Research Article

Full-Length Genome of an Africa-4 Lineage Wild-Type *Lyssavirus rabies* from a Stray Dog in Egypt, 2019

Amthal Ahmed Fouad[1], Basem Mohamed Ahmed[2], Momtaz Abdelhady Shahein[1], Hussein Aly Hussein[2*]

[1]*Department of Virology, Animal Health Research Institute, Agriculture Research Centre, Giza 12618, Egypt;*
[2]*Department of Virology, Faculty of Veterinary Medicine, Cairo University, Giza 12211, Egypt.*

Abstract | Background: Lyssavirus rabies is a major global zoonosis and is endemic in Egypt. Rabid dogs are the prime source for human and livestock *Lyssavirus rabies* (RABV) exposure with almost 100% case fatality. While all lyssavirus rabies genetic data -including whole genomes- from Egypt are clustered within a unique lineage (Africa-4 lineage of the cosmopolitan clade), no report provided detailed description of this lineage genome and the last published whole genome dates back to 2009. In this context, a complete RABV genome (5EG-QH19) was obtained from a stray dog, analyzed, and compared to other RABV sequences to update the present knowledge. Methods: Rapid immune-chromatography and direct fluorescent antibody test (DFAT) were applied to detect RABV antigens in the brain of rabid dog. Relevant reads from the Ion S5 sequencing system were assembled into a draft genome. The draft genome was analyzed, re-assembled, and compared to representative RABV genomes from Africa and the Middle East region. Results: The complete genome of 5EG-QH19 strain extends for 11919 nt including the standard RABV protein genes in the correct order. Twenty-nine synonymous and five non-synonymous mutations were evident in the protein genes. Deletions and insertions were observed in the intergenic spaces. It was highly similar to Egyptian RABV genomes and clustered within the Africa-4 (AF4) lineage of the cosmopolitan clade. It was submitted to NCBI GenBank under the access number OL314495. Conclusion: The study provides analysis for the 5EG-QH19 wild-type RABV genome, elucidates some key structural features of the AF4 lineage dominant in Egypt and supports the recently proposed division of AF4 lineage into two sub-lineages (AF4a and AF4b). Divergence from SAD B19 vaccine strain sequence was also reported but the exact effect of the observed genetic changes in an infectious RABV particle dictates in-vivo studies.

Keywords | *Lyssavirus rabies*, wild-type, genome, analysis, Africa 4 lineage, Egypt

INTRODUCTION

Lyssavirus rabies (RABV), is one of the 17 viral species within *genus Lyssavirus*, subfamily *Alpharhabdovirinae* of *family* rhabdoviridae (Walker et al., 2022). It is the etiologic agent of rabies which is an acute progressive almost fatal encephalomyelitis and is the deadliest lyssavirus to humans and warm-blooded animals. Bats are the reservoir of all known lyssaviruses, but dogs and wild canids are

the main source for susceptible hosts' exposure to RABV. RABV genome is about 12Kb negative sense monopartite RNA with 5 protein genes arranged in a fixed order, nucleoprotein (N)- phosphoprotein (P)- matrix protein (M)- glycoprotein (G)- polymerase (L), The genome also features a 58nt leader, 70nt tailer and 4 intergenic spaces. The last intergenic space, G-L, is the longest (400+ nt) (Kuzmin et al., 2008). The presence of single polyadenylation signal in the G-L intergenic space identifies wild type

viruses from Pasteur vaccine strains which are characterized by 2 polyadenylation sites but increased G-L length is believed to affect transcription efficiency (Mochizuki et al., 2009).

Dog bites account for more than 50 human deaths per year while the annual total number of bites exceeded 350000 with the socioeconomic burden for post-exposure prophylaxis and vaccination (Shwiff et al., 2013; WHO, 2014). Multiple reports detected rabies from livestock in Egypt with almost 100% case fatality (Botros and Moch, 1976; Abd El Rahman, 2015; Sultan et al., 2021).

Genomic studies identified 6 major clades of the dog related RABV with the cosmopolitan clade is widely spread via multiple lineages and sub-lineages in the old as well as new world countries (Dellicour et al., 2019; Marston et al., 2013; Brunker et al., 2020; David et al., 2007). Most clades/ lineages exerted a strong purifying non diversifying selective pressure and strong structural functional stability of RABV protein genes and regulatory sequences (Shwiff et al., 2013; Mochizuki et al., 2009), and the majority of RABV genome nucleotide changes are synonymous (Brunker et al., 2018). Studies involved complete RABV genomes from Egypt clustered Egyptian RABV genomes in the cosmopolitan Africa-4 (AF4) lineage and included no more than 3 sequences with no data about the existence of other clades/ lineages in Egypt (Fischer et al., 2018; Dellicour et al., 2019; David et al., 2007; Marston et al., 2013).

In this study, we report the analysis of a complete wild-type RABV genome (5EG-QH19), obtained directly from an Egyptian stray dog. Representative RABV genomes (n=60) from African and Middle Eastern clades/ lineages were included in the genomic analysis whilst all complete Egyptian RABV genomes (n=8) available in GenBank were included in the genetic analysis.

MATERIALS AND METHODS

ETHICS STATEMENT

The authors confirm that the proper ethical review committee approval has been received. Sampling and handling procedures were approved via Cairo University- Institutional animal care and use committee (CU-IACUC); approval document number: CU-II-F-C-17-21. Team members who collected and examined the sample(s) were fully vaccinated against *Lyssavirus rabies* infection according to WHO guidance.

SAMPLE COLLECTION AND PREPARATION

A whole native dog head with a history of violence and excessive biting incidents against humans and animals was submitted to the Department of Virology, AHRI for rabies virus detection. The brain tissue sample from this dog was prepared according to world organization of animal health (WOAH) regulations, part of the sample was homogenized to perform immunochromatography and molecular characterization. The other part was prepared as an impression smear from hippocampus and cerebellum for direct fluorescent antibody technique (DFAT).

SCREENING BY IMMUNOCHROMATOGRAPHY RAPID TEST AND DFAT

Homogenized brain sample was tested for the presence of *Lyssavirus rabies* antigen using rapid rabies antigen test kit (Shenzhen Zhenrui Biotechnology Co., Ltd., China) according to manufacturer's instructions. Briefly, part of the prepared sample was mixed well with kit sample buffer, 5 drops were loaded slowly to the sample hole in the detection device and read after 15-20 minutes. Appearance of a wine-red color on the device T-line was considered positive.

For direct fluorescent antibody test, impression smear from cerebellum and hippocampus were fixed in 80% acetone for 4 hours at -20°C, air dried and either stored at -70°C or brought to room temperature and stained with FITC Anti Rabies Monoclonal Globulin conjugate (FUJIRE-BIO Europe N.V., Belgium) according to manufacturer's instructions and examined under fluorescent microscope (WOAH, 2022).

RNA EXTRACTION AND DIRECT WHOLE GENOME SEQUENCING

Viral RNA was extracted using QIAamp Viral RNA Mini Kit (Qiagen, GMBH) according to manufacturer's instruction. Fresh RNA was quantified using SPECTRO-star Nano microplate reader (BMG LABTECH, Germany) and then was sent to Colors lab (Maadi, Egypt) for direct sequencing and assembly. Kits and software used are summarized in (Tables 1, 2).

SEQUENCE ANALYSIS

Draft genome was BLAST analyzed, then it was re-assembled. Final assembly, ORFs detection, nucleotide as well as deduced amino acid identities were obtained using BioEdit software version 7.0.5.3 (Hall 1999). Sixty whole genomes representing RABV clades, lineages, and sub-lineages reported from Middle Eastern and African countries were retrieved from NCBI nucleotide database to generate phylogenetic trees, sequence accession numbers and other details are provided in Table (S1). Phylogenetic analysis for the assembled rabies genome as well as separate protein gene trees were created using maximum likelihood algorithm of MEGA X version 11.0.10 (Tamura et al., 2021).

Table 1: Different kits and devices used to obtain whole genome reads from sample RNA.

No.	Step	Kit
1	Measure concentration of RNA (QC Step)	Qubit™ RNA HS Assay Kit
2	Removal of ribosomal RNA	RiboMinus™ Eukaryote Kit v2
3	Verify rRNA depletion (QC Step)	Agilent RNA Nano kit
4	RNA library preparation	Ion Total RNA-Seq Kit v2
5	Measure concentration of the amplified cDNA (QC Step)	Qubit™ DNA HS Assay Kit
6	Assess the yield and size distribution of the amplified cDNA (QC Step)	Agilent High Sensitivity DNA kit
7	Template preparation, loading chips, and sequencing	OT2-ES System
8		Ion S5 Device

Table 2: different software used to check, assemble reads, and construct whole genome.

No.	Step	versions
1	QC step to ensure the quality of produced read and prepare reads for assembly (FASTQC, MultiQC)	FASTQC: 0.11.9 MultiQC: 1.11
2	Reference-based assembly (spades, Velvet, BWA)	SPAdes: 3.15.3 Velvet: 1.1 BWA: 0.7.17
3	QC step to ensure the quality of assembly (QUAST)	QUAST: 5.1
4	Reference-based draft genome construction (Ragoo)	RaGoo: 1.1

RESULTS

INITIAL SCREENING BY IMMUNOCHROMATOGRAPHY AND DFAT

The obtained sample developed a wine-red color at the T line of rapid rabies antigen test device indicating presence of rabies virus antigen in the sample which was further confirmed by the positive result after direct fluorescent antibody test (DFAT) (Figures 1a, b).

WHOLE *LYSSAVIRUS RABIES* GENOME

The assembled genome of the 5EG-QH19 sample was 11919 nt long and followed the same organization of other rabies viruses with 5 genes arranged N-P-M-G-L in addition to leader, tailer and 4 intergenic spaces. The genome was 9 bases shorter than the SAD B19 vaccine strain and 5 to 4 bases shorter than other *Lyssavirus rabies* genomes published from Egypt. These bases account for 7 lost bases from the start of the leader sequence, one deletion in

the N-P intergenic space and one insertion in the G-L intergenic space compared to the longest G-L intergenic space recorded in AF4 genomes. The GC content of the obtained genome was 45.21% compared to 45.06% for the SAD B19 vaccine strain and a range of 45.28 to 45.46 for other Egyptian whole genomes. Assembled 5EG-QH19 RABV genome was submitted to NCBI GenBank under the access number OL314495. Genome components lengths are presented in (Table 3) compared to the SAD B19 and other Egyptian RABV genomes.

Figure 1: A) positive *Lyssavirus rabies* detection using rapid immunochromatography. B). positive Direct fluorescence indicative for RABV antigen in brain homogenate.

NUCLEOTIDE, DEDUCED AMINO ACID IDENTITIES AND MUTATIONS

The 5EG-QH19 genome was 89.7% identical to SAD B19 nucleotide sequence, while it scored identities between 96.7% with the 86092EGY sequence and 99.2% with RV2322 sequence (Table 3). Different protein gene sequences showed different identity values, where higher identity values were linked to deduced amino acid sequences. Detailed identities are presented in (Table 4) for protein genes, leader, tailer, and G-L intergenic space.
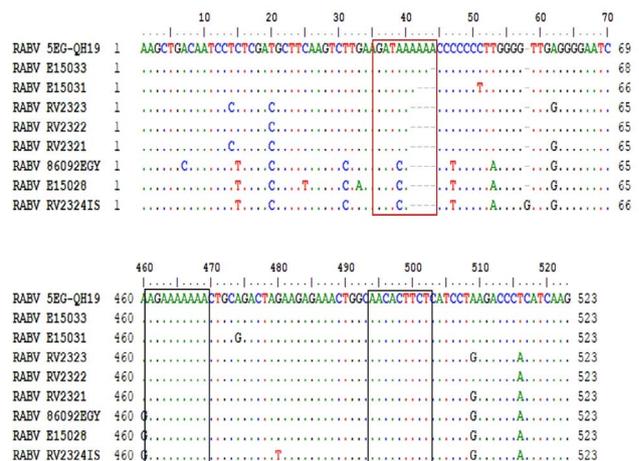
Figure 2: G-L intergenic space nucleotide features. The red box identifies the variable region where adenine (A) residues are shown to be inserted progressively. Black boxes identify the G-L TTP (G mRNA termination polyadenylation signal (left) and L mRNA transcription initiation signal (right)).

Table 3: different genome components of the study genome compared to SAD B19 RABV vaccine sequence and AF4 lineage sequences.

Genome	Length	GC content	year	host	leader	N	N-P	P	P-M	M	M-G	G	G-L	L	Tailer
5EG-QH19	11919	0.4521	2019	dog	63	1353	89	894	88	609	211	1575	522	6384	131
SADB19†	11928	0.4506	ND	vaccine	70	1353	90	894	88	609	212	1575	522	6384	131
RV2323	11923	0.4544	1999	dog	70	1353	90	894	88	609	211	1575	518	6384	131
RV2322	11923	0.4528	1998	dog	70	1353	90	894	88	609	211	1575	518	6384	131
RV2321	11923	0.4546	1998	dog	70	1353	90	894	88	609	211	1575	518	6384	131
86092EGY	11923	0.4537	1979	human	70	1353	90	894	88	609	211	1575	518	6384	131
RV2324IS	11924	0.454	1950	dog	70	1353	90	894	88	609	211	1575	519	6384	131
Complete coding RABV sequences from Egypt lacking leader and tailer sequences															
E15033	11725	0.4548	2009	dog	NA	1353	90	894	88	609	211	1575	521	6384	NA
E15031	11723	0.4551	2009	donkey	NA	1353	90	894	88	609	211	1575	519	6384	NA
E15028	11722	0.4547	2007	dog	NA	1353	90	894	88	609	211	1575	518	6384	NA

† *Lyssavirus rabies* vaccine from vaccine lineage of the cosmopolitan clade.

Table 4: nucleotide and deduced amino acid identities of the study sequence (5EG-QH19) compared to SAD B19 vaccine strain and other Egyptian RABV strains from 1950-2009.

5EG-QH19 RABV														
Identity percent	Whole genome	N		P		M		G		L		Leader	Tailer	G-L
		Nt	aa	Nt	aa	Nt	Aa	Nt	Aa	Nt	Aa			
RABV SADB19	89.7	91.7	98.2	88.9	92.2	89.8	92	91.1	93.8	90.4	96.5	87.3	80.1	77.5
RABV RV2323	98.4	98.7	100	99.4	99.3	98.8	99	99.4	99.4	99.3	99.9	98.4	98.4	98.8
RABV RV2322	99.2	99	100	98.9	99.3	98.8	99	99.5	99.6	99.3	99.9	100	97.7	98
RABV RV2321	98.4	96.7	100	95.9	96.9	97.3	99	96.3	98.6	96.7	99.4	100	99.2	90.6
RABV 86092EGY	96.7	98.8	100	98.2	98.6	98.1	98.5	98.7	99.4	98.5	99.6	96.8	94.6	96.1
RABV RV2324IS	96.9	98.8	99.5	99.2	98.9	99.5	99	99.4	99.4	99.3	99.8	98.4	94.6	97.8
RABV E15033	NA	98.9	100	98.2	98.6	98.1	98.5	98.7	99.4	98.4	99.6	NA	NA	96.1
RABV E15031	NA	96.6	99.7	95.9	96.9	97.5	98.5	97.1	98.8	97.2	99.5	NA	NA	91.5
RABV E15028	NA	96.9	99.3	96	96.2	97.7	98	97.2	98	97.4	99.3	NA	NA	91.9

RABV: Rabies virus, Nt: nucleotide, aa: amino acid, NA: not available.

Intergenic spaces also showed variable identity values.

The M-G intergenic space had the highest conservation ratio between all untranslated regions in the study sequence. It was 91.9% identical to the SAD B19 sequence, and it was 99 to 99.5% identical to other Egyptian RABV Sequences. Contrarily, the G-L intergenic space had the highest variation in length and identity (Table 4). The 5EG-QH19 G-L sequence had the same length as that of the SAD B19 G-L sequence, but it showed only 77.5% identity. The same sequence was 90.6 to 98.8% identical to other Egyptian RABV sequences. An extra adenine (A) residue (6As) compared to 5As or 3As in 2009 RABV sequences from Egypt (E15033 and E15031) in the G-L intergenic space was observed. One TTP, 470 nt from G stop codon, is conserved between all sequences (Mochizuki et al., 2009). It should be mentioned also that it seems to be a progressive feature of adding A residues at this region

(Figure 2).

The N-P and P-M intergenic spaces scored almost similar values. They were 84.8% and 81.8% identical to the SAD B19 sequence, respectively. The N-P intergenic space scored 92.2 to 96.6% identity while the P-M intergenic had 93.1 to 98.8% identity to other Egyptian sequences, respectively.

Transcription initiation/ poly adenylation signals were mostly conserved within the 5EG-QH19 genome, SAD B19 vaccine strain and members of the AF4 clade. Initiation signals followed the previously published RABV consensus sequence of AACAYYHCT (Conzelmann et al., 1990; Mochizuki et al., 2009). Termination/ poly adenylation signals were also conserved and followed the consensus TG(A)7 except N gene signal, TG(A)6, which lacked the seventh Adenine residue and G gene signal which was AG(A)7 (Figure 3).
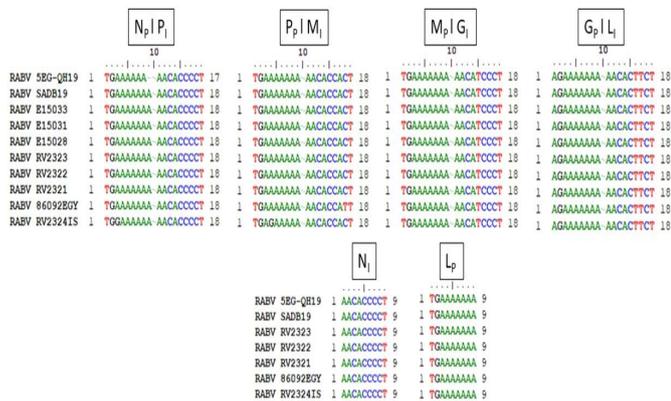

Figure 3: Transcription initiation and termination polyadenylation signals of 5EG-QH19 compared to SAD B19 and other genomes from AF4 lineage. Genes are presented as single letters (N-P-M-G-L). Subscript I stand for initiation while subscript P stand for polyadenylation.

A total of 34 mutations were observed in the genes of the 5EG-QH19 genome and were distributed as N (3), P (5), M (3), G (3), and L (20). Only 5 of these mutations changed their respective amino acids, 2 in P, 2 in M, and one in G protein.

**STRUCTURAL FEATURES OF THE 5EG-QH19 GENOME**

The N protein of the 5EG-QH19 RABV, it was the most conserved gene (Mochizuki et al. 2009; Jiao et al., 2011; Oem et al., 2013) with only 3 synonymous nucleotide changes; G675A, C795T, and C885T. B-cell epitope (antigenic site I) and T-cell epitope (antigenic site IV) (Goto et al., 2000) as well as the putative casein type phosphorylation site (389S) (Faber et al., 2004) were all conserved for the 5EG-QH19 and also for the lineage (Figure 4).

The P protein showed 3 synonymous mutations (C84T, T147C, A828G) besides two non-synonymous mutations; A200G which changed the neutral non-polar glycine into acidic polar aspartate D67G in the variable domain 1 (VD1-positions 61-80) and G418T which changed either the neutral nonpolar phenylalanine or the neutral polar serine into the neutral nonpolar valine V140F/S in VD2 (positions 134-180) which is located in the N binding site (69-177 and 268-297). VD2 also contained the dynein-tail light chain (LC8) interaction motif [(K/R)XTQT] which plays a key role in the axoplasmic transport of RABV nucleocapsid and it was expressed as KSTQT in the 5EG-QH19 positions (144-148). The previously described conserved domains CD1 (positions 1-50) containing L binding site (positions 1-19) (Delmas et al., 2008) and CD2 (positions 201-245) (Nadin-Davis et al., 1997, 2002) were found conserved in the study sequence and within AF4 lineage. P phosphorylation residues (162S, 210S, 271S) were highly conserved. Multiple methionine residues responsible for translation initiation and P protein polymorphism (20M, 53M, 83M) were conserved but the previously reported 69M residue (Kuzmin et al., 2008; Mochizuki et al., 2009) was replaced by nonpolar valine 69V in all sequences of the AF4 lineage (Figure 5).
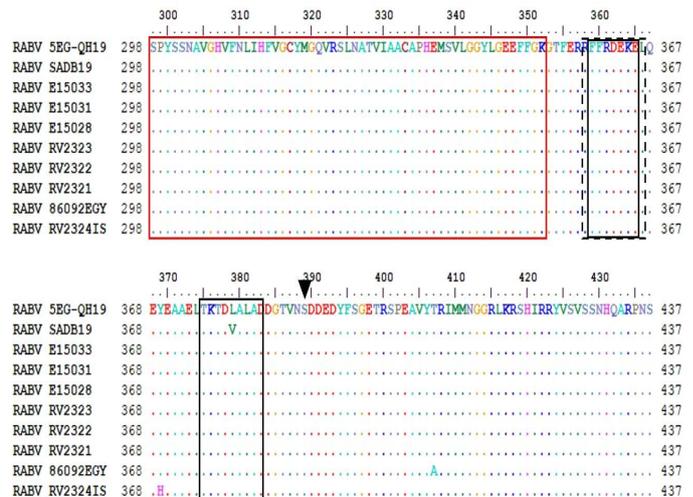
Figure 4: amino acid features of N. red box identify RNA binding domain. Black dashed box identifies antigenic site I. black solid boxes identify antigenic site IV. Black arrowhead for the 389S phosphorylation residue.

The M protein connects G cytoplasmic domain with the RNP and stabilizes viral envelope (Jiao et al., 2011), its involved in the regulation of RNA synthesis and viral budding. The M protein of the 5EG-QH19 genome one synonymous A42G and two non-synonymous A39T and G239A mutations which
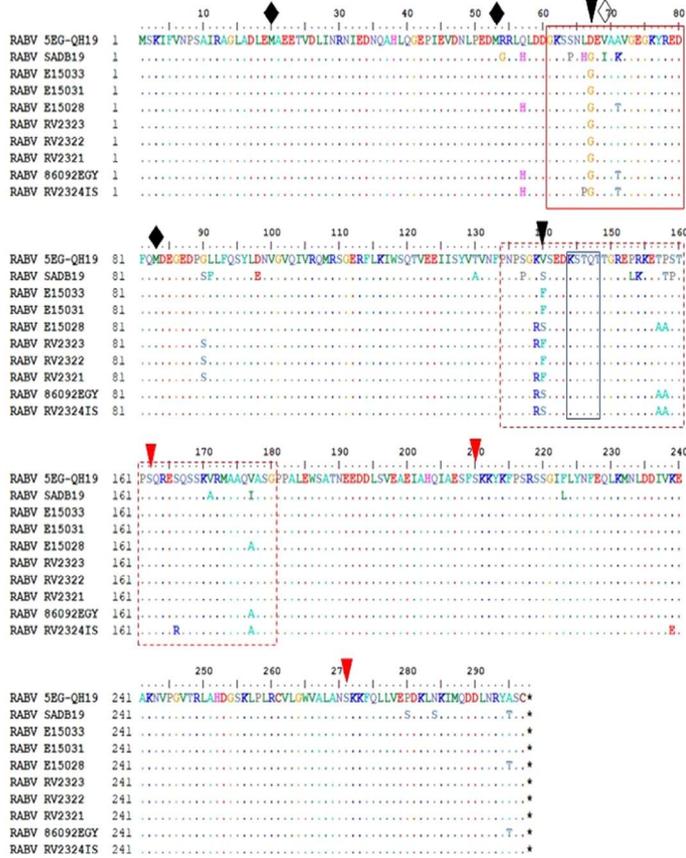
Figure 5: amino acid features of P protein. Variable domains (VD1) marked by red box, (VD2) marked by dashed red box. Black arrow heads indicate unique substitution in the studied sequence. Red arrowheads indicate P phosphorylation residues. Solid diamonds indicate in-frame M residues while empty diamond indicate the disrupted M69V.
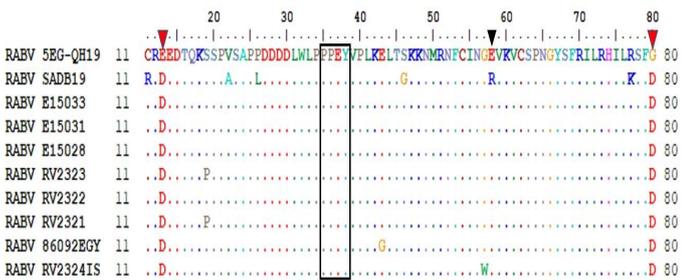
Figure 6: 5EG-QH19 M protein N-terminus showing the proline rich budding motif (black box), E58R substitution (black arrowhead) and the two unique mutations (red arrowheads).

changed the acidic polar aspartate into acidic polar glutamate E13D and into neutral nonpolar glycine G80D, respectively. The proline rich (PPXY) motif involved in viral budding and interaction with WW domains of the cells (Conzelmann et al., 1990; Mochizuki et al., 2009; Jiao et al., 2011) was conserved and expressed as PPEY (positions 35-38) in the study sequence and in the sequences

of the AF4 lineage (Figure 6). Position 58 included in the regulation of RABV RNA synthesis was expressed as 58E in 5EG-QH19 genome unlike the 58R of the SAD B19 vaccine strain but similar to the AF4 lineage sequences (Figure 6).

G protein, the sole surface viral glycoprotein responsible for host immune responses, contained 2 synonymous (A177G and T840C), and one nonsynonymous (T1456C) mutation which changed the non-polar neutral proline into the polar neutral serine S486P like that of the SAD B19 vaccine strain, the mutation was in the cytoplasmic domain of the G protein. Considerable level of conservation was seen in the G protein of the AF4 lineage but varied at least in 30 positions from that of the SAD B19 vaccine which might reflect a necessity for updating the current vaccine. Antigenic sites I (residue 231), II (residues 34-42,198-200), III (residues 330-338), IV (residue 264) and site a (residue342) were all conserved in the 5EG-QH19 genome as well as all AF4 lineage sequences. Two potential N-glycosylation sites 37NLS39 and 319NKT321 were conserved in the 5EG-QH19 sequence as well as all AF4 lineage sequences, a third previously reported potential N-glycosylation site 158NCS160 (Conzelmann et al., 1990) was ablated by the replacement of the neutral polar asparagine 158N into basic polar lysine 158K residue. This change was clear in all AF4 lineage sequences including study sequence. Pathogenicity related residues (242A, 255D, 268I, and 333R), and neuronal binding and motoneuron infection residue (330K), were also conserved (Figure 7) as previously reported (Tuffereau et al., 1989; Coulon et al., 1998; Takayama-Ito et al., 2006; Jiao et al., 2011).
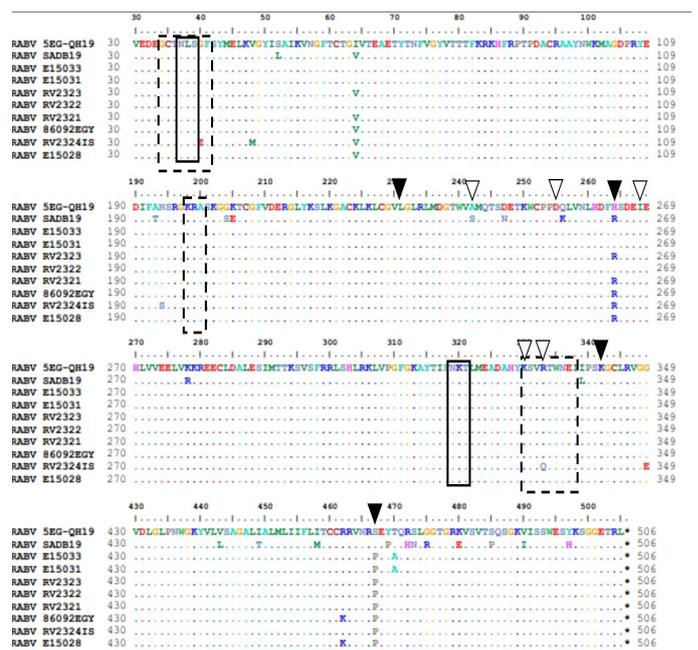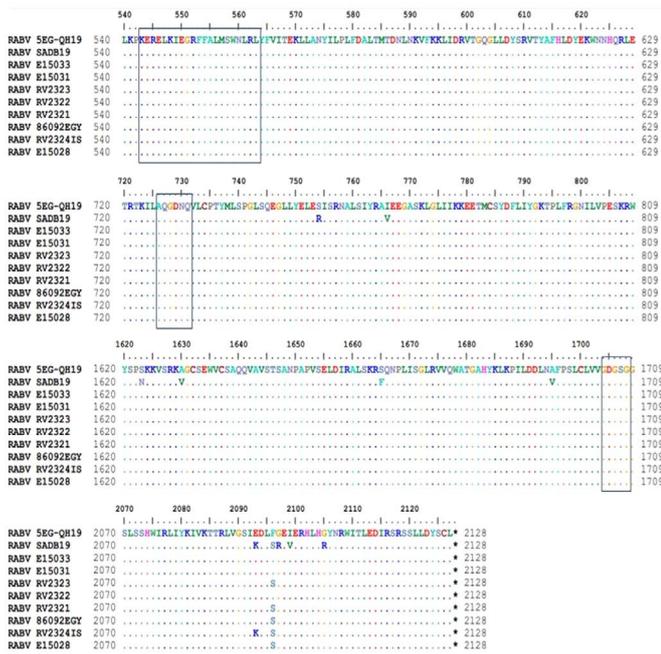

Figure 7: Structural features of G protein. Solid boxes show glycosylation sites. Dashed boxes for antigenic sites II and

III. Black arrow heads indicate remaining antigenic sites. White arrowheads indicate pathogenicity related residues.

L protein, the main component of RABV RNP and the viral polymerase, showed comparable conservation to the N protein, and composed 2127 amino acid residues distributed into six conserved motifs (I-VI) to the level of mononegaviruses (Poch et al., 1990). RNA binding domain positions (543-563) of motif II, polymerase active site of motif III positions (726-731) (Schnell and Conzelmann, 1995), and polyadenylation/protein kinase related glycine-rich domain of motif VI positions (1704-1708) (Poch et al., 1990) were all conserved in the 5EG-QH19 sequence as well as other AF4 lineage sequences (Figure 8).

Figure 8: functional domains and carboxyl end of the L protein, RNA binding domain (543-563), polymerase active site (726-731), and polyadenylation/protein kinase related glycine-rich domain (1704-1708) in boxes, respectively.

PHYLOGENETIC ANALYSIS
A phylogenetic tree was constructed using maximum likelihood method of MEGA version 11.0.10 (Tamura et al., 2021). It included 60 complete RABV sequences that represented known RABV clades/ lineages in Africa and Middle East (Table S1) and (Figure 9C). Leader and tailer sequences were excluded from the analysis as some sequences lacked them. The tree revealed that the 5EG-QH19 clustered with other Egyptian viruses (from 1950 to 2009) which belonged to the Africa 4 lineage (AF4) (figure 9C) of the cosmopolitan clade of *Lyssavirus rabies* first described in (David et al., 2007). These data were further supported by genetic analysis of partial N and G sequences from Egypt available in the GenBank database
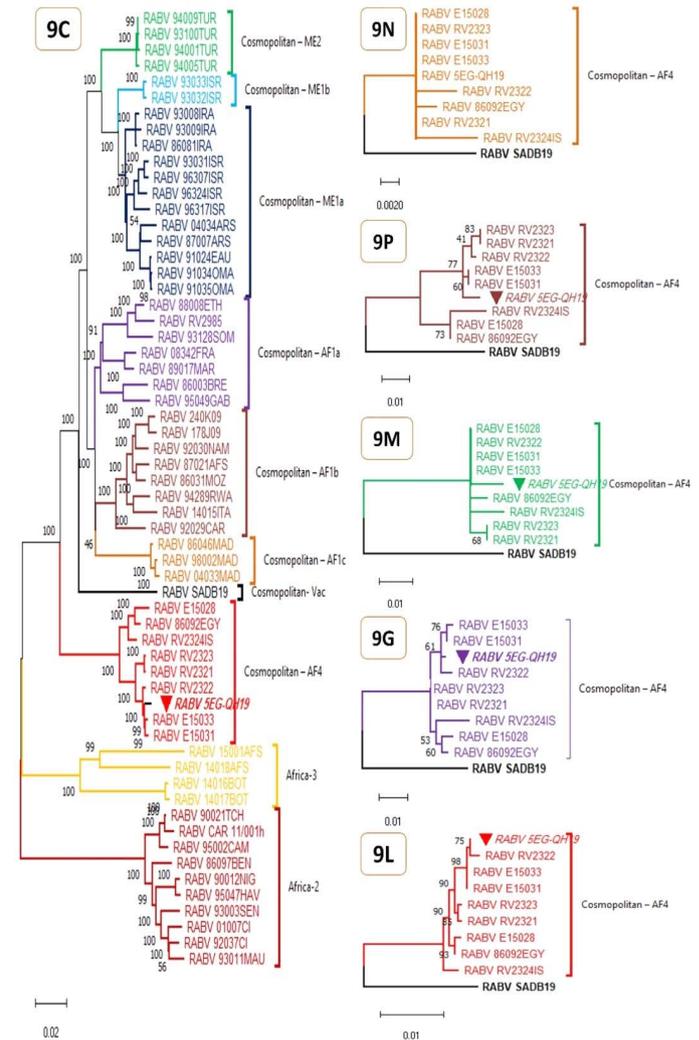
(data not shown). Further phylogenetic analysis relied only on AF4 lineage sequences with SAD B19 vaccine strain sequence as an outgroup (Figures 9N- 9L). Nucleoprotein and matrix protein followed by L showed the highest amino acid conservation level followed by P and G which showed more variation. Genomic as well as genetic trees based on nucleotide sequence (data not shown) supported the presence of two RABV sub-populations/ variants in Egypt; the old one contains sequences from 1950, 1979 and 2007 (putative AF4a) while the newer one included sequence from 1998 to 2019 (putative AF4b) (Figures 9C, 9P, 9G and 9L).

Figure 9: Maximum Likelihood phylogenetic analyses for complete RABV genomes (9C) and protein genes (9N, 9P, 9M, 9G, 9L). The evolutionary history was inferred by using Tamura-Nei model (Tamura and Nei 1993) for genomic tree (9C) or JTT matrix-based model (Jones et al. 1992) for protein genes (9N, 9P, 9M, 9G, 9L). The percentage of trees in which the associated taxa clustered together is shown next to the branches and is supported by 1K bootstrap test replicates for 9C and 100 bootstrap test replicates for 9N, 9P, 9M, 9G, and 9L. Initial tree(s) were obtained automatically by applying Neighbor-Join

and BioNJ algorithms to a matrix of pairwise distances estimated using either the Tamura-Nei model or the JTT model, and then the topology with superior log likelihood value is selected. The trees are drawn to scale, with branch lengths measured in the number of substitutions per site. All trees are midpoint rooted and clade/ lineage name is shown right to clusters. These analyses involved 60 nucleotide sequences (table S1) and 11740 positions (9C), or 10 amino acid sequences and 450, 279, 202, 524, and 2127 positions for 9N, 9P, 9M, 9G, and 9L respectively. The studied genome and its genes are marked with inverted solid triangle. Evolutionary analyses were conducted in MEGA11 (Tamura et al. 2021).

DISCUSSION

In this study, the complete 5EG-QH19 RABV genome from an Egyptian stray dog was analyzed a decade after the last published sequence (Dellicour et al., 2019). It was 19119nt long, obtained directly from brain tissue homogenate and assembled only from 851 relevant reads out of 422149 total reads (≤0.2%). The assembled genome was first compared to 60 RABV genomes representing almost all known clades/ lineages in the Middle East and Africa. 5EG-QH19 RABV genome clustered within the cosmopolitan AF4 lineage with other RABV genomes from Egypt. Further analysis relied on comparison with other AF4 genomes and SAD B19 vaccine strain. Excluding the lost 7 highly conserved nucleotides from the leader of 5EG-QH19 sequence, variation in the complete genome length was attributed mainly to deletions and insertions in the intergenic regions (1-2 nt). ORF sizes were fixed compared to either SAD B19 vaccine strain or to other AF4 genomes from Egypt (Table 3).

Like most of wild-type RABV, 5EG-QH19 genome and all members of AF4 lineage lacked the additional TTP signal reported in the G-L intergenic space of Pasteur vaccine (PV) and related strains (Conzelmann et al., 1990). Dependence on the second conserved TTP signal and utilization of the G-L intergenic space as a long untranslated region upstream to L gene was linked to reduced efficiency of L transcription, expression, and enhanced pathogenicity of wild-type viruses (Marston et al., 2007). The added adenine insertion in the G-L intergenic space of the 5EG-QH19 genome (Figure 2) could also be linked to enhanced pathogenicity of this wild-type virus. Transcription initiation, termination, and polyadenylation (TTP) signals were conserved as previously reported (Conzelmann et al., 1990; Mochizuki et al., 2009; Jiao et al., 2011) except N mRNA polyadenylation signal TG(A)6 which lacked the 7th A residue (figure 2). It was previously reported that additional adenine in the polyadenylation signals might affect the function of these signals, and the best length in 7

A residues (Conzelmann et al., 1990). A similar effect related to deletion cannot be excluded but studies are needed to prove it.

N was the highest conserved gene on nucleotide and amino acid levels with only three synonymous nucleotide changes. Antigenic sites I and IV were completely conserved for the 5EG-QH19 genome like other sequences of the AF4 lineage (Figure 4). RNA binding domain positions (298-350) and the phosphorylation residue 389S were also highly conserved for the 5EG-QH19 genome and for AF4 sequences indicating their biological importance in viral activities. A total of 8 amino acid changes from SAD B19 vaccine and only one residue in the antigenic site IV. A recent report describing changes in the antigenic sites of N in some Egyptian isolates might account for different viral antigenic groups within the lineage (Sultan et al., 2021).

The P protein was the most divergent with 2 amino acid substitutions in the variable domains VD1 and VD2 (Figure 5). One substitution V140F/S was in the context of N binding domain, its effect is expected to be minimal since a neutral amino acid (valine) replaced another neutral amino acid(s) (phenylalanine or serine), but the exact effect is to be elucidated. Substitution of an in-frame methionine V69M that was reported previously to contribute to translation initiation and P protein polymorphism (Chenik et al., 1995) in the 5EG-QH19 P and AF4 P sequences indicates ablation of the -68 P protein. Reported correlation between the first 52 positions in P amino terminal in the interaction with L indicates that other P forms specially those initiated at 53M, 69M and 83M might have no effect on RABV polymerase functions (Chenik et al., 1998). L binding domain was highly conserved in the 5EG-QH19 supporting the evidence that P interaction with L is essential for transcription (Mochizuki et al., 2009). LC8 interaction motif involved in the cross-neuronal transport of infective Rabies particles was also conserved in SAD B19 vaccine and the members of AF4 lineage (Raux et al., 2000; Lo et al., 2001).

M protein is known to control the shift from RABV RNA transcription to replication (Finke and Conzelmann, 2003). Single Arginine (58R) residue linked to this regulatory function was substituted in 5EG-QH19 (E58R) and in AF4 lineage sequences, a similar change was previously reported in wild type and vaccinal rabies strains (Mochizuki et al., 2009; Jiao et al., 2011). Experimental substitution of the basic arginine by the neutral Glycine in this position impaired the RNA regulatory function but had no effect on infectivity or virus composition (Finke and Conzelmann 2003). The observed shift to acidic glutamate might have a similar or different biological effect. The proline rich (PPEY) motif involved in viral budding

was found conserved. Two unique mutations in the 5EG-QH19 M protein N-terminus might influence the interaction of cellular components and viral budding (Figure 6).

It was known that RABV interacts with cellular receptors and fuses with the plasma membrane via G protein (Tuffereau et al., 1989; Whitt et al., 1991). G protein contains virulence related residues, and stimulates host immunity (Lafay et al., 1996; Badrane and Tordo 2001; Huang et al., 2017; Yang et al., 2020). Antigenic sites I, II, III, and site (a) were found conserved within the 5EG-QH19 like other AF4 lineage genomes. Whereas in antigenic site VI, 264H, in the study sequence replaced 264R in SAD B19 sequence. This change was also seen in AF4 genomes from 2009 and one genome from 1998 (Figure 7). The shift from basic polar arginine to the basic polar histidine is expected to have minimal effect on antigenicity and virulence. Virulence related residues (242A, 255D, 268I, 330K, and 333R) were also conserved in the 5EG-QH19 and AF4 G proteins while SAD B19 possessed 242S suggesting higher virulence of 5EG-QH19 and AF4 lineage members. Positions 330K and 333R were reported responsible for efficient motoneuron infection (Tuffereau et al., 1989; Coulon et al., 1998). The poorly glycosylated 37NLS39 and the efficiently glycosylated 319NKT321 N-glycosylation sites were conserved in 5EG-QH19 and AF4 G proteins. existence of any of these sites is enough for G protein glycosylation and surface expression (Kasturi et al., 1997). The previously reported efficiently glycosylated site 158NCS160 was absent in the structure of 5EG-QH19 G protein or AF4 lineage G proteins suggesting an adaptive response to the host and/or environment (Jiao et al., 2011). Collectively, the functional and structural elements of 5EG-QH19 and AF4 G protein exposed (Ecto) domains shown considerable level of conservation. On the other hand, transmembrane domain of 5EG-QH19 G was highly conserved, the sole amino acid change S467P was seen in the cytoplasmic (Endo) domain was also present in the SAD B19 G endo domain. As RBAV pathogenicity is shown to be related to changes in P, M, and G proteins (Yamada et al., 2006; Shimizu et al., 2007), the effect of reported mutations in the 5EG-QH19 proteins needs to be studied further.

Viral polymerase L, the key player in RNA transcription and replication (Poch et al., 1990; Morimoto et al., 1998), was highly conserved in the 5EG-QH19 like other AF4 lineage sequences specially the functional domains of RNA binding, polymerase active site, and polyadenylation/protein kinase glycine-rich motif (Figure 8). Domain I, encoding the conserved tripeptide GHP, and domain IV, rich in Proline, interacting with N binding sites were conserved. Domain V with numerous Cysteine and Histidine residues was also conserved among AF4 lineage sequences,

supporting the evidence that these domains play an important functional role (Poch et al., 1990). As previously reported, the 1499N cap methylation residue was conserved in 5EG-QH19 and other AF4 lineage sequences (Mochizuki et al., 2009). This high rate of conservation reflects high structural function of L protein.

Phylogenetic analysis of RABV genomes (Figure 9C) and deduced amino acid sequences (Figure 9N-9L) supported the recently proposed division of AF4 lineage into two sub-lineages (Sultan et al., 2021). The AF4 lineage was calculated to diverge from cosmopolitan clade around 1730 (Cile Troupin et al., 2016). We expect divergence into 2 sub-lineages (AF4a and AF4b) had occurred around 1980. Both sub-lineages cocirculated for approximately 3 decades then from 2009 only AF4b related viruses are being detected. The limited number of available sequences make these dates mostly uncertain. More research is required to support these speculations.

CONCLUSION

The present study analyzed the complete genome of 5EG-QH19 RABV as a member of cosmopolitan AF4 lineage, shed some light on the structural features of the lineage, and supported the recent proposed division of 2 sub-lineages from the currently dominant lineage in Egypt. The known strong purifying non diversifying selective pressure on RABV explains the observed high conservation rates in the study genome and other AF4 lineage viruses and supports strong structural functional agreement of RABV proteins (Cile Troupin et al., 2016). we tried to explain the possible effects of the amino acid changes observed in the protein genes, but the exact effects of these substitutions in an infectious RABV particle requires in-vivo studies. The study was limited by multiple factors including limited funds to apply whole genome sequencing for multiple specimens, high cost of whole genome sequencing and the safety requirements for in-vivo studies. Divergence from SAD B19 vaccine strain sequence was reported. Continuous RABV surveillance specially on border governorates, in wild canids, and in bats is required to investigate the presence of clades/ lineages other than AF4 in Egypt.

ACKNOWLEDGEMENTS

CONFLICT OF INTEREST

The authors have no relevant financial or non-financial in-

terests to disclose.

NOVELTY STATEMENT

In this study, authors updated the available knowledge about lyssavirus rabies situation in Egypt, provided the first detailed description of africa-4 lineage genome characteristics and supported the recent evidence that 2 subpopulations are present with the Africa-4 lineage.

AUTHORS CONTRIBUTION

Hussein Aly Hussein conceived the study, advised the whole work, corrected the initial manuscript. Momtaz Abdelhady Shahein and Basem Mohamed Ahmed approved the concept and co-advised the work. Momtaz Abdelhady Shahein supervised the work and supplied materials and workspace. Amthal Ahmed Fouad collected samples, performed the experiments and analysis, and drafted the initial manuscript. Basem Mohamed Ahmed analyzed the data and drafted the first manuscript. All authors had commented on earlier versions of the manuscript, revised and approved the final manuscript.

REFERENCES

Abd El Rahman S (2015). Detection of Rabies Virus and its Pathological Changes in Brain of Buffaloes in Egypt. Adv. Anim. Vet. Sci. 3:588–593. https://doi.org/10.14737/journal.aavs/2015/3.11.588.593

Al-Youm A-M (2017). Egypt addresses OIE over spread of rabies. Egypt Indep.

Badrane H, Tordo N (2001) Host switching in Lyssavirus history from the Chiroptera to the Carnivora orders. J. Virol. 75:8096–8104. https://doi.org/10.1128/JVI.75.17.8096-8104.2001

Botros BA, Moch RW (1976). Rabies in Egypt: a report of four cases showing some deviation from classical rabies. Bull Anim. Health Prod. Afr. 24:29–34

Brunker K, Jaswant G, Thumbi SM, et al (2020). Rapid in-country sequencing of whole virus genomes to inform rabies elimination programmes. Wellcome Open Res. 5:3. https://doi.org/10.12688/wellcomeopenres.15518.2

Brunker K, Nadin-Davis S, Biek R (2018). Genomic sequencing, evolution and molecular epidemiology of rabies virus. Rev. Sci. Tech. 37:401–408. https://doi.org/10.20506/rst.37.2.2810

Chenik M, Chebli K, Blondel D (1995) Translation initiation at alternate in-frame AUG codons in the rabies virus phosphoprotein mRNA is mediated by a ribosomal leaky scanning mechanism. J. Virol. 69:707–712. https://doi.org/10.1128/jvi.69.2.707-712.1995

Chenik M, Schnell M, Conzelmann KK, Blondel D (1998). Mapping the interacting domains between the rabies virus polymerase and phosphoprotein. J. Virol. 72:1925–1930. https://doi.org/10.1128/JVI.72.3.1925-1930.1998

Cile Troupin C, Dacheux L, Tanguy M, et al (2016). Large-Scale Phylogenomic Analysis Reveals the Complex Evolutionary History of Rabies Virus in Multiple Carnivore Hosts. PLoS Pathog. 12:e1006041. https://doi.org/10.1371/journal.ppat.1006041

Conzelmann KK, Cox JH, Schneider LG, Thiel HJ (1990). Molecular cloning and complete nucleotide sequence of the attenuated rabies virus SAD B19. Virology. 175:485–499. https://doi.org/10.1016/0042-6822(90)90433-R

Coulon P, Ternaux J-P, Flamand A, Tuffereau C (1998). An Avirulent Mutant of Rabies Virus Is Unable To Infect Motoneurons In Vivo and In Vitro. J. Virol. 72:273. https://doi.org/10.1128/JVI.72.1.273-278.1998

David D, Hughes GJ, Yakobson BA, et al (2007). Identification of novel canine rabies virus clades in the Middle East and North Africa. J. Gen. Virol. 88:967–980. https://doi.org/10.1099/vir.0.82352-0

Dellicour S, Troupin C, Jahanbakhsh F, et al (2019). Using phylogeographic approaches to analyse the dispersal history, velocity and direction of viral lineages — Application to rabies virus spread in Iran. Mol. Ecol. 28:4335–4350. https://doi.org/10.1111/mec.15222

Delmas O, Holmes EC, Talbi C, et al (2008). Genomic Diversity and Evolution of the Lyssaviruses. PLoS One. 3:. https://doi.org/10.1371/JOURNAL.PONE.0002057

Faber M, Pulmanausahakul R, Nagao K, et al (2004). Identification of viral genomic elements responsible for rabies virus neuroinvasiveness. Proc Natl. Acad. Sci. U SA. 101:16328–16332. https://doi.org/10.1073/PNAS.0407289101/ASSET/37091699-2F91-4FC0-856E-8F8311A43674/ASSETS/GRAPHIC/ZPQ0460464480005.JPEG

Finke S, Conzelmann K-K (2003). Dissociation of rabies virus matrix protein functions in regulation of viral RNA synthesis and virus assembly. J. Virol. 77:12074–12082. https://doi.org/10.1128/JVI.77.22.12074-12082.2003

Fischer S, Freuling CM, Müller T, et al (2018). Defining objective clusters for rabies virus sequences using affinity propagation clustering. PLoS Negl. Trop. Dis. 12:e0006182. https://doi.org/10.1371/journal.pntd.0006182

Goto H, Minamoto N, Ito H, et al (2000). Mapping of epitopes and structural analysis of antigenic sites in the nucleoprotein of rabies virus. J. Gen. Virol. 81:119–127. https://doi.org/10.1099/0022-1317-81-1-119

Hall TA (1999). BIOEDIT: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. Nucleic Acids Symp. Ser.

Huang J, Zhang Y, Huang Y, et al (2017). The ectodomain of rabies virus glycoprotein determines dendritic cell activation. Antiviral. Res. 141:1–6. https://doi.org/10.1016/J.ANTIVIRAL.2017.01.022

Jiao W, Yin X, Li Z, et al (2011). Molecular characterization of China rabies virus vaccine strain. Virol. J. 8:521. https://doi.org/10.1186/1743-422X-8-521

Jones DT, Taylor WR, Thornton JM (1992). The rapid generation of mutation data matrices from protein sequences. Comput. Appl. Biosci. 8:275–282. https://doi.org/10.1093/BIOINFORMATICS/8.3.275

Kasturi L, Chen H, Shakin-Eshleman SH (1997). Regulation of N-linked core glycosylation: Use of a site-directed mutagenesis approach to identify Asn-Xaa-Ser/Thr sequons that are poor oligosaccharide acceptors. Biochem. J. 323:415–419. https://doi.org/10.1042/bj3230415

Kuzmin I V., Wu X, Tordo N, Rupprecht CE (2008). Complete genomes of Aravan, Khujand, Irkut and West Caucasian bat viruses, with special attention to the polymerase gene

and non-coding regions. Virus Res. 136:81–90. https://doi.org/10.1016/J.VIRUSRES.2008.04.021

Lafay F, Benmansour A, Chebli K, Flamand A (1996). Immunodominant epitopes defined by a yeast-expressed library of random fragments of the rabies virus glycoprotein map outside major antigenic sites. J. Gen. Virol. 77 (Pt 2):339–346. https://doi.org/10.1099/0022-1317-77-2-339

Lo KWH, Naisbitt S, Fan JS, et al (2001). The 8-kDa dynein light chain binds to its targets via a conserved (K/R)XTQT motif. J. Biol. Chem. 276:14059–14066. https://doi.org/10.1074/JBC.M010320200

Marston DA, McElhinney LM, Ellis RJ, et al (2013). Next generation sequencing of viral RNA genomes. BMC Genom. 14:1–12. https://doi.org/10.1186/1471-2164-14-444/FIGURES/3

Marston DA, McElhinney LM, Johnson N, et al (2007). Comparative analysis of the full genome sequence of European bat lyssavirus type 1 and type 2 with other lyssaviruses and evidence for a conserved transcription termination and polyadenylation motif in the G–L 3′ non-translated region. J. Gen. Virol. 88:1302–1314. https://doi.org/10.1099/vir.0.82692-0

Marston DA, Wise EL, Ellis RJ, et al (2015). Complete genomic sequence of rabies virus from an ethiopian wolf. Genom. Announc. 3:. https://doi.org/10.1128/genomeA.00157-15

Mochizuki N, Kobayashi Y, Sato G, et al (2009). Complete genome analysis of a rabies virus isolate from Brazilian wild fox. Arch. Virol. 154:1475–1488. https://doi.org/10.1007/s00705-009-0475-9

Morimoto K, Akamine T, Takamatsu F, Kawai A (1998). Studies on rabies virus RNA polymerase: 1. cDNA cloning of the catalytic subunit (L protein) of avirulent HEP-flury strain and its expression in animal cells. Microbiol. Immunol. 42:485–496. https://doi.org/10.1111/J.1348-0421.1998.TB02314.X

Nadin-Davis SA, Abdel-Malik M, Armstrong J, Wandeler AI (2002). Lyssavirus P gene characterisation provides insights into the phylogeny of the genus and identifies structural similarities and diversity within the encoded phosphoprotein. Virology. 298:286–305. https://doi.org/10.1006/viro.2002.1492

Nadin-Davis SA, Huang W, Wandeler AI (1997). Polymorphism of rabies viruses within the phosphoprotein and matrix protein genes. Arch Virol. 142:979–992. https://doi.org/10.1007/S007050050133

Oem JK, Kim SH, Kim YH, et al (2013). Complete genome sequences of three rabies viruses isolated from rabid raccoon dogs and a cow in Korea. Virus Genes. 47:563–568. https://doi.org/10.1007/s11262-013-0923-1

Poch O, Blumberg BM, Bougueleret L, Tordo N (1990). Sequence comparison of five polymerases (L proteins) of unsegmented negative-strand RNA viruses: theoretical assignment of functional domains. J. Gen. Virol. 71 (Pt 5):1153–1162. https://doi.org/10.1099/0022-1317-71-5-1153

Raux H, Flamand A, Blondel D (2000). Interaction of the rabies virus P protein with the LC8 dynein light chain. J. Virol. 74:10212–10216. https://doi.org/10.1128/JVI.74.21.10212-10216.2000

Schnell MJ, Conzelmann KK (1995). Polymerase activity of in vitro mutated rabies virus L protein. Virology. 214:522–530. https://doi.org/10.1006/VIRO.1995.0063

Shimizu K, Ito N, Mita T, et al (2007). Involvement of nucleoprotein, phosphoprotein, and matrix protein genes of rabies virus in virulence for adult mice. Virus Res. 123:154–160. https://doi.org/10.1016/J.VIRUSRES.2006.08.011

Shwiff S, Hampson K, Anderson A (2013). Potential economic benefits of eliminating canine rabies. Antiviral Res. 98:352–356. https://doi.org/10.1016/J.ANTIVIRAL.2013.03.004

Sultan S, Ahmed SAH, Abdelazeem MW, Hassan S (2021). Molecular characterisation of rabies virus detected in livestock animals in the southern part of Egypt during 2018 and 2019. Acta Vet. Hung. 69:80–87. https://doi.org/10.1556/004.2021.00005

Takayama-Ito M, Ito N, Yamada K, et al (2006). Multiple amino acids in the glycoprotein of rabies virus are responsible for pathogenicity in adult mice. Virus Res. 115:169–175. https://doi.org/10.1016/J.VIRUSRES.2005.08.004

Tamura K, Nei M (1993). Estimation of the number of nucleotide substitutions in the control region of mitochondrial DNA in humans and chimpanzees. Mol. Biol. Evol. 10:512–526. https://doi.org/10.1093/OXFORDJOURNALS.MOLBEV.A040023

Tamura K, Stecher G, Kumar S (2021). MEGA11: Molecular Evolutionary Genetics Analysis Version 11. Mol. Biol. Evol. 38:3022–3027. https://doi.org/10.1093/MOLBEV/MSAB120

Tuffereau C, Leblois H, Bénéjean J, et al (1989). Arginine or lysine in position 333 of ERA and CVS glycoprotein is necessary for rabies virulence in adult mice. Virology. 172:206–212. https://doi.org/10.1016/0042-6822(89)90122-0

Walker PJ, Freitas-Astua J, Bejerman N, et al (2022). ICTV Virus Taxonomy Profile: Rhabdoviridae 2022. J. Gen. Virol. 103:. https://doi.org/10.1099/JGV.0.001689

Whitt MA, Buonocor L, Prehaud C, Rose JK (1991). Membrane fusion activity, oligomerization, and assembly of the rabies virus glycoprotein. Virology. 185:681–688. https://doi.org/10.1016/0042-6822(91)90539-N

WHO (2014). WHO Guide for Rabies Pre and Post-exposure Prophylaxis in Humans. World Heal Organ 1–21.

Yamada K, Ito N, Takayama-Ito M, et al (2006). Multigenic relation to the attenuation of rabies virus. Microbiol. Immunol. 50:25–32. https://doi.org/10.1111/J.1348-0421.2006.TB03767.X

Yang F, Lin S, Ye F, et al (2020). Structural Analysis of Rabies Virus Glycoprotein Reveals pH-Dependent Conformational Changes and Interactions with a Neutralizing Antibody. Cell Host Microbe. 27:441-453.e7. https://doi.org/10.1016/J.CHOM.2019.12.012

WOAH (2022). Rabies (infection with rabies virus and other lyssaviruses), Manual of Diagnostic Tests and Vaccines for Terrestrial Animals 2022. Available at: https://www.woah.org/fileadmin/Home/eng/Health_standards/tahm/3.01.18_RABIES.pdf (Accessed: March 26, 2023).