

Molecular characterization of KGH, the first human isolate of rabies virus in Korea

Jun-Sun Park · Chi-Kyeong Kim · Su Yeon Kim ·
Young Ran Ju

Received: 12 September 2012 / Accepted: 12 November 2012 / Published online: 15 December 2012
© Springer Science+Business Media New York 2012

Abstract The complete genome sequence of the KGH strain of the first human rabies virus, which was isolated from a skin biopsy of a patient with rabies, whose symptoms developed due to bites from a raccoon dog in 2001. The size of the KGH strain genome was determined to be 11,928 nucleotides (nt) with a leader sequence of 58 nt, nucleoprotein gene of 1,353 nt, phosphoprotein gene of 894 nt, matrix protein gene of 609 nt, glycoprotein gene of 1,575 nt, RNA-dependent RNA polymerase gene of 6,384 nt, and trailer region of 69 nt. Sequence similarity was compared with 39 fully sequenced rabies virus genomes currently available, and the result showed 70.6–91.6 % at the nucleotide level, and 82.8–97.9 % at the amino acid level. The deduced amino acids in the viral protein were compared with those of other rabies viruses, and various functional regions were investigated. As a result, we found that the KGH strain only had a unique amino acid substitution that was identified to be associated either with host immune response and pathogenicity in the N protein, or with a related region regulating STAT1 in the P protein, and related to pathogenicity in G protein. Based on phylogenetic analyses using the complete genome of 39 rabies viruses, the KGH strain was determined to be closely related with the NNV-RAB-H strain and transplant rabies virus serotype 1, which are Indian isolates, and was

confirmed to belong to the Arctic-like 2 clade. The KGH strain was most closely related to the SKRRD0204HC and SKRRD0205HC strain when compared with Korean animal isolates, which was separated around the same time and place, and belonged to the Gangwon III subgroup.

Keywords Rabies virus · KGH · Complete genome sequence · Phylogenetic analysis

Introduction

Rabies is a major threatening zoonotic disease of humans and animals, which has been reported since 2300 BC [1, 2]. In 1802, rabies was reported as a communicable disease that could be transferred from rabid to healthy dogs via the saliva [3], and was verified to be transmitted from one animal species to another [4, 5]. Although rabies virus vaccines have been developed, human rabies is often misunderstood as a neglected infectious disease, and hence, results in some 55,000 deaths annually, mainly in Asia and Africa [6].

The rabies virus is a member of the family Rhabdoviridae, genus *Lyssavirus*. *Lyssavirus* has seven genotypes. The most widespread classical rabies viruses belong to genotype 1, and virus isolates derived from Africa, Asia, and European bat species are genotypes 2–7 [7, 8]. The rabies virus genome is a negative-sense, single-stranded RNA, and the RNA genome is tightly encapsidated by nucleoprotein [6]. The genome is approximately 12 kb, and is encoded with five structural proteins, including nucleoprotein (N), phosphoprotein (P), matrix protein (M), glycoprotein (G), and RNA-dependent RNA polymerase (L).

The N protein forms the ribonucleoprotein complex together with the P and L proteins and viral genomic RNA

J.-S. Park · Y. R. Ju (✉)
187 Osongsaengmyeong2-ro, Osong-eup, Chungcheongbuk-do,
Cheongwon-gun 363-951, South Korea
e-mail: juyran@korea.kr

J.-S. Park
e-mail: uavcom@hanmail.net

C.-K. Kim · S. Y. Kim · Y. R. Ju
Division of Zoonoses, Korea National Institute of Health,
Osong, South Korea

to regulate RNA transcription and replication, evade activity of RIG-I-mediated signaling which is a host innate immune response, and affect pathogenicity [9]. The G protein mediates easy entry of the virus into cells, affects viral spread and uptake speed, and plays an important role in neurovirulence and selection of antigenic sites [6, 10]. Phylogenetic analyses have focused on the use of partial sequences of the N or G genes in other rabies virus strains. The rabies virus N gene is mostly used for evolutionary studies, because it is well conserved. The G gene is mostly used for epidemiological and evolutionary studies, because it changes amino acids at antigenic sites to escape from the host immune system [11, 12]. As the G–L non-coding region is not affected by positive selection, it is used for phylogenetic analyses [12]. Host-specific features can be confirmed through determining the complete genome sequence, and very subtle changes in unknown regions of the viral genome can be identified [13]. These methods of analysis contribute to understanding the origin and transmission pattern of the disease, and provide an efficient strategy for regulating rabies based on molecular epidemiological data [14].

Rabies was reported for the first time in Korea in 1907. Since then, an average of 500 cases per year were reported until 1984, when the incidence dropped to 30 cases in 1985 following the widespread vaccination of dogs. No incidence of rabies was reported between 1985 and 1992, but animal rabies cases were reported during the period of 1993–2004 [15]. No human rabies case was reported between 1985 and 1998, but a single case in 1999, and five further cases up until 2004 were reported. Partial sequences of the N, G, P, and G–L non-coding regions have been analyzed, but the complete genome sequence analysis is reported for the first time here [15, 16]. Through maximum-parsimony analysis, strains of rabies viruses, which are prevalent in Asia, were divided into five distinct clusters, including Asian 1, Asian 2, Asian 3, Arctic-related, and Cosmopolitan clusters [17]. The Arctic-like lineage is composed of the Arctic-like 1 and Arctic-like 2 clades [18]. According to a recent classification, Korean isolates belong to the Arctic-like 2, and consist of the four sub-groups of Gangwon I, Gangwon II, Gangwon III, and Gyeonggi [16, 18].

In the present study, we have determined the complete genome sequence of the KGH strain, which was the first human rabies virus isolated from a skin biopsy of a Korean patient with rabies in 2001. The KGH strain was compared to 39 fully sequenced rabies virus strains at the molecular level. A complete genome sequence analysis is a basis for understanding the molecular characteristics and genetic diversity of rabies viruses prevalent in Korea. In particular, a protein substitution associated with the host immune response, pathogenicity, and signal transducers and activators of transcription (STAT)1 regulation was observed in

the KGH strain. Our results will aid the understanding of the immune response of rabies viruses prevalent in Korea, as well as the identification of their relationship with pathogenicity.

Materials and methods

Virus isolation

The KGH rabies virus strain was isolated from the hair follicles of a 68-year old patient who was bitten by a rabid raccoon dog in Hwacheon-gun, Gangwon-do in 2001 and that had an 11-week incubation period without PEP treatment. The patient was observed for symptoms of hyperesthesia, headache, nausea, photophobia, and hydrophobia. This rabies virus strain was isolated from the brains of suckling mice that were intracerebrally injected with homogenized hair follicles of the patient. The brain tissues were 50 % homogenated in a tube with PBS and ceramic beads.

Reverse transcription and amplification

Viral RNA was extracted from 200 μ l of a 10 % brain homogenate from the infected mice using Trizol LS reagent (Invitrogen, Carlsbad, CA, USA) according to the manufacturer's recommendations. Extracted viral RNA was resuspended in 30 μ l of diethylpyrocarbonate-treated water (Invitrogen, USA). Superscript III (Invitrogen, USA) was used for the first-strand cDNA synthesis according to the manufacturer's instructions. Three synthesized cDNA fragments (RV24F, RV3866F, and RV6325F) were amplified by polymerase chain reaction (PCR) using Pyrobest DNA polymerase (Takara Bio Inc., Shiga, Japan), and the primers are listed in Table 1. PCR primer sets for amplifying the RV24F amplicon (nt 1–4178) were RV24F and RV4178R, those for amplifying the RV3866F amplicon (nt 4068–8003) were RV4068F and RV8003R, and those for amplifying the RV6325F amplicon (nt 7623–11928) were RV7623F and RV11928R. The PCR parameters included denaturation at 94 °C for 30 s, annealing at 52 °C for 30 s, extension at 72 °C for 5 min, and a final extension at 72 °C for 10 min. After electrophoresis with 1 % agarose gels, all PCR products were purified using QIAquick Gel Extraction kit (Qiagen, Hilden, Germany). Direct sequencing was conducted twice in both directions.

5'-Random amplification of cDNA ends (RACE) PCR and RLM-3'-RACE

A standard 5'-RACE kit (Invitrogen, USA) was used to determine the 5'-terminal ends of the genomic RNA. The reaction was performed according to the manufacturer's

Table 1 Primers used for cDNA synthesis, PCR amplification, and RLM-RACE

Primer	Sequence (5'–3')	Position ^a	Polarity	Use
RV 24F	ACGCTTAACAACAAAATCAGAAAA	1–24	Sense	cDNA
RV 3877F	CTACTGCCCTACCAACCATGA	3878–3898	Sense	cDNA
RV 6325F	AAGTTATTAAGATATTAGAGCCATA	6300–6324	Sense	cDNA
RV 4178R	CCTCCACAACGAGATGTTTCGAT	4179–4200	Antisense	PCR
RV 32F	GGACTTAGACTTATGGATGGAACG	4071–4094	Sense	PCR
RV 7R	GTAGTGAAAGACTGCCTGT	7981–7999	Antisense	PCR
RV 7623F	TAACCAGGTTCTATGTCCGACAT	7600–7622	Sense	PCR
RV 11928R	ACGCTTAACAAAAACAATAAAGA	11904–11928	Antisense	PCR
RVL F	TCATCTCACTGGATCAGG	11627–11644	Sense	5'RACE
RVL FN	TACAACAGGTGGATTACC	11729–11748	Sense	5'RACE
RV 470R	GACTAGAGACGCATGCTCGGGGAC	470–493	Antisense	3'RACE
RVN 6R	AAGCCCAATCCCTTCTACATCAG	411–434	Antisense	3'RACE
TB 3'P-End	GACAACAACTGGCATATTTCTTGCTGA			3'RACE
TB 3'end-rev	AGCAAGAAATATGCCAGTT			3'RACE

^a Nucleotide positions are based on the complete genome sequence of the KGH strain

instructions. For the first-strand cDNA synthesis, the RVL F gene specific primer with Superscript III RT (Invitrogen, USA) was used with 8 μ l of viral RNA. The synthesized cDNA was purified in a 30 μ l final volume using a S.N.A.P. column purification kit (Invitrogen, USA). A 10 μ l aliquot of purified cDNA was used for the TdT-tailing reaction. For the first-round PCR, 5 μ l of dC-tailed cDNA and an anti-sense abridged anchor primer included in the 5'RACE kit were used with the RVL F positive sense primer. The RVL FN internal positive sense primer and abridged anchor primer were used for the second-round PCR with a pair of primers. All PCR reactions were amplified with Pyrobest DNA polymerase with the following thermocycling parameters: 25 cycles (for the first-round PCR) and 30 cycles (for the second-round PCR) of 30 s at 94° C, 30 s at 60° C, 30 s at 72° C, and a final extension of 10 min at 72° C. The PCR products were inserted into the pJET vector (Fermentas, Burlington, ONT, Canada) and cloned in ECOS *E. coli* DH10B competent cells (Yeastern, Taipei, Taiwan). We confirmed that 50 independent clones were randomly selected for sequencing.

Genomic RNA was identified by 3'-RLM-RACE to determine the 3'-end. Viral RNA and the 5'-phosphorylated oligonucleotide 3'-P end were ligated using T4 RNA ligase (New England Biolabs, Ipswich, MA, USA) at 16° C for 2 h. RNA-3'-P end ligated RNA was recovered with TRIzol LS reagent and was used for cDNA synthesis with 3'End Rev, which is a complementary oligonucleotide. PCR was amplified with Pyrobest DNA polymerase using a pair of primers, 3'End Rev and the R470 negative-sense primer. The thermocycling parameters were 30 cycles of 30 s at 94° C, 30 s at 55° C, 50 s at 72° C, and a final extension for 10 min at 72° C. The PCR products were inserted into the pJET vector and cloned in *E. coli* DH10B

competent cells. We confirmed that 80 independent clones were randomly selected for sequencing.

Phylogenetic and genetic distance analyses

As a full genome sequence dataset, 39 full-length nucleotide sequences of other rabies virus strains listed on available in Genbank, and 23 glycoprotein gene sequence of rabies virus strains separated in Korea, were used as shown in Table 2. Full sequences were assembled using the Seqman program of DNASTAR, version 5. Multiple sequence alignments were performed using the Megalign tool (DNASTAR). Phylogenetic trees were constructed, and evolutionary distances analyses were conducted using MEGA version 5.05 using the neighbor-joining algorithm with the Kimura two-parameter model. Bootstrap values were statistically evaluated with 1,000 replicates.

Results

Comparison of nucleotide and deduced amino acids sequence of the KGH strain

To characterize the KGH strain, we determined the complete genome sequence of the first Korean rabies virus strain. The full-length RNA genome of the KGH strain was confirmed to have a length of 11,928 nt, and was composed of 58 nt genomic RNA at the 3'-terminal and 69 nt genomic RNA at the 5'-terminal. Five coding regions consisted of 1,353 nt for the N gene (450 amino acids), 894 nt for the P gene (297 amino acids), 609 nt for the M gene (202 amino acids), 1,575 nt for the G gene (524 amino acids), and 6,384 nt for the L gene (2127 amino acids). The intergenic

Table 2 History of the rabies virus strains used in this study

Virus strain ^a	Accession no.	Place	Years	Source	Virus strain	Accession no.	Place	Years	Source
KGH		Korea(Hwacheon)	2001	Human	SKRBV9801YC	DQ076105	Korea(Yeoncheon)	1998	Cattle
07240FRA	EU626552	France	2007	Cat	SKRRD9901PJ	DQ076103	Korea(Paju)	1999	Raccoon dog
8743THA	EU293121	Thailand	1983	Human	SKRRD9902PJ	DQ076102	Korea(Paju)	1999	Raccoon dog
8764THA	EU293111	Thailand	1983	Human	SKRRD9903YG	DQ076099	Korea(Yanggu)	1999	Raccoon dog
9147FRA	EU293115	France	1991	Fox	SKRDG9901GY	DQ076100	Korea(Goyang)	1999	Dog
BD06	EU549783	China	2008	Dog	SKRDG9902GY	DQ076101	Korea(Goyang)	1999	Dog
CTN181	EF564174	China	1956	Human	SKRRD0204CW	DQ076094	Korea(Cheorwon)	2002	Raccoon dog
D01	FJ712193	China	2008	Dog	SKRDG0203CW	DQ076104	Korea(Cheorwon)	2002	Dog
D02	FJ712194	China	2008	Dog	SKRBV0403CW	DQ076095	Korea(Cheorwon)	2004	Cattle
RV1333	EF157977	United Kingdom	2002	Human	SKRDG0204HC	DQ076093	Korea(Hwacheon)	2002	Dog
F02	FJ712195	China	2008	Ferret	SKRRD0205HC	DQ076096	Korea(Hwacheon)	2002	Raccoon dog
F04	FJ712196	China	2008	Ferret	SKRBV0404HC	DQ076097	Korea(Hwacheon)	2004	Cattle
FJ008	FJ866835	China	2008	Dog	SKRRD0406CC	DQ076098	Korea(Chuncheon)	2004	Raccoon dog
FJ009	FJ866836	China	2008	Dog	KRVR0801	GU937024	Korea(Sokcho)	2008	Raccoon dog
HN10	EU643590	China	2006	Dog	KRVC0802	GU937025	Korea(Inje)	2008	Dog
Nishigahara	AB044824	Japan	2001	Human	KRVR0803	GU937026	Korea(Sokcho)	2008	Raccoon dog
NNV-RAB-H	EF437215	India	2006	Human	KRVR0804	GU937027	Korea(Sokcho)	2008	Raccoon dog
Rabies virus serotype 1	AY956319	Germany (imported from India)	2004	Human	KRVR0901	GU937028	Korea(Goseong)	2009	Raccoon dog
RRV ON-99-2	EU311738	Canada	1999	Raccoon	KRVB0902	GU937029	Korea(Goseong)	2009	Cattle
SHBRV-18	AY705373	USA	1983	Bat	KRVB0904	GU937031	Korea(Goseong)	2009	Cattle
JX08-45	GU647092	China	2008	Ferret	KRVB0903	GU937030	Korea(Hongcheon)	2009	Cattle
GX4	GU358653	China	1994	Dog	KRVB0905	GU937032	Korea(Inje)	2009	Cattle
SH06	GU345748	China	2006	Dog	KRVB0907	GU937034	Korea(Inje)	2009	Cattle
J	GU345747	China	1986	Human	91RABN0730	U11748	Canada	1991	Red fox
CQ92	GU345746	China	1992	Dog	90RABN9239	U11750	Canada	1990	Red fox
CTN-1	FJ959397	China	1956	Human	90RABN9341	U11752	Canada	1990	Red fox
BR-Px1	AB362483	Brazil	2002	Fox	Hudson Bay-1	U03767	Canada		Dog
H-08-1320	AB569299	SriLanka	2008	Human	9901NEP	EU086153	Nepal	1998	Dog
RbE3-15-5	EU182347				9902NEP	EU086154	Nepal	1998	Goat
RBE3-15	EU182346				04029AFG	EU086128	Afghanistan	2004	Dog
Flury-HEP	GU565704		1948		CHAND03	AY987478	India	1999	Dog
Flury-LEP-C	GU565703		1948		RVD	AY237121	India	1998	Dog
ERA	EF206707	USA		Vaccine	NEP1-DG	AF325489	Nepal	1989	Dog
Ni-CE	AB128149	Japan	2003	Vaccine	94257SRI	EU086156	SriLanka	1986	Dog
PV	M13215	France	1882	Vaccine	04030PHI	EU086155	Philippines	2004	Human
RC-HL	AB009663	Japan	2000	Vaccine	HM65	AY257980	Thailand		Human

Table 2 continued

Virus strain ^a	Accession no.	Place	Years	Source	Virus strain	Accession no.	Place	Years	Source
RV-97	EF542830	Russia		Vaccine	HM208	AY257983	Thailand	Human	Vaccine
SAD B19	M31046	USA	1990	Vaccine	9910LAO	EU086152	Laos	1999	Dog
SRV9	AF499686	China	2002	Vaccine	9912CBG	EU086132	Cambodia	1998	Dog
NeiMeng1025C	EU284098	China(Jilin)	2007	Raccoon dog	9908CBG	EU086130	Cambodia	1999	Dog
NeiMeng1025B	EU284097	China(Jilin)	2007	Raccoon dog	01017VNM	EU086160	Viet Nam	2001	Dog
NeiMeng927A	EU284095	China(Jilin)	2007	Raccoon dog	01016VNM	EU086159	Viet Nam	2001	Dog
NeiMeng927B	EU284096	China(Jilin)	2007	Raccoon dog					

^a 39 full genome sequenced rabies virus strains are indicated in boldface type

region among the sites, at which the transcription sequence starts and stops, as well as the five genes are summarized in Table 3.

Eighty and fifty independent clones were randomly selected before analyzing their base sequences to confirm the 3'-end and 5'-end sequences of genomic RNA, respectively. As a result, 10 % of the sequences were truncated, whereas the remaining sequences showed complete alignment, and consequently, complete leader and trailer sequences were determined. Two conserved sequences in the 3' leader and 5' trailer were in the negative-strand RNA genomes, and they are known as cyclization motifs, which play an important role in viral RNA replication [20]. The 3'-end genomic sequence (¹ACGCTTAACAAC¹²) and the 5'-end genomic sequence (¹¹⁹²⁸ACGCTTAACAAA¹¹⁹¹⁷) were obtained, and the first 11 nucleotides of the 3' and 5' terminal regions were confirmed to be complementarily conserved.

We compared the complete KGH strain genome sequence with sequences of each of the 39 other fully sequenced rabies virus strains (Table 2) which are available from GenBank. Sequence similarity at the nucleotide level was 70.6–91.6 %, and that at the amino acid level was 82.8–97.9 % (Table 4). In particular, the KGH strain showed high similarity with NNV-RAB-H (91.6 %) and transplanted rabies virus serotype 1 (91.6 %), which are two rabies virus strains isolated in India. Less similarity was observed in RRV ON-99-2 (80.1 %) and SHBRV-10 (81.6 %) of genotype 1. When each gene of the N, P, M, G, and L proteins of the KGH strain and the 40 rabies strains were compared, the N protein showed 75–92.6 % (87.8–99.3 %), the P protein showed 63.8–91.5 % (61.6–94.6 %), the M protein showed 77.3–92.1 % (84.2–97.5 %), the G protein showed 67.3–91.2 % (72–96.8 %), and the L protein showed 74–91.9 % (86.9–98.5 %) identity ranges of nucleotide and amino acids, respectively.

Structure–function analyses

Substitutions at amino acid positions that play an important role in five viral protein-coding regions were compared using the KGH strain and the 39 fully sequenced rabies virus strains (Table 5). When the N protein of the KGH strain was compared, a unique substitution was observed only at amino acid position 394 (Tyr to Cys, described as a mutation from another rabies virus strain to the KGH strain). A total of 273 and 394 amino acids of the N protein play an important role in host innate immune response and pathogenicity, respectively [9, 21]. Furthermore, a substitution occurred in the KGH strain and in an Indian isolate (NNV-RAB-H, rabies virus serotype 1), vaccines strains (PV, RbE3-15, SRV 9, SAD B19, RV97), and RRV ON-99-2 isolates at amino acid position 135 (Ser to Phe) which is a T-cell epitope in humans and mice [22]. Antigenic sites (residues 313–337, 358–367, 374–383, and

Table 3 Genomic features and protein characteristics of KGH

Genes	mRNA characteristics (nt)				Intergenic sequence (nt)	mRNA			
	Total length	5'UTR	ORF	3'UTR		Start	Position	Stop	Position
3'UTR					58		1–58		
N	1423	11	1353	59	N–P:2 (1483–1484)	AACACCCCT	59–67	TGAAAAAAAA	1474–1482
P	990	28	894	68	P–M:5 (2476–2480)	AACACCCCT	1485–1493	TGAAAAAAAA	2467–2475
M	807	14	609	184	M–G:5 (3289–3293)	AACACCACT	2481–2489	TGAAAAAAAA	3280–3288
G	2065	26	1575	464	G–L:24 (5360–5383)	AACATCCCT	3294–3302	AGAAAAAAAA	5351–5359
L	6474	29	6384	61		AACACTTCT	5384–5392	TGAAAAAAAA	11850–11858
5'UTR					69		11859–11928		

410–413), a putative casein-type phosphorylation site (Ser 389), a residue Asn 157 and Pro 435 related to immune system evasion, and the T-helper cell epitope (residue 404–418) were conserved [9, 23–26].

Two unique substitutions were observed in the P protein (Table 5). Amino acid residues 69–177 and 268–297 are N protein binding sites [27, 28], and substitutions were observed only at amino acids 268 and 296 (Ala to Gly and Thr to Val) of the KGH strain. Five phosphorylation sites in the P protein target separate cellular kinases, and serine residue positions 63/64 code RABV protein kinase C, whereas serine positions 162/210/271 code protein kinase C [27, 29]. Among them, a substitution was uniquely observed only at residues 61/62 (Gly to Lys, Lys to Glu) in the KGH strain. Among five Met residues in the P protein, Met 1/Met 20 remains in the cytoplasm according to the CRM1-dependent nuclear export signal (NES), whereas Met 53/Met and 69/Met 83 remain in the nucleus according to the nuclear localization signal (NLS). The Met 69 residue of the KGH strain was substituted with Val from Met [30, 31]. The cytoplasmic dynein light chain (LC8)-binding motif (residues 144–148), NES (residues 49–58 and 227–232), and NLS (residues 211–214 and 260R) are conserved [31, 32].

Unique amino acid substitutions were observed in the G protein (Table 5). The amino acid residue 164–303 region plays an important role in pathogenicity [33], and substitutions were observed only at amino acids 164 (Val, Arg, Gln to Ile) of KGH and India isolated strains, and at 179 (Met to Leu) and 245 (Thr to Ala) residues in KGH strains. In the G protein, 37 and 319 residues in N-glycosylated sites and most of the antigenic sites (I, II, and III) were conserved in the KGH strain. The amino acid of G protein 333, which plays an important role in determining the virulence of the rabies virus, was Arg in the KGH strain. At 194 amino acids affecting viral pathogenicity were also conserved.

Proline-rich (PPEY) motif in M protein is called as a late budding domain or L domain and it plays an important role in virion assembly and budding and has been conserved in Rhabdoviridae, Retroviridae, and Filoviridae

families [34]. Amino acid residue 58 regulates RNA synthesis, and the highly hydrophobic region (residues 89–107) is related to apoptosis and pathogenesis [35, 36]. In these functional domains, the KGH strain has been conserved.

There are binding sites of six domains and various functional motifs in the L protein. Domain I includes tripeptide GHP (residues 372–374) and putative leucine zipper motif, LX6LX6LX6 (residues 237–258) [37–39]. Domain II (residues 544–563) and Domain III (residues 728–732), which are major functional domains form “polymerase module” [38, 39]. Domains IV and V showed the availability of host, a rNTP binding site or tyrosine kinase activity [37, 39, 40]. Dipeptide GG (residues 899–900) and DP (residues 916–917) are known as the first conserved motif of polymerase, domain IV has been conserved mostly in rhabdoviruses [37, 40]. Domain VI (residues 1705–1710) known as glycine-rich motif (GXGXG) is strongly conserved in all animal rhabdoviruses [41]. These functional six domains have been conserved in the KGH strain.

Phylogenetic analysis

The phylogenetic relationship was analyzed using the nt sequence of the complete genome (Fig. 1). Chinese isolates and two phylogenetic clades of Arctic-like viruses were constructed as main branches. The China I and II lineages appeared with the Chinese isolates. The Arctic-like viruses were divided into two clades using the N protein nt, which were the Arctic-like-1 (AL1) clade and Arctic-like-2 (AL2) clade [18]. KRH2-04, which is a rabies virus strain isolated from Korean animals, belongs to Arctic-like-2 [16, 18, 19]. When the nt sequence of the complete KGH strain genome was compared, it was found to be closely related to the Indian isolate (NNV-RAB-H) strain, which belongs to Arctic-like-1 rather than to Japanese, Russian, or Chinese isolated strains.

Table 4 Comparison of the complete genome sequence of RV KGH strain

Strain	Sequence identity (%)	
	Nucleotide	Amino acid
07240FRA	70.6	82.8
8743THA	84.6	95.9
8764THA	84.4	96.0
9147FRA	88.1	97.0
BD06	84.9	95.6
CTN181	84.9	96.0
D01	84.8	95.3
D02	84.6	95.1
RV1333	71.3	84.5
F02	84.7	95.3
F04	84.6	95.0
FJ008	84.9	95.8
FJ009	84.9	95.6
HN10	85.1	96.5
Nishigahara	86.9	95.0
NNV-RAB-H	91.6	97.9
Rabies virus serotype1	91.6	97.8
RRV ON-99-2	80.1	93.0
SHBRV-18	81.6	93.1
JX08-45	84.9	96.0
GX4	85.0	96.3
SH06	84.9	95.7
J	85.1	96.3
CQ92	85.1	96.1
CTN-1	85.1	96.3
BR-Pfx1	87.0	96.8
H-08-1320	84.9	96.2
RbE3-15-5	87.3	95.6
RBE3-15	87.4	95.6
Flury-HEP	87.7	96.3
Flury-LEP-C	87.8	96.4
ERA	87.3	95.6
Ni-CE	86.7	94.6
PV	87.2	95.5
RC-HL	86.4	94.2
RV-97	86.7	94.8
SAD B19	87.2	95.3
SRV9	87.1	95.2

The phylogenetic tree was constructed based on glyco-protein gene (Fig. 2). It was closely related to the Mongolia isolate strain, which belongs to Arctic-like-2 clade. The KGH strain was constructed with Korean rabies virus strains and is composed of four distinct sub-groups; it was most closely related with Korean strains which were isolated from the same region at same time period [16, 18].

Table 5 Comparison of amino acid substitution in the KGH strain with other rabies virus strains

Protein	Amino acid position	KGH ^a	NNV-RAB	Serotype1	Other rabies strains ^b		
N	60	L	L	L	M		
	371	T	T	T	A		
	394	C	Y	Y	Y, H, L, F		
	61	K	G	G	G, R, E, D		
P	62	E	K	K	K, Q		
	68	G	G	G	E		
	91	V	L	L	L, F		
	268	G	G	G	A		
	296	V	T	T	T, A, K, L		
	M	15	G	D	D	D	
		G	156	N	S	S	R, G
			164	I	I	I	V, R, Q
			179	L	M	M	M
			245	A	T	T	T
414			N	D	D	D	
443			V	V	V	M, L, I	
462			G	G	R	R, K	
L			156	D	E	E	E
			202	E	E	E	D
	316		R	R	R	K	
	351	G	G	G	R, K		
	431	S	T	T	A, P		
	1134	S	P	P	T		
	1280	R	K	K	K		
	1566	D	N	N	N, T		

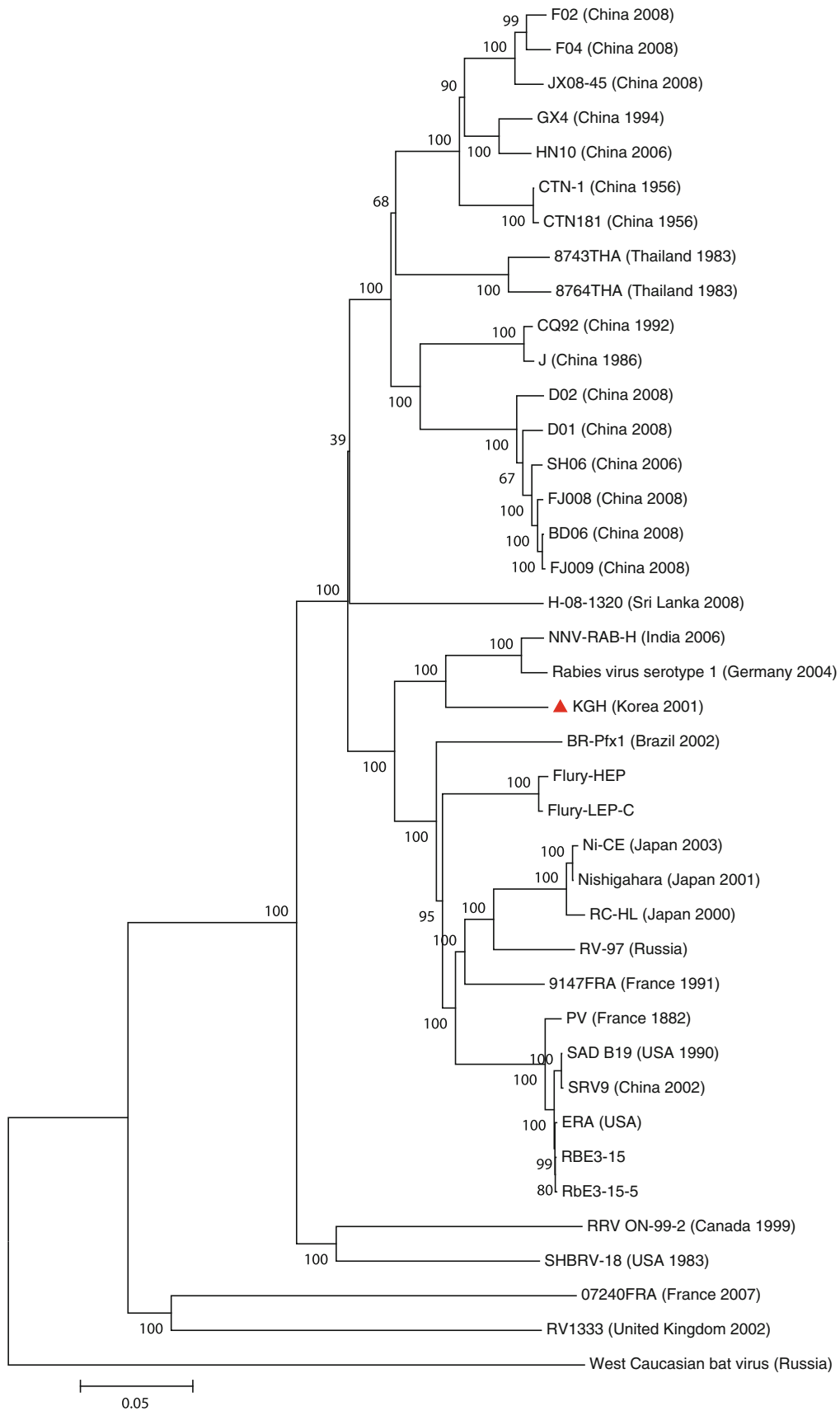
^a Unique amino acid substitution are indicated in boldface type

^b Multiple alignment of 40 fully sequenced rabies virus strains

Discussion

The purpose of this study was to identify the complete genome sequence of the KGH strain, which is a rabies virus separated from the hair follicles of a Korean patient with rabies, and to compare it with 39 available fully sequenced rabies virus strains. The KGH strain had a nt length of 11,928, which is the same as NNV-RAB-H, SAD B19, and SRV9. The KGH strain was longer than H-08-1320 (11,926 nt), BR-Pfx1(11,924 nt), and Japanese isolates (11,926 nt), but was shorter than the vaccine strains of PV(11,932 nt), ERA (11,931 nt), RBE3-15(11,931nt), and RV-97(11,932nt). Diverse variations and lengths in the G–L region affect not only full genome length but also virus proliferation by regulating gene expression [42].

A unique substitution was observed in residue 394 of the KGH strain. When the avirulent Ni-CE strain was prepared through passage of the virulent Nishigahara strain (Ni) from Japan, the 273 (Phe to Leu), 394 (Tyr to His), and 395 (Phe to Leu) residues of the N protein changed [36]. Amino



◀ **Fig. 1** Phylogenetic tree was constructed with the nucleotide sequence of the full-length genome. The tree was constructed using the neighbor-joining algorithm with the Kimura two-parameter in MEGA version 5.05. Numbers below branches are bootstrap values calculated from 1,000 replicates. West Caucasian bat virus was used as an outgroup. The KGH Korean isolate strain is marked with a *red triangle* (Color figure online)

acids 273 and 394 of the N protein play an important role in evading RIG-I-mediated antiviral responses and have an important relationship with pathogenicity in adult mice [9, 21]. Residue 394 of the Ni and Ni-CE strains was a hydrophilic Tyr or His, whereas the KGH strain had a hydrophobic Cys. This may be due to the effects on N protein structure and residues 373–395 [22], which are T-cell epitopes, and subsequently, interferon (IFN)- β and the expression of CXCL10 [43], which is regulated by IRF-3, may be affected. Host antiviral responses may also be affected, as residue 394 of the N protein includes the P protein binding domain [36]. Rabies virus strains isolated from Korean animals include the 394 residues of the N protein with subgroup A (Gyeonggi-do province) of Tyr and subgroup B (Gangwon-do province) of Cys [15]. The effects of substituting residue 394 in the KGH strain on immune response will require further investigation.

Two unique substitutions were observed in the P protein of the KGH strain. In rabies viruses, the N and P proteins are expressed, N–P complexes are formed, and viral genomic RNA is combined to inhibit N protein phosphorylation [44]. N phosphorylation is important for regulating viral RNA transcription and replication [45]. The P protein is present in two independent N-binding domains. One P protein functions as a chaperone by combining with newly formed non-RNA bound N protein (N⁰) at residues 69–177 of weaker N-binding sites, whereas another one is positioned at residues 268–297 and combines with N of the N-RNA complex [30, 46].

Ito et al. [32] confirmed the STAT1-binding domain at amino acid positions 267–297. Substitutions at residues 268 and 296 of the P protein in the KGH strain were found. Residue 268 is substituted from Ala to Gly only in the Korean and Indian isolates. Only residue 296 of the KGH strain was substituted from Thr to Val. Therefore, IFN production and its relationship with microtubule-dependent mechanisms may have been affected according to STAT1 regulation by substitution of these residues [47].

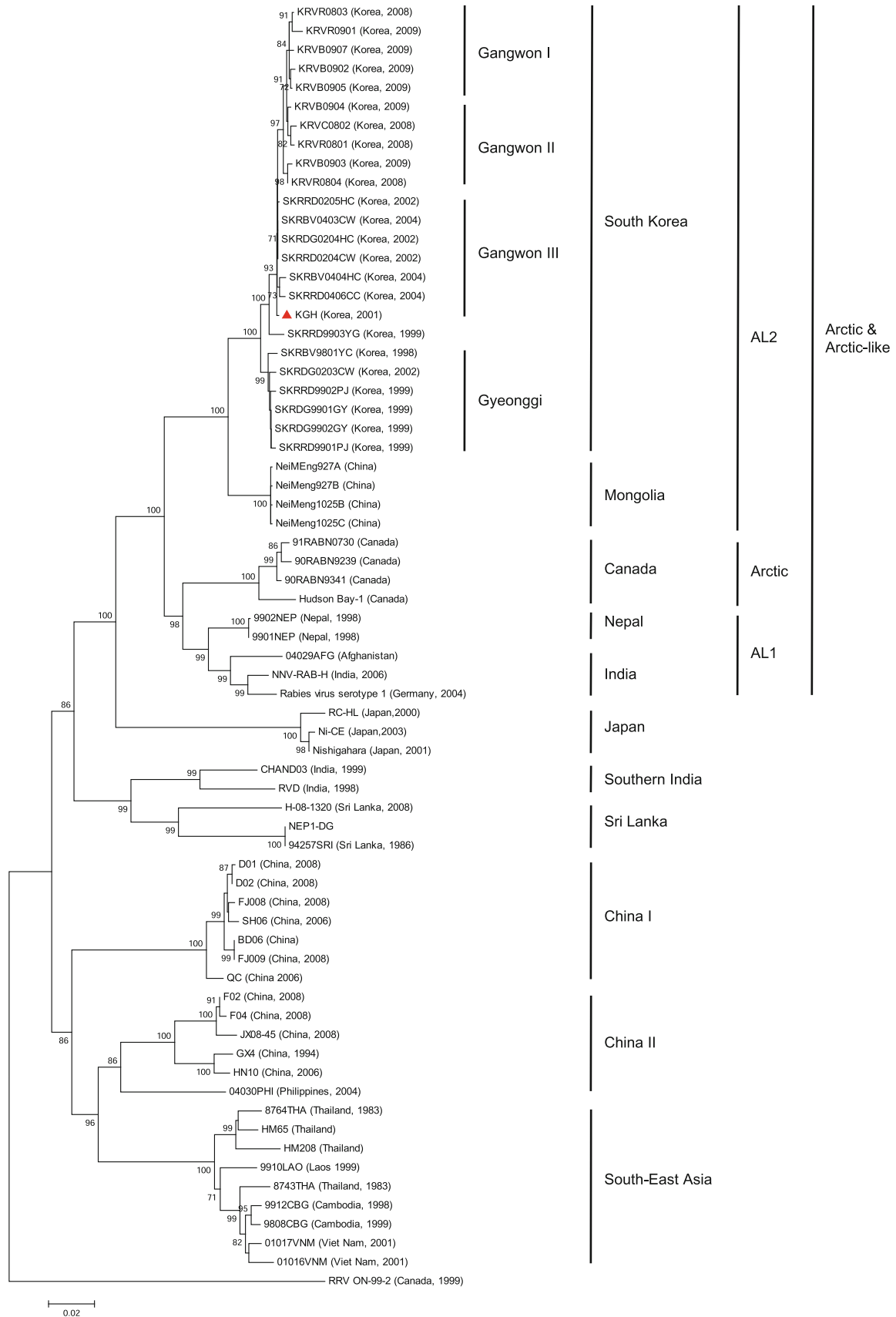
Takayama-Ito et al. [33] reported that the region at the position 164–303 of glycoprotein is related to pathogenicity in adult mice. Since then, another study has shown that glycoprotein 268 (Ile) residue plays a very important role in pathogenicity, and 242 (Ala), 255 (Asp), and 268 (Ile) may cooperate to the virulence of the Nishigahara strains [48]. Further studies are needed to elucidate the association of pathogenicity upon the substitution of glycoprotein

164 (Ile), 179 (Leu), and 245 (Ala) residues in KGH strain. No specific substitutions were observed in the M or L proteins.

According to Kuzmin et al. and Shao et al. [18, 19], the Arctic-like-1 clade is composed of regions in Iraq, Iran, Pakistan, Nepal, and India, whereas the Arctic-like-2 clade is composed of areas of south-eastern Siberia, Russian Far East, Japan, Inner Mongolia, and Korea. According to the complete genome sequence comparison in the present study, the Indian isolates of NNV-RAB-H and rabies virus serotype 1 (AY956319) were confirmed to have 91.6 % homology. In contrast, the G nt sequences of the rabies virus strain isolated in Korea have been confirmed to be more closely related to, and transmitted from, the NeiMeng1025C, NeiMeng1025B, NeiMeng927A, and NeiMeng927B isolates in northeastern China (Jilin province in China), than those in India (Fig. 2). Based on this result, China was verified as the primary source of Asian rabies events. Additionally, northern India and adjacent countries, such as Afghanistan, Nepal, Inner Mongolia, and Korea, have been confirmed to have common virus ancestors with the Arctic-like strain [16, 17, 49].

In Korea, animal rabies viruses that are found in limited areas of Gangwon-do and Gyeonggi-do provinces comprise four subgroups (Gangwon I, II, III, and Gyeonggi) [16]. Among the multiple animal rabies virus isolates, partial sequences of the N, P, and G genes as well as the G–L region have been reported [15, 50]. When the KGH strain was compared with the 23 glycoprotein gene sequence of rabies virus strains separated in Korea, the most closely related were determined to be SKRDG0204HC and SKRRD0205HC, which were isolated from dogs and raccoon dog in the Hwacheon area in 2002, around the same period when the patient was bitten (2001), and which belongs to the Gangwon III subgroup (Fig. 2). These results suggest that the rabies virus circulating in South and North Korea is a variety of viral variants, with features based on factors such as topography, animal species, and circulation year [18]. Although the complete genome sequence of Korean animal isolates is unknown, specific mutations in the N, P, and G genes were confirmed, and unique mutations were observed only in 342 nt (A to G) and 854 nt (C to T) in the G–L region.

In conclusion, we have provided the complete nucleotide and deduced amino acid sequence of a KGH rabies virus strain isolated from a Korean patient for the first time, and full sequence and genetic features of diverse rabies virus genomes were compared. In the N, P, and G protein functional region, the KGH strain was confirmed to have a unique substitution at sites related to the host immune response and pathogenicity. These results allow for an assessment of the interaction with host cells at the molecular level, and aid the evaluation of viral protein functions through direct manipulation of the viral genome. This



◀ **Fig. 2** Phylogenetic tree constructed with the G gene sequence of 68 selected rabies virus strains isolated from Asia. The KGH Korean isolate strain is marked with a *red triangle*. The trees were used together with the Kimura two-parameter neighbor-joining algorithm in MEGA version 5.05. Numbers below branches are bootstrap values calculated from 1,000 replicates (Color figure online)

result indicates that diverse genomic modifications occur among circulating rabies virus strains in other countries as well as in Korea. We expect that the complete genome analysis of the KGH strain will help us, not only to understand the phylogenetic relationships among rabies viruses strains all over the world, but also to select a candidate for vaccine development.

Acknowledgments This study supported by grant no. 2010-N53001-00 from the Korea National Institute of Health.

References

1. A.C. Jackson, W.H. Wunner, in *Rabies*, ed. by G.M. Baer, (Academic press, Elsevier, London, 2007)
2. H. Bourhy, J.-M. Reynes, E.J. Dunham, L. Dacheux, F. Larrous, V.T.Q. Huong, G. Xu, J. Yan, M.E.G. Miranda, E.C. Holmes, *J. Gen. Virol.* **89**(Pt 11), 2673–2681 (2008)
3. E.C. Holmes, C.H. Woelk, R. Kassis, H. Bourhy, *Virology* **292**, 247–257 (2002)
4. R.H. Fischman, F.E. Ward 3rd, *Am. J. Epidemiol.* **88**(1), 132 (1968)
5. M.J. Schnell, J.P. McGettigan, C. Wirblich, A. Papaneri, *Nat. Rev. Microbiol.* **8**, 51–61 (2010)
6. World Health Organization, WHO Expert consultation on rabies. *World. Health. Organ. Tech. Rep. Ser.* **931**, 1–88 (2005)
7. A.R. Gould, A.D. Hyatt, R. Lunt, J.A. Kattenbelt, S. Hengstberger, S.D. Blacksell, *Virus Res.* **54**, 165–187 (1998)
8. X. Wu, R. Franka, A. Velasco-Villa, C.E. Rupprecht, *Virus Res.* **129**, 91–103 (2007)
9. T. Masatani, N. Ito, K. Shimizu, Y. Ito, K. Nakagawa, M. Abe, S. Yamaoka, M. Sugiyama, *Virus Res.* **155**, 168–174 (2011)
10. L. Tao, J. Ge, X. Wang, H. Zhai, T. Hua, B. Zhao, D. Kong, C. Yang, H. Chen, Z. Bu, *J. Virol.* **84**(17), 8926–8936 (2010)
11. Y. Kobayashi, Y. Suzuki, T. Itou, A.A.B. Carvalho, E.M.S. Cunha, F.H. Ito, T. Gojobori, T. Sakai, *Infect. Genet. Evol.* **10**, 278–283 (2010)
12. Y.Z. Zhang, C.L. Xiong, X.D. Lin, D.J. Zhou, R.J. Jiang, Q.Y. Xiao, X.Y. Xie, X.X. Yu, Y.J. Tan, M.H. Li, Q.S. Ai, L.J. Zhang, Y. Zu, C. Huang, Z.F. Fu, *Infect. Genet. Evol.* **9**, 87–96 (2009)
13. A.G. Szanto, S.A. Nadin-Davis, B.N. White, *Virus Res.* **136**, 130–139 (2008)
14. T. Nagaraja, S. Madhusudana, A. Desai, *Virus Genes* **36**, 449–459 (2008)
15. B.H. Hyun, K.K. Lee, I.J. Kim, K.W. Lee, H.J. Park, O.S. Lee, S.H. An, J.B. Lee, *Virus Res.* **114**, 113–125 (2005)
16. D.K. Yang, Y.N. Park, G.S. Hong, H.K. Kang, Y.I. Oh, S.D. Cho, J.Y. Song, *J. Vet. Sci.* **12**, 57–63 (2011)
17. S. Meng, Y. Sun, X. Wu, J. Tang, G. Xu, Y. Lei, J. Wu, J. Yan, X. Yang, C.E. Rupprecht, *Virology* **410**, 403–409 (2011)
18. I.V. Kuzmin, G.J. Hughes, A.D. Botvinkin, S.G. Gribencha, C.E. Rupprecht, *Epidemiol. Infect.* **136**, 509–519 (2008)
19. X.Q. Shao, X.J. Yan, G.L. Luo, H.L. Zhang, X.L. Chai, F.X. Wang, J.K. Wang, J.J. Zhao, W. Wu, S.P. Cheng, F.H. Yang, X.C. Qin, Y.Z. Zhang, *Epidemiol. Infect.* **139**, 629–635 (2010)
20. J. Du, Q. Zhang, Q. Tang, H. Li, X. Tao, K. Morimoto, S.A. Nadin-Davis, G. Liang, *Virus Res.* **135**, 260–266 (2008)
21. T. Masatani, N. Ito, K. Shimizu, Y. Ito, K. Nakagawa, Y. Sawaki, H. Koyama, M. Sugiyama, *J. Virol.* **84**, 4002–4012 (2010)
22. B. Dietzschold, M. Lafon, H. Wang, L. Otvos Jr, E. Celis, W.H. Wunner, H. Koprowski, *Virus Res.* **8**(2), 103–125 (1987)
23. H. Goto, N. Minamoto, H. Ito, N. Ito, M. Sugiyama, T. Kinjo, A. Kawai, *J. Gen. Virol.* **81**, 119–127 (2000)
24. I.V. Kuzmin, L.A. Orciari, Y.T. Arai, J.S. Smith, C.A. Halon, Y. Kameoka, C.E. Rupprecht, *Virus Res.* **97**, 65–79 (2003)
25. L. Geue, S. Schares, C. Schnick, J. Kliemt, A. Beckert, C. Freuling, F.J. Conraths, B. Hoffmann, R. Zanoni, D. Marston, L. McElhinney, N. Johnson, A.R. Fooks, N. Tordo, T. Müller, *Vaccine* **26**, 3227–3235 (2008)
26. A.A.V. Albertini, A.K. Wernimont, T. Muziol, R.B.G. Ravelli, C.R. Clapier, G. Schoehn, W. Weissenhorn, R.W.H. Ruigrok, *Science* **313**, 360–363 (2006)
27. M. Chenik, K. Chebli, Y. Gaudin, D. Blondel, *J. Gen. Virol.* **75**, 2889–2896 (1994)
28. A.K. Gupta, D. Blondel, S. Choudhary, A.K. Banerjee, *J. Virol.* **74**, 91–98 (2000)
29. O. Delmas, R. Assenberg, J.M. Grimes, H. Bourhy, *RNA Biol.* **7**(3), 322–327 (2010)
30. G. Castel, M. Chtéoui, G. Caignard, C. Prêhaud, S. Méhouas, E. Réal, C. Jallet, Y. Jacob, R.W.H. Ruigrok, N. Tordo, *J. Virol.* **83**, 10808–10820 (2009)
31. D. Pasdeloup, N. Poisson, H. Raux, Y. Gaudin, R.W.H. Ruigrok, D. Blondel, *Virology* **334**, 284–293 (2005)
32. N. Ito, G.W. Moseley, D. Blondel, K. Shimizu, C.L. Rowe, Y. Ito, T. Masatani, K. Nakagawa, D.A. Jans, M. Sugiyama, *J. Virol.* **84**, 6699–6710 (2010)
33. M. Takayama-Ito, N. Ito, K. Yamada, N. Minamoto, M. Sugiyama, *J. Neurovirol.* **10**, 131–135 (2004)
34. A. Guleria, M. Kiranmayi, R. Sreejith, K. Kumar, S.K. Sharma, S. Gupta, *J. Biosci.* **36**, 929–937 (2011)
35. S. Finke, K.K. Conzelmann, *J. Virol.* **77**, 12074–12082 (2003)
36. K. Shimizu, N. Ito, T. Mita, K. Yamada, J. Hosokawa-Muto, M. Sugiyama, N. Minamoto, *Virus Res.* **123**, 154–160 (2007)
37. O. Poch, B.M. Blumberg, L. Bougueleret, N. Tordo, *J. Gen. Virol.* **71**, 1153–1162 (1990)
38. D.A. Marston, L.M. McElhinney, N. Johnson, T. Müller, K.-K. Conzelmann, N. Tordo, A.R. Fooks, *J. Gen. Virol.* **88**, 1302–1314 (2007)
39. I.V. Kuzmin, X. Wu, N. Tordo, C.E. Rupprecht, *Virus Res.* **136**, 81–90 (2008)
40. M.A. McClure, *J. Perrault, Virology* **172**, 391–397 (1989)
41. P. Ming, J. Du, Q. Tang, J. Yan, S.A. Nadin-Davis, H. Li, X. Tao, Y. Huang, R. Hu, G. Liang, *Virus Res.* **143**, 6–14 (2009)
42. M.J. Schnell, T. Mebatsion, K.K. Conzelmann, *EMBO J.* **13**, 4195–4203 (1994)
43. K. Honda, T. Taniguchi, *Nat. Rev. Immunol.* **6**, 644–658 (2006)
44. P. Liu, J. Yang, X. Wu, Z.F. Fu, *J. Gen. Virol.* **85**, 3725–3734 (2004)
45. X. Wu, X. Gong, H.D. Foley, M.J. Schnell, Z.F. Fu, *J. Virol.* **76**, 4153–4161 (2002)
46. S. Finke, K. Brzóka, K.K. Conzelmann, *J. Virol.* **78**, 12333–12343 (2004)
47. G.W. Moseley, X. Lahaye, D.M. Roth, S. Oksayan, R.P. Filmer, C.L. Rowe, D. Blondel, D.A. Jans, *J. Cell Sci.* **122**, 3652–3662 (2009)
48. M. Takayama-Ito, N. Ito, K. Yamada, M. Sugiyama, N. Minamoto, *Virus Res.* **115**, 169–175 (2006)
49. T. Matsumoto, K. Ahmed, O. Wimalaratne, K. Yamada, S. Nanayakkara, D. Perera, D. Karunanayake, A. Nishizono, *Arch. Virol.* **156**, 659–669 (2011)
50. Y.J. Park, M.K. Shin, H.M. Kwon, *Virus Genes* **30**, 341–347 (2005)