequency, indicating a significant level of genetic variability within this gene. Interestingly, the N and M proteins exhibit comparable

mutation frequencies. On the other hand, the G and L proteins show similar mutation frequencies despite the striking difference in the gene length.

The SNP-silent is the Major Mutation Class in the RABV

The SNPs identified from 417 sequences belonging to four clades were categorised into different classes using an R script (see method section). The genome of the virus is highly compact, which limits the occurrence of mutations that can have a huge impact on the virus by purifying selection (29). Even though about 87% (420,717) of the SNPs are observed in the protein-coding region, most nucleotide variations did not result in an amino acid change. Around 90% of the genome codes for proteins in RABV. Depending on protein sequence variation, these mutations encompass various categories, such as insertion frameshift, deletion frameshift, SNP-silent and SNP. Around 90% (382,328) of SNPs in the protein-coding region are synonymous at the amino acid level, referred to as SNP-silent mutations. The remaining 9.1% (38,389) result in an amino acid change (non-synonymous) referred to as SNP mutation, while other classes of mutations like insertion frameshift and deletion frameshift make up less than 1% of the total. The percentage of these mutation classes varies among different clades.

Interestingly, around 12% (58,101) of SNPs are found in the extragenic region of the genome. These regions do not code for proteins but harbour critical elements like 5' untranslated regions (UTRs) and 3' UTRs of different genes. There is also a pseudogene in the G-L intergenic region, which is considered redundant (30). The mutations in these regions can significantly impact mRNA stability (31). For instance, the 5' UTR of RABV is known to play a crucial role in RNA stability (32**).** Interestingly, RABV pathogenicity has been reported to increase upon making modifications in non-coding regions (33).

Figure 2 shows the occurrence of different classes of mutations in the four major clades of RABV. The percentage contribution of the SNP-silent class is highest across all the clades, ranging from 76% to 87% per cent. The highest percentage of SNP-silent is seen in the Asian clade, followed by Africa-2. The occurrence of mutations is comparable in Cosmopolitan and Arctic, but deletion frameshift and insertion frameshift are lowest in the Cosmopolitan clade. The differences in the events of mutations across clades imply different selection pressures on the clades, but to get the holistic picture of the differences in evolutionary trajectories, the interactions of the dog host with other hosts must be studied.

Cytosine to Thymine is the Most Common Nucleotide Transition across all RABV Clades

The most common change in the genome of RABV is SNP-silent mutations. Although these changes do not alter the amino acid in the coding protein, they can have an impact on protein translation (34) and mRNA stability (35). Cytosine to Thymine (C>T) is the most common nucleotide change in RABV across clades, leading to subtle variations in the genetic code. However, the order of other nucleotide changes differs across various clades. For instance, in the African and Cosmopolitan clades, T>C represents the second most frequent alteration, while in the Arctic and Asian clades, it is A>G (Figure 3A). This observation highlights clade-specific evolutionary patterns and suggests distinct selection pressure on nucleotides. Most synonymous changes were associated with the amino acid Leucine. Furthermore, our

data demonstrate that these nucleotide-level changes happen majorly in codons encoding hydrophobic and polar amino acids (data not shown).

Our analysis demonstrates heterogeneity in mutation frequencies across the genome, highlighting distinct regions characterised by elevated mutation rates, commonly referred to as 'Hotspots' (Figure 3B). While the majority of genomic regions remain conserved, specific positions exhibit conserved mutational patterns, elucidating a nuanced landscape of mutation dynamics among clades. Remarkably, within the SNP-silent class, a pronounced predilection for mutations targeting the second nucleotide of codons emerges, constituting a substantial 83.8% of the total mutations (330,675 occurrences). Conversely, mutations at the third nucleotide position manifest with lower frequency, accounting for 15.9% (62,887 occurrences) of SNP-silent alterations. Notably, the first position remains devoid of nucleotide changes within the SNP-silent mutation class. Strikingly, multiple nucleotide alterations within a single codon exceed 14,000 occurrences (approximately 3.5% of the SNP-silent class), unveiling intricate mutational patterns. These mutations exhibit specific clade associations, prominently appearing in significant genomes of respective clades. Such distinct mutations and SNP-silent variations may intricately contribute to viral fitness dynamics.

The SNP Mutations are Predominantly Unique and Non-Overlapping in RABV Clades

A total of 38,389 SNPs, accounting for 8% of the total mutations, resulted in non-synonymous amino acid changes referred to as SNP mutations. And the frequency of the occurrence is not the same across the clades. These mutations predominantly arise from single nucleotide variations. Remarkably, the L-protein emerges as the most mutation-prone, displaying a substantial 18,462 amino acid modifications, encompassing approximately 48% of total SNP mutations, despite constituting just over 50% of the genome. Across different clades, the contribution of the L-protein exhibits variability, spanning from 46% to 52%, thereby underscoring clade-specific selection pressures on the gene. The Asian clade displayed the highest percentage, with approximately 52% of the total SNPs occurring in the L-protein, followed by the Cosmopolitan, Arctic, and Africa-2 clades, where the L-protein contributes 46% of the mutations. The G-protein ranks second in terms of total contribution, ranging from 19% to 30% across the clades. The P-protein follows, with a percentage ranging from 13.6% to 17.8%. The N-gene and M-protein exhibit less than 14% contributions in each clade (Figure 4A). It is important to note that the length of M and N differs, which suggests that the N protein is under high negative selection.

Regarding uniqueness, a substantial number of SNP mutations are specific to particular clades. For instance, the Cosmopolitan clade showcases 546 unique mutations, followed by the Asian clade. The Arctic and Africa-2 clades exhibit a comparable number of unique mutations. Among the clades, the Asian and Cosmopolitan share the highest number of mutations, with 85 amino acid changes in common. This is followed by the Arctic and Cosmopolitan clades, which share a considerable number of mutations. Additionally, there are 80 amino acid changes observed across all the clades, highlighting a set of mutations shared universally. The most common nucleotide change observed in SNP mutations occurs at the first and second nucleotide, mostly the first nucleotide (26304 in numbers). The Venn diagram provides an overview of the number of SNP mutation and their distribution across different clades (Figure 4B).

138 Mutations in RABV Have Become Stabilised Across the RABV Clades

A total of 138 mutations were observed, with a frequency of more than 90% across all the clades. These mutations belonged to three classes: SNP mutation class, SNP-silent mutation class, and deletion-frameshift mutation class. Among these, approximately 74% represented the SNP-silent mutation class. Interestingly, the L protein exhibits the highest number of universal mutations, accounting for half of the total variations. The SNP mutation class follows a similar trend, with a majority of mutations observed in the L protein. However, when considering the distribution based on percentages, the G protein and P protein showed a decrease in contribution to the SNP mutation class (Figure 5A&B).

Furthermore, a subset of variants were identified that occur with a frequency of 100% across all the clades. Out of the six variants, three are located in the M protein. Notably, no variant with 100% frequency across all the clades was observed in the L protein. Among the identified variants with 100% frequency, four belonged to the SNP-silent mutation class, and two variants, M_R165Q and P_I108M, resulted in an amino acid change. The variant M_R165Q is primarily caused by a single nucleotide change; however, in several sequences, two nucleotide changes are observed. For instance, all NEE subclade sequences of the cosmopolitan clade and most of the Asian sequences exhibit two nucleotide changes from GG to AA. The variant P_I108M is attributed to a single nucleotide change consistently observed across all the sequences. These mutations were found in all 417 analyzed sequences. In approximately 92.6% (386 sequences) of cases, the R165Q mutation occurs due to a G to A transition at the second nucleotide position of the respective codon. However, in specific sequences, particularly those belonging to the Asian clade (SEA3 (19 sequences) and SEA2a (5 sequences) from China), as well as a few Cosmopolitan clade sequences (three sequences of NEE and one sequence of AF1b), change at positions two and three of the codon lead to this amino acid changes. Additionally, an Africa-2 sequence from Guinea isolated in 1990 shows a unique mutation at position two. The I168M variation arises from a G to A mutation at the second residue (Figure 5C).

Africa-2 Exhibited the Highest Clade-Specific Mutations

The clade-specific mutations are non-overlapping mutations uniquely present in the respective clades with a frequency of at least 90%. In the Africa-2 clade, fifteen clade-specific variants belonging to three classes (SNP-mutation, SNP-silent and Deletion-frameshift) were observed. Four of these variants, N65N, K139K, W354W, and Q745Q, resulted in synonymous amino acid changes in the P, G, M and L proteins, respectively. The only deletion-frameshift clade-specific mutations observed in the dataset were found in M protein, N2 and L4. The Non-synonymous changes were observed in the four proteins N, P, G and L. In the P protein, a single mutation at position 159 (P159L) is observed in 33 out of 36 sequences. The N protein also exhibited a single mutation from Phenylalanine to Tyrosine at position 213 (F213Y). Three non-synonymous mutations were identified in the glycoprotein at positions 14 (F14L), 263 (Q263R), and 484 (N484K). The L protein in the Africa-2 clade showed four non-synonymous amino acid changes at positions E52D, M93I, K1624R, and T1961A. In the Arctic clade, five mutations were observed in the N, G, P and L proteins. Three of these mutations lead to non-synonymous and two to synonymous changes. N protein showed a single SNP-silent mutation, and P and L proteins showed two clade-specific mutations each. In the P protein, a single nucleotide change at nucleotide position 494 (A>C)

resulted in an amino acid change at position 165 (E165D) of the mature protein. A whole codon change (799-802, CCT>GTC) at position 267 (A267G) in the mature protein leads to another amino acid change. The E165D residue lies in variable domain II of the phosphoprotein, which is known to interact with complex I of mitochondria. The G protein of the Arctic clade showed one synonymous (A182A) and one non-synonymous change (V183I). V183I mutation in G protein substituted Valine with Isoleucine. This residue lies in the highly variable snake toxin-like domain, which is involved in binding to the AChR receptor (36). In the Asian clade, three mutations are reported in total, all occurring in the L protein. Two non-synonymous mutations are L424I and N1839E. Additionally, a synonymous change was identified at position 558 (S558S), where a single nucleotide change from adenine to thymine is observed at the third position in the codon (Figure 6).

No clade-specific SNPs were found in the Cosmopolitan clade, possibly due to its high genetic diversity and limitations in the reference sequence.

Discussion

The present study is based on 417 whole genome sequences collected from dogs over a temporal range of 90 years and represents the first comprehensive comparison of genetic diversity at the nucleotide level across different clades of RABV. This study provides valuable insights into the SNP profile across the genome of four RABV clades. The study of SNPs is crucial to understanding various aspects of the virus's biology, such as viral fitness, immune evasion, virulence, and evolution.

The phylogenetic analysis based on the maximum likelihood method showed that the oldest sequences belong to the Bat clade, which lies very close to the outgroup sequences, confirming that RABV in canines came from bats. Four major clades show high diversity within them, ultimately leading to the formation of many subclades. Our study also confirms the previous reports on the presence of multiple clades circulating in a single country. While most of the clades show geographical localization, some incidences of transboundary migration are also observed (37,38). Although dogs do not travel long distances normally, infected dogs are seen to travel larger distances, and their interaction with other hosts could play a crucial role in the spread (39). Therefore, whole genome sequences from other hosts must be analysed to comprehensively understand virus spread and evolution.

In this study, we have identified several types of mutations. Many mutations are not extensively detailed in this study because of their low frequency. However, several of these mutations are found within critical regions of various proteins. Despite the sporadic appearance of these mutations, it remains crucial to monitor them closely and investigate their potential impact on the fitness of the virus. Notably, three clades showed some mutations that are characteristic of the clade and are non-overlapping with any other clade. In addition to playing crucial biological roles, these mutations could be used to track and identify clades. Although further research is needed to validate this approach, similar strategies have demonstrated success in tracking viral lineages in other infectious diseases, such as COVID-19, where specific PCR variants have been employed (40,41). The reason Cosmopolitan did not show any clade-specific mutation could be attributed to the high diversity in the clade.

Diagnosis of rabies has mainly relied on clinical signs and is broadly characterised into furious and paralytic forms. The possible reasons for such distinctions are attributed to several factors, such as the host immune response, vectors, and site of the bite. In addition to these factors, it is also important to study the role of the virus genotype, which could be related to disease outcomes. Current diagnostic technologies are not designed to differentiate between different RABV clades. However, utilising clade-specific mutations could facilitate the development of assays capable of not only identifying the presence of infection but also determining the specific clade involved. This approach would be instrumental in understanding the true burden of different clades within a country and formulating appropriate national and international policies. Additionally, it could aid in tracking the movement of clades across borders.

Vaccination is the only effective way to prevent infection and eradicate the disease. An approach like hauling the hosts is not recommended as it leads to an increase in the immigration of dogs from other areas and later was not proven effective. Vaccinating 70% of the dog population has eradicated dog-mediated rabies from multiple countries. But, in those countries, other hosts have now taken over, such as raccoons, skunks, bats, and raccoon dogs. This has led to the reemergence of the infection in several areas (42,43). Considering the binary nature of disease outcome, the presence of multiple clades, the incidence of vaccine-induced rabies (44) and several cases of post-exposure prophylaxis failure (45,46) highlight the need to evaluate the efficacy of current vaccines for each clade separately. The findings of this study have important implications for public health policies; the discovery of the most common clades in a particular region suggests the priority of that clade or subclade for vaccination and surveillance efforts.

In conclusion, this study not only highlights the significance of these mutations in various aspects of the virus's biology but also emphasises the importance of their application in surveillance and eradication efforts. Continued research and validation of these findings will contribute to a more comprehensive understanding of rabies transmission and aid in developing targeted control measures.

ACKNOWLEDGMENTS

The authors would like the whole UT lab for the discussions and their input.

AUTHOR CONTRIBUTIONS

UT and AK conceptualised the idea. AK curated the data, analysed the data, and wrote the manuscript. MK helped in data analysis and manuscript writing; SS helped in manuscript writing. UT supervised the whole study.

FUNDING

UT acknowledges the DBT-IISc partnership. AK thanks CSIR for financial support. No external funding was obtained to support the project.

DATA AVAILABILITY STATEMENT

The data of this study are available in the manuscript and supplementary file.

CONFLICT OF INTEREST

The authors declare no conflict of interest.

References

1. Černe D, Hostnik P, Toplak I, Presetnik P, Maurer-Wernig J, Kuhar U. Discovery of a novel bat lyssavirus in a Long-fingered bat (Myotis capaccinii) from Slovenia. Gilbert AT, editor. PLoS Negl Trop Dis. 2023 Jun 29;17(6):e0011420.
2. Marston DA, Banyard AC, McElhinney LM, Freuling CM, Finke S, de Lamballerie X, et al. The lyssavirus host-specificity conundrum — rabies virus — the exception not the rule. Curr Opin Virol. 2018 Feb;28:68–73.
3. Rieder M, Conzelmann K-K. Interferon in Rabies Virus Infection. In 2011. p. 91–114.
4. BANYARD AC, TORDO N. Rabies pathogenesis and immunology. Rev Sci Tech l'OIE. 2018 Aug 1;37(2):323–30.
5. Sparkes J, McLeod S, Ballard G, Fleming PJS, Körtner G, Brown WY. Rabies disease dynamics in naïve dog populations in Australia. Prev Vet Med. 2016 Sep;131:127–36.
6. Brunker K, Mollentze N, Brunker@glasgow KA, Uk (K, Brunker). Rabies Virus. Trends Microbiol. 2018 Oct 1;26(10):886–7.
7. Yale G, Sudarshan S, Taj S, Patchimuthu GI, Mangalanathan BV, Belludi AY, et al. Investigation of protective level of rabies antibodies in vaccinated dogs in Chennai, India. Vet Rec Open. 2021 Dec 5;8(1).
8. TU C, FENG Y, WANG Y. Animal rabies in the People's Republic of China. Rev Sci Tech l'OIE. 2018 Aug 1;37(2):519–28.
9. Nagaraja T, Madhusudana S, Desai A. Molecular characterization of the full-length genome of a rabies virus isolate from India. Virus Genes. 2008 Jun 1;36(3):449–59.
10. Morin B, Liang B, Gardner E, Ross RA, Whelan SPJ. An In Vitro RNA Synthesis Assay for Rabies Virus Defines Ribonucleoprotein Interactions Critical for Polymerase Activity. Lyles DS, editor. J Virol. 2017 Jan;91(1).
11. Ogino M, Ito N, Sugiyama M, Ogino T. The Rabies Virus L Protein Catalyzes mRNA Capping with GDP Polyribonucleotidyltransferase Activity. Viruses. 2016 May 21;8(5):144.
12. Morimoto K, Hooper DC, Carbaugh H, Fu ZF, Koprowski H, Dietzschold B. Rabies virus quasispecies: Implications for pathogenesis. Proc Natl Acad Sci. 1998 Mar 17;95(6):3152–6.
13. Okada K, Ito N, Yamaoka S, Masatani T, Ebihara H, Goto H, et al. Roles of the Rabies Virus Phosphoprotein Isoforms in Pathogenesis. Lyles DS, editor. J Virol. 2016 Sep 15;90(18):8226–37.
14. Morimoto K, Shoji Y, Inoue S. Characterization of P gene-deficient rabies virus: Propagation, pathogenicity and antigenicity. Virus Res. 2005 Jul 1;111(1):61–7.
15. Finke S, Conzelmann KK. Replication strategies of rabies virus. Virus Res. 2005 Aug

1;111(2):120–31.

16. Pulmanausahakul R, Li J, Schnell MJ, Dietzschold B. The Glycoprotein and the Matrix Protein of Rabies Virus Affect Pathogenicity by Regulating Viral Replication and Facilitating Cell-to-Cell Spread. J Virol. 2008 Mar;82(5):2330.

17. Mebatsion T, Weiland F, Conzelmann K-K. Matrix Protein of Rabies Virus Is Responsible for the Assembly and Budding of Bullet-Shaped Particles and Interacts with the Transmembrane Spike Glycoprotein G. J Virol. 1999 Jan;73(1):242–50.

18. Badrane H, Tordo N. Host Switching in Lyssavirus History from the Chiroptera to the Carnivora Orders . J Virol. 2001 Sep;75(17):8096–104.

19. de Souza DN, Carnieli P, Macedo CI, de Novaes Oliveira R, de Carvalho Ruthner Batista HB, Rodrigues AC, et al. Phylogenetic analysis of rabies virus isolated from canids in North and Northeast Brazil. Arch Virol. 2017 Jan 26;162(1):71–7.

20. Kissi B, Tordo N, Bourhy H. Genetic Polymorphism in the Rabies Virus Nucleoprotein Gene. Virology. 1995 Jun;209(2):526–37.

21. Wan S, Kumar D, Ilyin V, Al Homsi U, Sher G, Knuth A, et al. The effect of protein mutations on drug binding suggests ensuing personalised drug selection. Sci Rep. 2021 Jun 29;11(1):13452.

22. Campbell K, Gifford RJ, Hill V, Toole O', Hampson K, Brunker &. Making Genomic Surveillance Deliver: A Lineage Classification and Nomenclature System to Inform Rabies Elimination.

23. Katoh K. MAFFT: a novel method for rapid multiple sequence alignment based on fast Fourier transform. Nucleic Acids Res. 2002 Jul 15;30(14):3059–66.

24. Nguyen L-T, Schmidt HA, von Haeseler A, Minh BQ. IQ-TREE: A Fast and Effective Stochastic Algorithm for Estimating Maximum-Likelihood Phylogenies. Mol Biol Evol. 2015 Jan;32(1):268–74.

25. Letunic I, Bork P. Interactive Tree Of Life (iTOL) v5: an online tool for phylogenetic tree display and annotation. Nucleic Acids Res. 2021 Jul 2;49(W1):W293–6.

26. Delcher AL. Fast algorithms for large-scale genome alignment and comparison. Nucleic Acids Res. 2002 Jun 1;30(11):2478–83.

27. Mercatelli D, Giorgi FM. Geographic and Genomic Distribution of SARS-CoV-2 Mutations. Front Microbiol. 2020 Jul 22;11:1800.

28. Kissi B, Tordo N, Bourhy H. Genetic Polymorphism in the Rabies Virus Nucleoprotein Gene. Virology. 1995 Jun 1;209(2):526–37.

29. Nadin-Davis SA, Real LA. Molecular Phylogenetics of the Lyssaviruses—Insights from a Coalescent Approach. In 2011. p. 203–38.

30. Tordo N, Poch O, Ermine A, Keith G, Rougeon F. Walking along the rabies genome: is the large G-L intergenic region a remnant gene? Proc Natl Acad Sci. 1986 Jun;83(11):3914–8.

31. Jia L, Mao Y, Ji Q, Dersh D, Yewdell JW, Qian S-B. Decoding mRNA translatability and stability from the 5′ UTR. Nat Struct Mol Biol. 2020 Sep 27;27(9):814–21.

32. Palusa S, Ndaluka C, Bowen RA, Wilusz CJ, Wilusz J. The 3′ Untranslated Region of the Rabies Virus Glycoprotein mRNA Specifically Interacts with Cellular PCBP2 Protein and Promotes Transcript Stability. Tian B, editor. PLoS One. 2012 Mar 16;7(3):e33561.

33. Virojanapirom P, Yamada K, Khawplod P, Nishizono A, Hemachudha T. Increased pathogenicity of rabies virus due to modification of a non-coding region. Arch Virol. 2016 Nov 24;161(11):3255–61.

34. Rauscher R, Ignatova Z. Timing during translation matters: synonymous mutations in human pathologies influence protein folding and function. Biochem Soc Trans. 2018 Aug 20;46(4):937–44.

35. Chamary J V., Hurst LD. Evidence for selection on synonymous mutations affecting stability of mRNA secondary structure in mammals. Genome Biol. 2005 Aug 16;6(9):1–12.

36. Hueffer K, Khatri S, Rideout S, Harris MB, Papke RL, Stokes C, et al. Rabies virus modifies host behaviour through a snake-toxin like region of its glycoprotein that inhibits neurotransmitter receptors in the CNS. Sci Rep. 2017 Oct 9;7(1):12818.

37. Lushasi K, Hayes S, Ferguson EA, Changalucha J, Cleaveland S, Govella NJ, et al. Reservoir dynamics of rabies in south□east Tanzania and the roles of cross□species transmission and domestic dog vaccination. J Appl Ecol. 2021 Nov;58(11):2673–85.

38. Bannazadeh Baghi H, Alinezhad F, Kuzmin I, Rupprecht C. A Perspective on Rabies in the Middle East—Beyond Neglect. Vet Sci. 2018 Jul 17;5(3):67.

39. Mancy R, Rajeev M, Lugelo A, Brunker K, Cleaveland S, Ferguson EA, et al. Rabies shows how scale of transmission can enable acute infections to persist at low prevalence. Science (80-). 2022 Apr 29;376(6592):512–6.

40. Wang H, Jean S, Eltringham R, Madison J, Snyder P, Tu H, et al. Mutation-Specific SARS-CoV-2 PCR Screen: Rapid and Accurate Detection of Variants of Concern and the Identification of a Newly Emerging Variant with Spike L452R Mutation. J Clin Microbiol. 2021 Aug 1;59(8).

41. Nasir A, Aamir UB, Kanji A, Bukhari AR, Ansar Z, Ghanchi NK, et al. Tracking SARS-CoV-2 variants through pandemic waves using RT-PCR testing in low-resource settings. Dhanasekaran V, editor. PLOS Glob Public Heal. 2023 Jun 1;3(6):e0001896.

42. Botto Nuñez G, Becker DJ, Plowright RK. The emergence of vampire bat rabies in Uruguay within a historical context. Epidemiol Infect. 2019 Apr 22;147:e180.

43. Sharma B, Dhand NK, Timsina N, Ward MP. Reemergence of Rabies in Chhukha District, Bhutan, 2008. Emerg Infect Dis. 2010 Dec;16(12):1925–30.

44. Smreczak M, Orłowska A, Müller T, Freuling CM, Kawiak-Sadurska M, Trębas P. Vaccine-induced rabies in a red fox in Poland. J Vet Res. 2022 Nov 15;66(4):473–7.

45. Quiambao BP, DyTioco HZ, Dizon RM, Crisostomo ME, Laot TM, Teuwen DE. Rabies post-exposure prophylaxis in the Philippines: Health status of patients having received purified equine F(ab′)2 fragment rabies immunoglobulin (favirab). PLoS Negl Trop Dis. 2008 May;2(5).

46. Scholand SJ, Quiambao BP, Rupprecht CE. Time to Revise the WHO Categories for

Severe Rabies Virus Exposures–Category IV? Viruses 2022, Vol 14, Page 1111. 2022 May 22;14(5):1111.

Figures

Figure 1

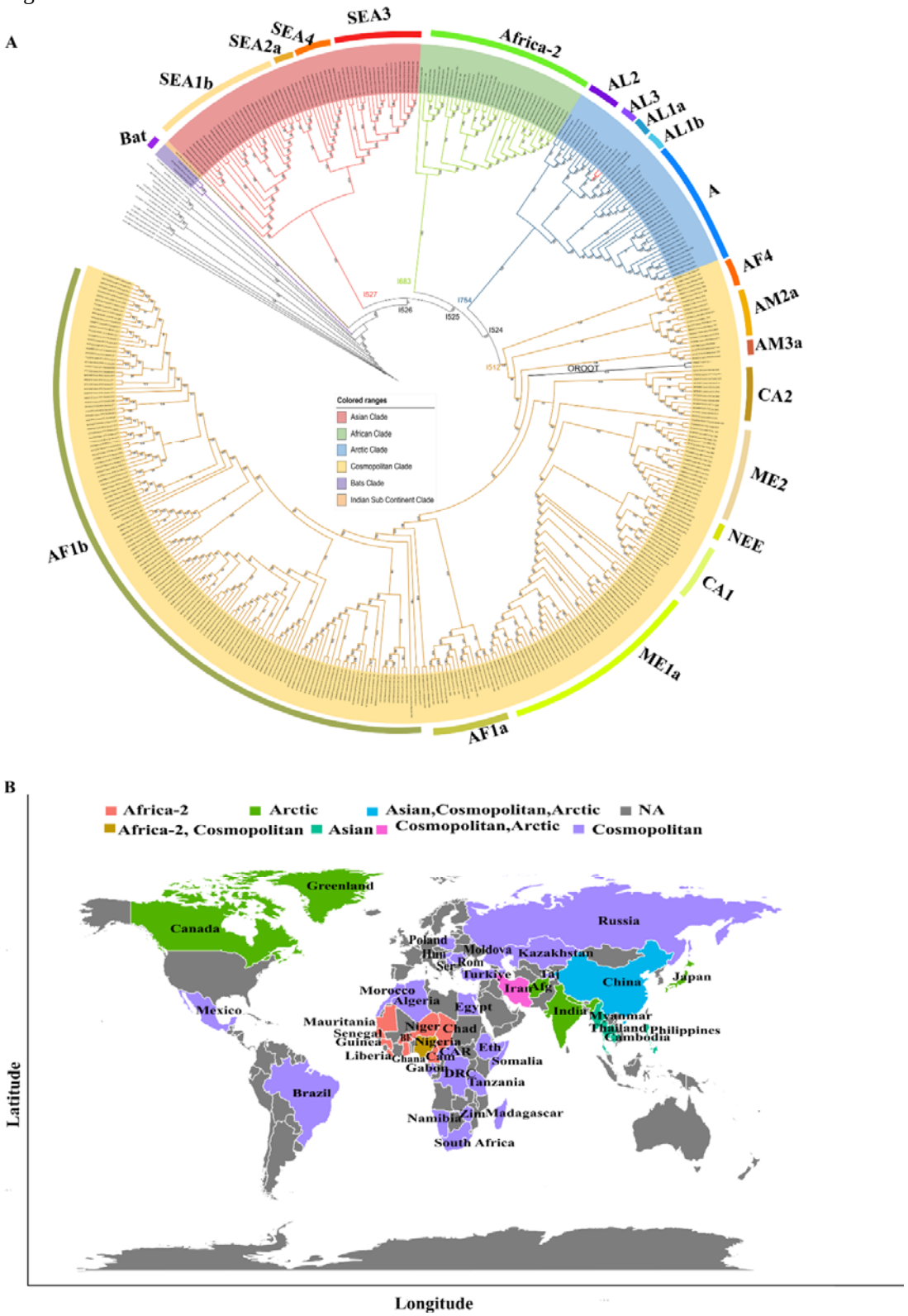


Figure 2

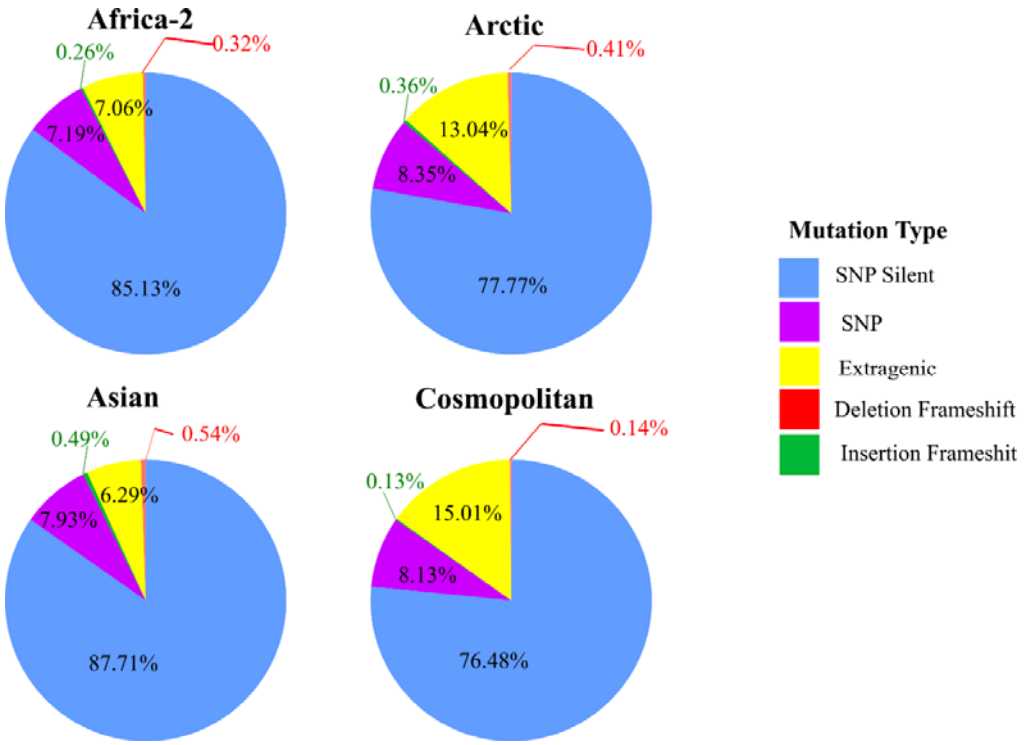

Africa-2
0.26%
0.32%
7.06%
7.19%
85.13%
Arctic
0.41%
0.36%
13.04%
8.35%
77.77%
Asian
0.49%
0.54%
6.29%
7.93%
87.71%
Cosmopolitan
0.14%
0.13%
15.01%
8.13%
76.48%
Mutation Type
SNP Silent
SNP
Extragenic
Deletion Frameshift
Insertion Frameshit

Figure 3

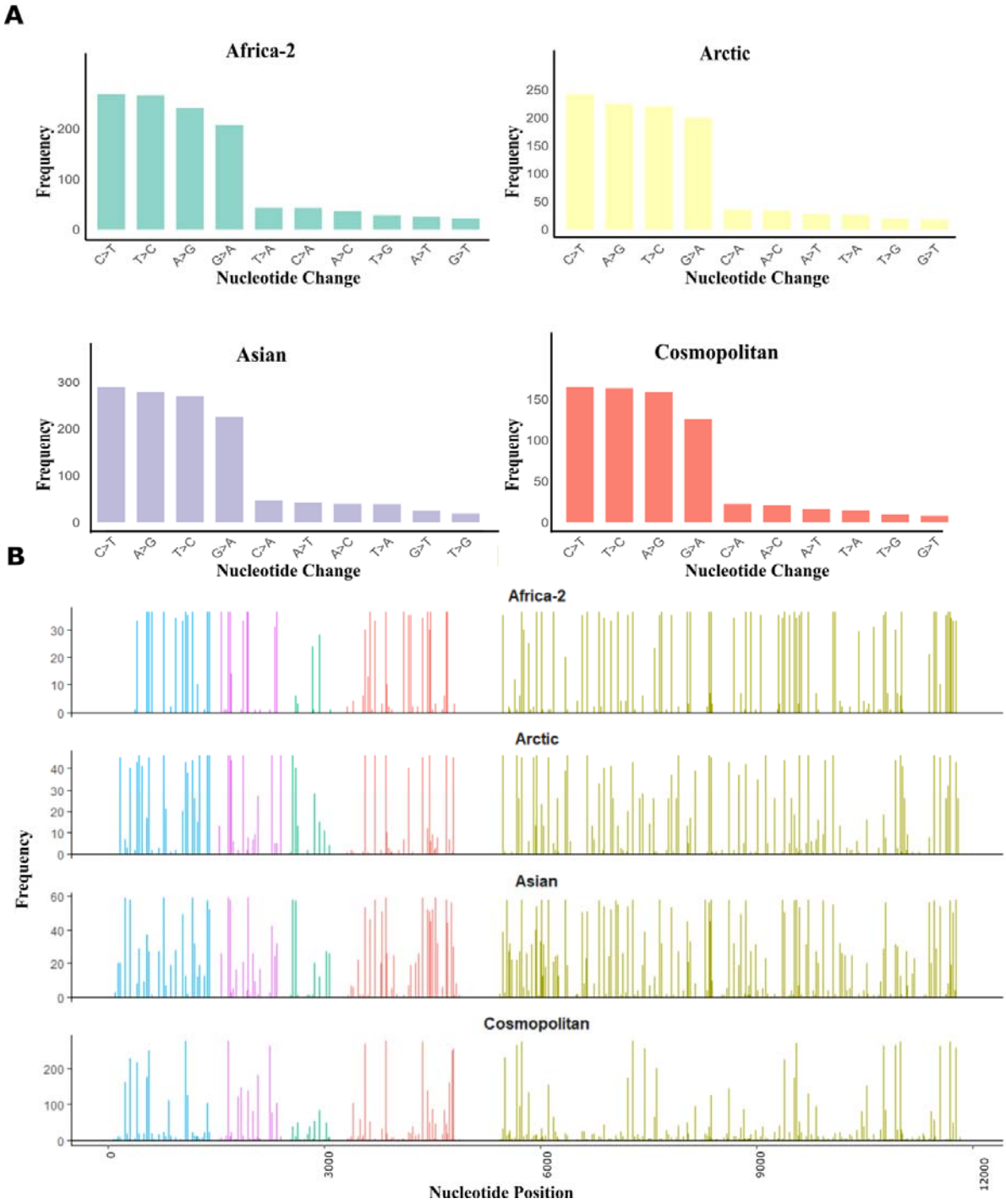

A
Africa-2
Arctic
Asian
Cosmopolitan
Frequency
Nucleotide Change
C>T
T>C
A>G
G>A
T>A
C>A
A>C
T>G
A>T
G>T
B
Africa-2
Arctic
Asian
Cosmopolitan
Frequency
Nucleotide Position
0
3000
6000
9000
12000

Figure 4

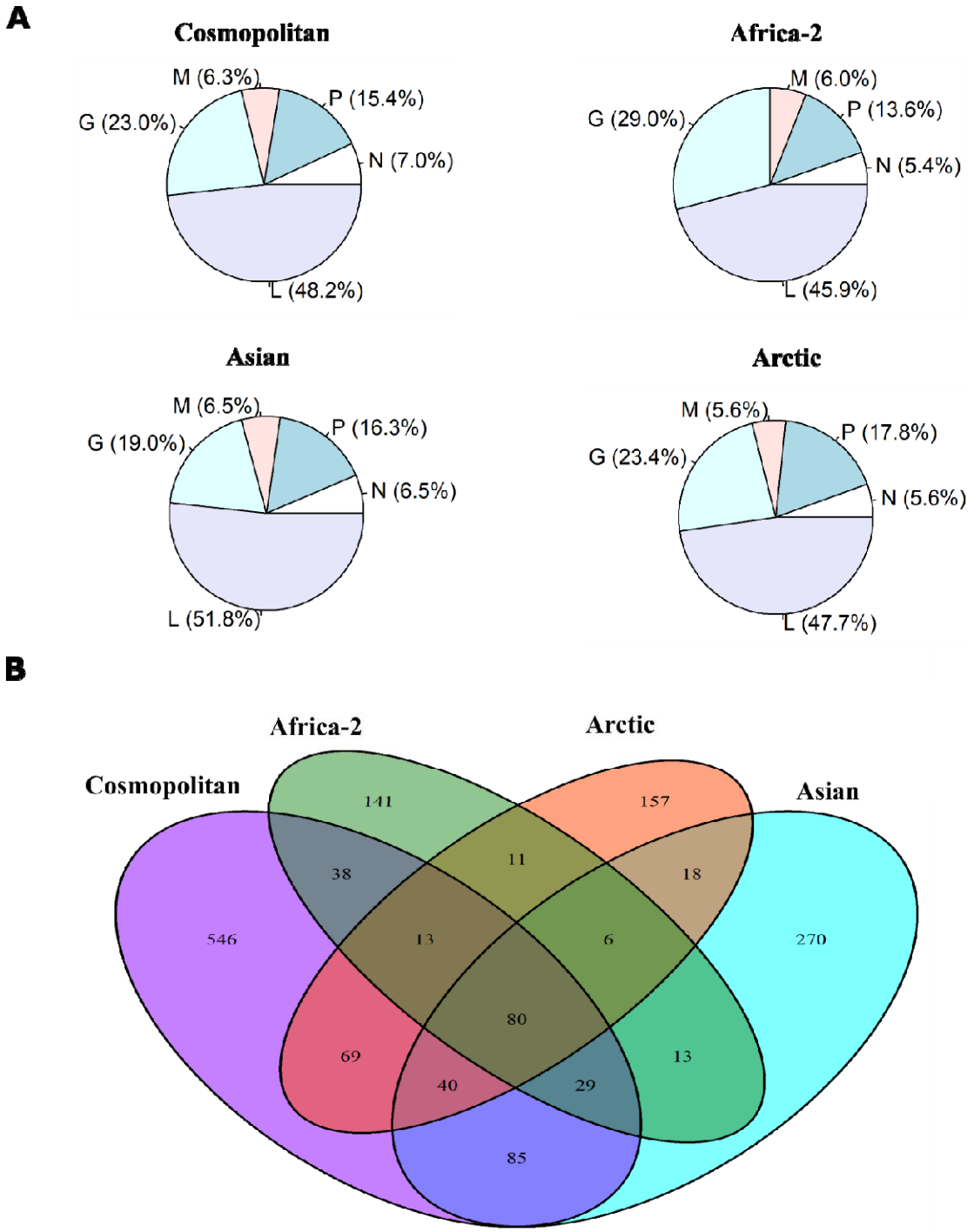

A
Cosmopolitan
M (6.3%)
P (15.4%)
G (23.0%)
N (7.0%)
L (48.2%)
Africa-2
M (6.0%)
P (13.6%)
G (29.0%)
N (5.4%)
L (45.9%)
Asian
M (6.5%)
P (16.3%)
G (19.0%)
N (6.5%)
L (51.8%)
Arctic
M (5.6%)
P (17.8%)
G (23.4%)
N (5.6%)
L (47.7%)
B
Africa-2
Arctic
Cosmopolitan
Asian
141
157
38
11
18
546
13
6
270
80
69
13
40
29
85

Figure 5

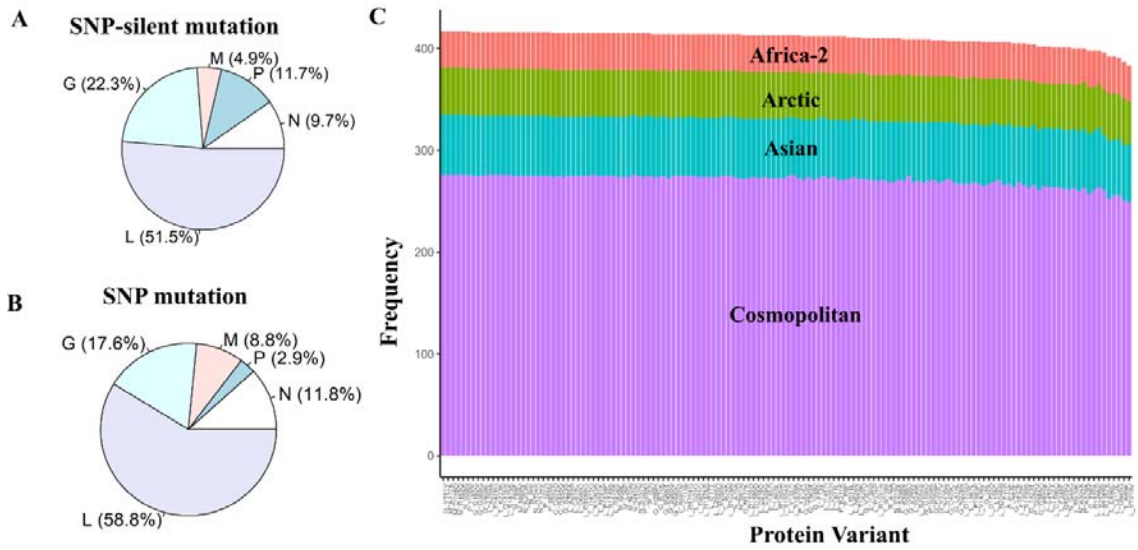
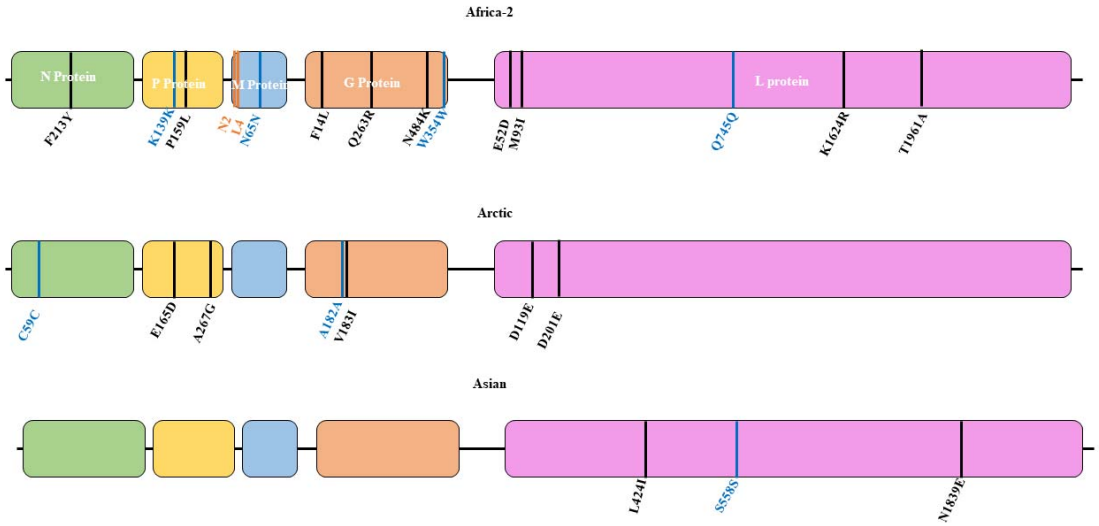


Figure 6



Legends:

Figure 1: Maximum-Likelihood Phylogenetic Tree of Dog-Origin RABV Sequences

(A) A maximum-likelihood-based phylogenetic tree illustrates the relationship of various clades of RABV isolated from dog hosts to each other, unveiling the emergence and interconnections of distinct clades. We employed sixteen recognised species from the Lyssavirus genus as an outgroup to elucidate their relationship to the RABV virus. Notably, major clades are vividly distinguished by varying colours, as denoted in the accompanying legend. These clades have further evolved into subclades, indicative of evolutionary divergence and are highlighted with distinct coloured arcs. Some subclades are represented by a few genomes, like SEA1a (1 genome), and AF1c (2 genomes) are not highlighted. Notably, only two sequences are documented from India, underscored by their representation of the leaves in red, and these two sequences belong to the Arctic AL1a subclade.

(B) This world map provides a depiction of the global distribution of RABV major clades, revealing simultaneous occurrences of multiple major clades across diverse regions. Notable nations such as China, Iran, Russia, and Nigeria showcase the co-occurrence of several major clades, emphasising the complex epidemiological dynamics of RABV. This geographical representation offers valuable insights into the widespread prevalence and diversity of RABV subclades, underscoring the significance of global surveillance and coordination in rabies control efforts.

Figure 2. The Pie Diagrams Show the Occurrence of Different Classes of Mutation.

The pie diagrams represent the prevalence of various mutation classes in different clades as percentages. Each mutation class is uniquely identified by distinct colours, as indicated in the legend. The dominant mutation class, SNP-silent, is depicted in blue, symbolising its high prevalence across all clades. Notably, the proportion of extragenic mutations, represented by the yellow-coloured pie, is higher in the Cosmopolitan and Arctic clades compared to the Africa-2 and Asian clades. The SNP mutation class, illustrated in purple, exhibits a range of 6 to 8 per cent across the different clades. Intriguingly, deletion frameshift (red colour) and insertion frameshift (green colour) mutation classes occur at significantly lower rates across all clades, with the Cosmopolitan clade exhibiting the lowest frequency among all. This comparative analysis highlights the variation in mutation class distributions across the different clades, shedding light on potential factors influencing the mutational landscape in different geographic and genetic contexts.

Figure 3: Most Common Nucleotide Changes and Hotspot Regions in the SNP-Silent Class

(A) This section displays the top ten prominent nucleotide changes in four RABV clades and is colour-coded accordingly. The nucleotide changes are arranged in descending order of frequency observed across the various clades, providing a clear understanding of the prevalent variations. C>T is consistently the most prevalent mutation across all the clades of RABV. T > C change is the second most prevalent change in Africa-2 and Cosmopolitan, while A>G is in the Arctic and the Asian clades.

(B) Here, we showcase the genomic hotspot regions exhibiting frequent nucleotide changes across distinct clades of the RABV genome. The x-axis represents genome positions, and the y-axis represents the frequency of occurrence. A specific colour for easy identification represents each gene: blue designates the N gene, purple signifies the P gene, green represents the M gene, red highlights the G gene, and dark yellow is used for the L gene. This colour-coded scheme enables efficient visualisation and analysis of the distribution of nucleotide changes within the different genes across the genome of RABV.

Figure 4: Occurrence of SNP Mutations within and Across RABV Clades

(A) This panel illustrates the percentage of SNP mutations occurring in distinct genes across the various Rabies clades. The percentage of mutations for each gene is provided within parentheses and corresponds to the respective protein. This data offers a comprehensive view of the occurrence of the gene-specific SNP mutation for each clade. The percentage of the mutation class is highest in the L protein. The values range from 45% (Africa-2) to 52% (Asian). The N protein and M protein show similar percentages in the Asian and Arctic clade, while in the Africa-2, the percentage of SNP mutation of N protein (5.4%) is less than M protein (6.0%) and in Cosmopolitan, it is the opposite. In the G protein, the mutation percentage ranged from 19% to 29% for the Asian and Africa-2, respectively.

(B) The Venn diagram depicts the instances of SNPs in numbers in the clades of RABV. This visual representation highlights the clade-specific nature of most proteins, underscoring the genetic diversity and lineage-specific mutation patterns observed in Rabies. The purple-shaded section, representing the Cosmopolitan clade, stands out with 546 unique SNP mutations. Sixty-nine mutations are shared with the Arctic clade (orange), whereas 40 mutations are common to all three clades: Cosmopolitan, Arctic, and Asian (blue). Notably, 85 mutations are shared between the Cosmopolitan and Asian clades, and 29 mutations are observed among the Cosmopolitan, Asian, and Africa-2 (green) clades. Moreover, 80 mutations are shared across all four clades. Additionally, 13 mutations are common to the Africa-2, Arctic, and Cosmopolitan clades, while 38 mutations are shared between the Cosmopolitan and Africa-2 clades. The Africa-2 clade possesses 141 mutations unique to its sequence, while 11 mutations are shared with the Arctic clade and six mutations with both the Arctic and Asian clades. Additionally, 13 mutations are shared between the Africa-2 and Asian clades and 18 between the Asian and Arctic clades. The Arctic clade has 157 unique mutations, and the Asian clade showcases 270 mutations exclusively unique to its sequence.

Figure 5: Occurrence of Universal Mutations in Different Proteins and Frequency Distribution in the Clades

(A) This panel illustrates the distribution of Universal SNP-silent mutations among various proteins, with the respective percentages indicated within parentheses. The L protein showed the highest number of Universal SNP-silent mutations, followed by G, P, N and M.

(B) This panel highlights the occurrence of universal SNP mutations in RABV. The L protein showed the highest mutations, followed by G, N, M and P. The P constituted only 3% of the Universal SNP mutation class.

(C) The bar graph represents different universal mutations in RABV. The x-axis of the graph represents the variant, and the y-axis represents the cumulative frequency of occurrence of a mutation. A specific colour visually distinguishes each clade: purple corresponds to Cosmopolitan (n = 268), (G_V181V, M_A177A, M_D146D, N_S58S, P_I108M)blue signifies Asian (n = 58), green represents Arctic (n = 46), and red designates Africa-2 (n = 36). The mutations are arranged in decreasing order of occurrence along the x-axis. There are six mutations (G_V181V, M_A177A, M_D146D, N_S58S, P_I108M) that are present with a frequency of 100% across all the clades.

Figure 6: Clade-Specific Mutations within Three Major RABV Clades.

This figure depicts clade-specific mutations within the three major RABV clades. The five RABV proteins are visually represented as rectangles, each assigned a distinct colour for easy identification: green (N protein), yellow (P protein), blue (M protein), orange (G protein), and pink (L protein). Within these protein rectangles, bars represent the three classes of mutations. SNP mutations, indicated by black bars, involve changes in the DNA sequence that result in altered amino acids in the encoded protein. SNP-silent mutations are represented by blue bars that don't lead to amino acid alterations in the protein sequence. Deletion Frameshift-displayed as orange bars, deletion frameshifts are a type of mutation that involves the removal of a segment of DNA, potentially leading to a shift in the reading frame during translation.

Supporting information

Supplementary file 1. Metadata with clade information for 421 canine whole genome sequences.

Supplementary file 2. A modified R script for mutational analysis.

Supplementary file 3. Results with all the SNPs found in RABV whole genome sequences belonging to four clades.