Published in final edited form as:

J Gen Virol. 2016 February ; 97(2): 389–402. doi:10.1099/jgv.0.000349.

Full genome characterization of the first G3P[24] rotavirus strain detected in humans provides evidence of interspecies reassortment and mutational saturation in the VP7 gene

Slavica Mijatovic-Rustempasic[1], **Sunando Roy**[1], **Elizabeth N. Teel**[1], **Geoffrey A. Weinberg**[2], **Daniel C. Payne**[1], **Umesh D. Parashar**[1], and **Michael D. Bowen**[1]

[1]Gastroenteritis and Respiratory Viruses Laboratory Branch, Division of Viral Diseases (DVD), National Center for Immunization and Respiratory Diseases (NCIRD), Centers for Disease Control and Prevention, 1600 Clifton Road NE, Atlanta, GA 30329-4027, USA

[2]Department of Pediatrics, University of Rochester School of Medicine and Dentistry, 601 Elmwood Avenue, Box 690, Rochester, NY 14642, USA

Abstract

During the 2008–2009 rotavirus season of the Centers for Disease Control and Prevention New Vaccine Surveillance Network, one case of paediatric acute gastroenteritis associated with a rotavirus G14P[24] strain was identified. This was the first detection of the genotype G14 and P[24] in humans, and the first detection of the G14P[24] combination. To gain an insight into the origins and the evolution of this strain, we determined the complete ORF sequences of all 11 genes. A majority of the genes identified were similar to the simian strain TUCH, except for the VP1 and VP7 genes that clustered only distantly with the bovine and equine strains, respectively. In addition, this strain carried AU-1-like NSP2 and NSP4 genes. Using codon-partitioning and protein-based phylogenetic approaches, we determined that the VP7 genotype of strain 2009727118 was actually G3; therefore, the proposed full genomic classification of the 2009727118 strain is G3-P[24]-I9-R2-C3-M3-A9-N3-T3-E3-H6. These findings indicate the possibility that the 2009727118 strain originated by interspecies transmission and multiple reassortment events involving human, bovine and equine rotaviruses, resulting in the introduction of some genes into the genome of simian rotaviruses. Additionally, we found evidence of mutational saturation in the third codon position of the VP7 ORF which presented an issue with homoplasy in phylogenetic analyses.

Introduction

Group A rotaviruses (RVA) are the major cause of diarrhoea in young children and animals worldwide, and are estimated to cause ∼450 000 deaths annually in children <5 years of age (Tate *et al.*, 2011). RVA belongs to the family *Reoviridae* and possesses a segmented dsRNA genome composed of 11 segments encoding six structural proteins (VPs) and five or six non-

Correspondence Michael D. Bowen mkb6@cdc.gov.

The GenBank/EMBL/DDBJ accession numbers for the 11 gene sequences of strain RVA/Human-wt/USA/2009727118/2009/ G3P[24] are KF541281–KF541291.

structural proteins (NSPs) (Estes & Kapikian, 2007). The traditional binomial classification of RVA was based upon the serological or genetic characterization of the two outer capsid proteins VP4 (protease sensitive protein, the P-type determinant) and VP7 (glycosylated protein, the G-type determinant) (Estes & Kapikian, 2007). Based on extensive serological and genomic studies, at least 28 G and 39 P genotypes have been identified globally (http://rega.kuleuven.be/cev/viralmetagenomics/virus-classification). However, only five common G types (G1–G4 and G9) and three common P types (P[4], P[6] and P[8]) account for 80 % of infections in humans (Gentsch *et al.*, 2005; Patel *et al.*, 2011; Santos & Hoshino, 2005).

An extended RVA classification and nomenclature system was established by the Rotavirus Classification Working Group (RCWG) to define genotypes for each genome segment (Matthijnssens *et al.*, 2011a). The notation for the VP7, VP4, VP6, VP1–3, NSP1–5/6 genome segments was defined as Gx-P[x]-Ix-Rx-Cx-Mx-Ax-Nx-Tx-Ex-Hx, with x indicating the numbers of the corresponding genotypes (Matthijnssens *et al.*, 2011a). Currently, at least 28 G, 39 P, 21 I, 14 R, 14 C, 13 M, 24 A, 14 N, 16 T, 21 E and 16H genotypes have been described (http://rega.kuleuven.be/ cev/viralmetagenomics/virus-classification). Worldwide, the majority of human RVA strains possess either the Wa-like genogroup 1 constellation (I1-R1-C1-M1-A1-N1-T1-E1-H1) or the DS-1-like genogroup 2 constellation (I2-R2-C2-M2-A2-N2-T2-E2-H2) that are believed to have a common origin with the porcine and the bovine RVAs, respectively (Matthijnssens *et al.*, 2008). A smaller group of RVA strains have the AU-1-like genogroup 3 constellation (I3-R3-C3-M3-A3-N3-T3-E3-H3) that is believed to have originated from feline RVAs (Matthijnssens *et al.*, 2008, 2010). Currently, genotypes are assigned based on percentage nucleotide identity of at least 50 % length of the ORF.

A novel genotype, G14 (prototype strain FI23), was designated based on cross-neutralization studies and VP7 sequence in 1991 (Browning *et al.*, 1991). The G14 genotype is commonly found in horses, usually in combination with the P[12] genotype. Recently, a study reporting the complete genome sequence of four equine RVA G14 strains (G14-P[12]-I2-R2-C2-M3-A10-N2-T3-E2/12-H7) suggested that the vast majority of currently circulating equine RVA strains found on different continents are highly conserved with only limited genetic diversity (Matthijnssens *et al.*, 2012).

The first P[24] strain identified was isolated from an asymptomatic rhesus macaque in 2002 (McNeal *et al.*, 2005). This strain, designated TUCH, was identified in combination with the G3 genotype. The full genome of strain TUCH has been determined to be G3-P[24]-I9-R3-C3-M3-A9-N1-T3-E3-H6 (Matthijnssens *et al.*, 2010). Full genome analysis revealed that a possible origin of this strain is a result of the reassortment events between feline, canine and human RVA, and that overall strain TUCH is more closely related to the feline-like human strain AU-1 than to Wa- and DS-1-like RVA. Nucleotide-based phylogeny presented by Matthijnssens *et al.* (2010) reports the TUCH strain as a part of the G3 genotype clade; however, inclusion in this clade was not supported by a strong bootstrap value (<50 %).

During active surveillance, we found a novel genotype combination, G14P[24], in the stool of a 3-year-old child with diarrhoea from Rochester, NY, USA (Weinberg *et al.*, 2013). The child had a possible animal exposure at a petting zoo and had not received rotavirus vaccine.

Enteric adenovirus was detected also in the stool sample. Genotype G14 was determined, using the current RCWG guidelines, by analysing the strain using RotaC 2.0 (http://rotac.regatools.be/). To gain insight into the origins and evolution of this strain, designated RVA/Human-wt/USA/2009727118/2009/G14P[24] (2009727118), we determined the complete ORF sequences of all 11 genes, and subjected them to genetic and phylogenetic analyses. We found that the strain was actually a G3P[24] genotype and that mutational saturation in the third codon position of the VP7 ORF presented an issue with homoplasy in phylogenetic analyses.

Results

Complete ORF sequences were determined for strain 2009727118 and are described in the following text.

VP7

The complete ORF for the VP7 gene (981 bp, 327 aa) of strain 2009727118 was determined. When the full VP7 ORF was analysed using BLAST, strain 2009727118 shared 87 % identity to human G3 strains with a potential bat origin (Xia *et al.*, 2014), G3P[10] CMH079/05 (Genbank accession number EU791924) and G3P[X] strain CMH222 (Genbank accession number AY707792); 86 % identity to animal G3 strains Rhesus (Genbank accession number HQ665466), simian RVA/Simian-tc/USA/RRV/ 1975/G3P[3] (Genbank accession number EU636932) and equine G3 strain ERV316 (Genbank accession number L49043); 84 % identity with some equine G14 strains and 79 % with TUCH (Genbank accession number FJ816615) strain. However, genotype G14 was assigned for this strain based on the RotaC 2.2b automated genotyping tool for RVA. The VP7 gene of this strain clusters distantly with G14 strains of equine origin, and does not cluster with G3 strains of human, equine and simian origin (Fig. 1). In addition, this strain has a higher nucleotide but lower amino acid identity with G14 reference strain E403 (84.1 and 88.7 %, respectively) compared with the G3 reference strain AU-1 (82.4 and 91.4 %, respectively). Nucleotide similarities between strain 2009727118 and both G14 and G3 strains fall within the 80 % cut-off value necessary for designating a genotype as defined by the extended classification system (Fig. 2). The nucleotide and amino acid identities between the VP7 genes of TUCH and AU-1 are 78.2 and 87.1 %, respectively.

Amino acid alignment of the VP7 gene

Examination of the antigenic regions of the VP7 protein showed that strain 2009727118 has 6 aa (92, 94, 96, 146, 147 and 221) in these regions that are almost identical to G3 strains and differ from G14 strains (Fig. 3).

Recombination analysis of the VP7 gene

Testing for potential recombination in the VP7 gene was accomplished by GARD analysis found in the Datamonkey webserver to detect breakpoints. A single breakpoint was identified at position 615 for the entire alignment, but the two different trees still showed identical clustering for 2009727118 strain so recombination was not indicated. In addition,

Simplot analysis of our sequence alignment confirmed that there was no evidence of recombination.

Measurements of phylogenetic conflict and saturation for the VP7 gene

Given that the strain analysed in this report clustered distantly with G14 strains positioned in the G3 cluster, we used the codon-partitioning and protein-based phylogenetic approaches in order to recover the underlying phylogeny. Fig. 4 shows unrooted phylogenetic trees of the 2009727118 strain and other strains included in Fig. 1. We hypothesized that the strains found on long branches observed in the unrooted tree inferred from all three codon positions (Fig. 4a) might have accumulated so many substitutions that the resolution of the internode would be random with respect to the historical evolution of the strains analysed.

To measure saturation in substitutions, we plotted corrected pairwise distances using the Kimura two-parameter distance model against the proportion of transitions and transversions taking into account all three codon positions, the first position, the second position and the third codon position for VP7 (Fig. 5a–d) and VP3 (Fig. 6 a–d) gene segments. The VP3 gene segment was chosen as a control for analysis and because there are a substantial number of full ORFs sequences available in GenBank. Transition distances for the VP7 gene segment are saturated with respect to distances when all three codons are analysed (Fig. 5a). They increase linearly for the first and second codon positions (Fig. 5b, c), but the saturation effect is observed strongly for the third codon of the VP7 gene (Fig. 5d), which affects the transition pattern when all three codons are analysed (Fig. 5a). In contrast, the transition and transversion distances for the VP3 gene increase linearly when all three, as well as individual first, second or third codons are analysed, and therefore no saturation was observed (Fig. 6a–d).

To determine if the third codon changes can skew the evolutionary relationship between our strain and other G3 and G14 strains, we compared unrooted phylogenetic trees using all three VP7 gene codon positions (Fig. 4a), first and second codon positions only (Fig. 4b), and amino acid sequences (Fig. 4c). The latter two analyses eliminated the contribution of saturated third codon position bases. As observed in Fig. 4(a), when all three codons are used, strain 2009727118 fell between the G3 and G14 clades. In the first and second codon position tree and amino acid tree, G3 and G14 strains formed distinct, bootstrap-supported clusters and our strain clustered within the G3 clade. Therefore, the effect of the third base saturation did appear to contribute to unresolved clustering of G3 and G14 strains in a VP7 tree (Figs 1 and 4a).

VP4

The complete ORF for the VP4 gene (2331 bp, 777 aa) of strain 2009727118 was determined. When the full ORF was analysed using BLAST, this strain had a maximum identity of 90 % with other P[24] strains in GenBank that were partially sequenced from original simian stool (GenBank accession number AY596189) and fully sequenced from the tissue culture-adapted strain RVA/Rhesus-tc/USA/ TUCH/2002/G3P[24] (GenBank accession number FJ816611). Genotype P[24] was assigned to the sequence by RotaC 2.2b. Based on the extended genotyping classification system and cut-off value for assigning the

genotype for the VP4 gene (>80 %), the genotype of strain 2009727118 is P[24]. The results of the panel show that VP4 clusters with the only P[24] strain recorded to date, TUCH (Fig. 7).

VP1

The VP1 gene (3267 bp, 1089 aa) of strain 2009727118 clusters with R2 strains of bovine, simian and human origin (WI79-9, SC2-9, WC3, BrB-9 and WI79-4) with nucleotide and amino acid similarities of 96.4 and 99.1– 98.9 %, respectively (Fig. 7).

NSP2

The complete ORF for the NSP2 gene (954 bp, 318 aa) is AU-1-like or belongs to the N3 genogroup with high nucleotide similarities to G3P[9] strain CU-365-KK from Thailand, equine strain E3198, human strain T152 and reference strain AU-1 (Fig. 7).

NSP4

The NSP4 gene (528 bp, 176 aa) clusters with two human E3 strains detected in Thailand, with the highest nucleotide and amino acid similarity of 90.3 and 99.4 %, respectively, to the CMH222 strain.

Other genes

All other genes, VP2 (2670 bp, 890 aa), VP3 (2508 bp, 836 aa), VP6 (1194 bp, 398 aa), NSP1 (1479 bp, 493 aa), NSP3 (942 bp, 314 aa) and NSP5 (594 bp, 198 aa), have the highest nucleotide (97.1, 96.6, 96.1, 96.0, 98.0 and 98.7 %) and amino acid similarities (99.0, 97.7, 100, 95.7, 97.4 and 98.0 %) to the TUCH strain, respectively, and, as shown in Fig. 7, they cluster with the TUCH strain in each case for these other genes.

Full genome constellation

Based on the extended classification system for RVA, the full genotype constellation for strain 2009727118 is G3-P[24]-I9-R2-C3-M3-A9-N3-T3-E3-H6.

Discussion

Although > 73 G/P combinations have been described to date (Matthijnssens *et al.*, 2009), this is the first detection of genotype G3P[24] in humans. Phylogenetic analyses of all 11 genes revealed a puzzling configuration of genes suggestive of extensive genetic reassortment events between various animal and human RVA to generate this unusual strain. The full genome classification of strain 2009727118 indicates that this strain was likely generated by genetic reassortment of VP1, VP7, NSP4 and NSP2 genes onto the RVA strains of simian origin in either a human or simian host. It is possible, however, that this strain may represent a conventional strain from a yet-to-be identified host rather than a multiple reassortant virus.

Overall, G3 and G14 genotypes are closely related, and differentiating between G3 and G14 genotypes for undefined strains can be quite challenging. The first G14 genotype, strain FI23, although distinct, was identified as having only 6 aa (92, 94, 96, 146, 147, and 221) in

the VP7 antigenic regions that differ from G3 equine RVA (Browning *et al.*, 1991). The strain characterized in this report, i.e. 2009727118, has amino acids identical to those of equine, simian and human G3 strains, but not equine G14 strains. Strain 2009727118 has seven genes that are closely related to the TUCH strain. In the original report characterizing the TUCH strain, nucleotide sequences were determined for VP4 and VP6 genes, whilst the VP7 gene was serotyped as G3 by ELISA with VP7-specific mAbs against G1, G2, G3 and G4 RVA strains (McNeal *et al.*, 2005). Phylogenetic analysis of the VP7 gene of tissue culture-adapted TUCH strain was ambiguous; therefore, it was assigned to the G3 genotype based on the previously mentioned serotyping data (Matthijnssens *et al.*, 2010). When we considered BLAST results and amino acid residues that differentiate G14 strains from G3 strains, we concluded that the VP7 genotype of 2009727118 was G3 like TUCH. Additionally, based on BLAST analysis, this strain has a higher maximum identity to G3 strains than to G14 strains. However, strain 2009727118 shares a higher nucleotide identity to G14 reference strain E403 (GenBank accession number JF12582) rather than the G3 reference strain AU-1 (GenBank accession number D86271) and, in this case, RotaC designated the VP7 of strain 2009727118 as G14. Based on the widely accepted extended genotyping classification system developed by RCWG and the RotaC classification tool, genotypes are designated by comparing nucleotide identity of the unknown genotype to a reference alignment. The RCWG's extended genotyping classification system is a useful tool because it standardizes the approach for genotyping rota-virus strains. However, we have to consider other approaches when strains like the one analysed here prove to be challenging. In rare cases (e.g. the strain from this study), additional genetic analysis will be necessary to ascertain the RVA genotype. The insertion of additional G3 reference strains (e.g. CMH079/05 and CMH222) to the RotaC database also may resolve this issue.

Nearly all RNA viruses exhibit very rapid evolutionary change in order to escape their host's immune system or to infect different species (Duchêne *et al.*, 2014). In the coding region, changes at the third codon position occur at a higher frequency compared with the first and second positions. Molecular analysis of viral evolution generally relies on estimated rates of nucleotide substitutions that may change by several orders of magnitude over time. More specifically, molecular data acquired over a short time frame demonstrate evolutionary changes at higher rates than those data sampled over longer time periods (Charrel *et al.*, 1999; Sanjuán *et al.*, 2010). Saturated data refers to a set of data for which the phylogenetic signal is overwhelmed by multiple changes at each site (Philippe *et al.*, 2011). In that case, similarities between strains are likely to be the result of homoplasy, rather than homology or common ancestry. In other words, when full saturation is reached the phylogenetic signal is obscured, meaning that sequences are no longer entirely informative about virus evolution. One way to avoid this problem is to exclude the third codon position from the analysis or to analyse corresponding amino acid sequences, as we have done in this study.

The evolution and genetic diversity of RVA are driven by point mutations, interspecies transmission, reassortment and intragenic recombination (Gentsch *et al.*, 2005). Genetic reassortment is common due to the segmented nature of RVA. Even though RVA strains are generally species specific, there are many reports of human infection by animal/human reassortant strains and a few examples of direct interspecies transmission of animal strains to humans (Degiuseppe *et al.*, 2013; Esona *et al.*, 2010; Matthijnssens *et al.*, 2011b; Mladenova

et al., 2012). In recent years, reports of reassortants having more than one gene segment from animal and human RVA have been increasing. Evidence of reassortment between human and animal RVA strains has been shown in different genotypes and different genome segments (Dóró *et al.*, 2015; Gentsch *et al.*, 2005; Ghosh & Kobayashi, 2011). This unusual G3P[24] RVA strain 2009727118 is possibly the result of direct interspecies transmission from a simian to a human via a strain carrying equine- and bovine-like genes. Another plausible theory is that this strain has a simian origin, was transmitted to humans, and reassorted with equine- and bovine-like genes during infections in humans. As the identity of the original host of strain 2009727118 is unknown, we can only speculate as to the source of this strain's genetic constellation and the role of reassortment in its production.

In conclusion, the findings in this report add to the growing body of evidence in support of zoonotic transmission of RVA and multiple reassortment events. The full genome studies are valuable for understanding the ecology of RVA, but also help in accumulating knowledge about the genetic diversity and origin of uncommon RVA genotypes, which in turn may be important for further development of the currently licensed RVA vaccines. The findings of this study highlight the need for continuous monitoring of RVA strains and timely recognition of novel or rare genotypes.

Methods

Stool sample

The stool sample was collected during the 2008–2009 RVA season from a 36-month-old girl at Golisano Children's Hospital at University of Rochester Medical Center (Rochester, NY, USA) and submitted to the CDC (Atlanta, GA, USA) for testing as part of the New Vaccine Surveillance Network (Weinberg *et al.*, 2013).

RNA extraction

After preparing a 10 % (v/v) faecal suspension of stool specimen in PBS, suspensions were clarified by centrifugation at 3000 r.p.m. for 10 min using Fisher Scientific accuSpin Micro 17 centrifuge with 24-place rotor and bio lid (9.5 L × 8.9 W × 13.8 H inches). RVA dsRNA was extracted from the clarified suspension using a silica-based method on the automated NulciSens extractor (bioMérieux) or the KingFisher Flex Magnetic Particle Processor (Thermo Fisher Scientific) following manufacturers' protocols. The extracted RNA was stored at –80 °C until testing.

Reverse transcription (RT)-PCR

The VP7 and VP4 genes were first amplified by RT-PCR using 9Con1-L with VP7-RDg primers and con3 with con2 primers, respectively (Das *et al.*, 1994; Gentsch *et al.*, 1992). Subsequently, individual gene fragments were amplified using gene-specific primers published previously or designed specifically for this study (Table 1) (Bányai *et al.*, 2009; Kerin *et al.*, 2007; Mijatovic-Rustempasic *et al.*, 2011). Additionally, the VP4F 1–9 forward primer (GGCTATAAA) with an M13 extension (TGTAAAACGACGGC-CAGT) was used in combination with reverse primer VP4R 341–321 (Table 1) in order to generate an amplicon which contains the 59 end of VP4 gene. All RT-PCRs were generated using a One-

Step RT-PCR kit (Qiagen) on a GeneAMP PCR System 9700 thermocycler (Applied Biosystems). The dsRNA was denatured at 97 °C for 5 min, followed by the addition of RT-PCR master mix according to the kit instructions. RT was performed at 42 °C for 30 min, followed by 15 min at 95 °C, and the PCR was performed for 30 cycles consisting of 30 s at 95 °C, 30 s at 42 °C and 1 min at 72 °C. Reactions were subjected to a final extension step of 72 °C for 7 min and then held at 4 °C.

Sequencing

All amplicons were analysed by electrophoresis in 1 % SeaKem agarose gels (Thermo Fisher Scientific), subsequently excised and purified with a QIAquick Gel Extraction kit (Qiagen) following the manufacturer's instructions. Cycle sequencing of each amplicon was performed with the same consensus primers used for RT-PCR, using a Big Dye Terminator 3.1 Cycle Sequencing Ready Reaction kit (Applied Biosystems). Additionally, primers homologous to internal regions were employed in order to obtain complete sequences of each amplified cDNA (Table 1). The M13 primer (TGTAAAACGACGGCCAGT) was used during sequencing for the amplicon generated by VP4F 1–9 and VP4R 341–321 primers, in order to obtain sequence that is close to the 5′ end. Cycle sequencing products were purified using a carboxyl bead purification method (Mijatovic-Rustempasic *et al.*, 2012). Automated separation and base-calling of cycle sequencing products were performed using an ABI3130xl sequencer (Applied Biosystems). Sequence chromatogram files were edited and sequence contigs were assembled using Sequencher 4.8 software (Gene Codes Corporation).

Analysis

The genotypes of the 11 genome segments under investigation were determined according to the genotyping recommendations of the RCWG (Matthijnssens *et al.*, 2011a) using the RotaC online classification tool (http://rotac.regatools.be/) (Maes *et al.*, 2009). Nucleotide similarity searches were performed using BLAST (http://blast.ncbi.nlm.nih.gov/Blast.cgi). Sequences were aligned using the MUSCLE program within MEGA version 5. Once aligned, the JModelTest 2 program was used to identify the optimal evolutionary model that best fit the sequence datasets (Posada, 2008). Using corrected Akaike Information Criterion (AICc), the following models were found the best fit for the sequence data for each gene: GTR +I + G (VP2, VP3, VP4 and VP6), TIM1 +I + G (NSP1), TrN + G (NSP2), GTR+G (NSP3),HKY + G (NSP4),TPM2uf+I + G (NSP5),TrN + I (VP1), and TIM2 + I + G (VP7), TPM1uf+I + G (VP7 first and second codon positions only) and JTT + G + F (VP7 amino acids). Using these models, maximum-likelihood trees were constructed using PhyML 3.0 along with aLRT statistics for branch support (Guindon *et al.*, 2010). Nucleotide distance matrices were prepared using the *p*-distance algorithm of MEGA 5. Recombination analysis was carried out using Simplot 3.5.1, and GARD implemented in the Datamonkey webserver was used to identify breakpoints. Selection analysis was carried out in the Datamonkey webserver using a combination of the SLAC (single-likelihood ancestor counting), fixed effects likelihood (FEL) and random effects likelihood (REL) methods (Delport *et al.*, 2010; Kosakovsky Pond & Frost, 2005, 2006; Lole *et al.*, 1999).

Acknowledgments

We wish to thank Jon Gentsch (CDC, Atlanta, GA, USA) for his scientific advice and guidance. The findings and conclusions in this report are those of the author(s) and do not necessarily represent the official position of the Centers for Disease Control and Prevention. Names of specific vendors, manufacturers, or products are included for public health and informational purposes; inclusion does not imply endorsement of the vendors, manufacturers, or products by the Centers for Disease Control and Prevention or the US Department of Health and Human Services.

References

Bányai K, Bogdán A, Domonkos G, Kisfali P, Molnár P, Tóth A, Melegh B, Martella V, Gentsch JR, Szucs G. Genetic diversity and zoonotic potential of human rotavirus strains, 2003–2006, Hungary. J Med Virol. 2009; 81:362–370. [PubMed: 19107981]

Browning GF, Fitzgerald TA, Chalmers RM, Snodgrass DR. A novel group A rotavirus G serotype: serological and genomic characterization of equine isolate FI23. J Clin Microbiol. 1991; 29:2043–2046. [PubMed: 1663521]

Charrel RN, De Micco P, de Lamballerie X. Phylogenetic analysis of GB viruses A and C: evidence for cospeciation between virus isolates and their primate hosts. J Gen Virol. 1999; 80:2329–2335. [PubMed: 10501484]

Das BK, Gentsch JR, Cicirello HG, Woods PA, Gupta A, Ramachandran M, Kumar R, Bhan MK, Glass RI. Characterization of rotavirus strains from newborns in New Delhi, India. J Clin Microbiol. 1994; 32:1820–1822. [PubMed: 7929782]

Degiuseppe JI, Beltramino JC, Millan A, Stupka JA, Parra GI. Complete genome analyses of G4P[6] rotavirus detected in Argentinean children with diarrhoea provides evidence of interspecies transmission from swine. Clin microbiol Infect. 2013; 19:E367–E371. [PubMed: 23586655]

Delport W, Poon AF, Frost SD, Kosakovsky Pond SL. Datamonkey 2010: a suite of phylogenetic analysis tools for evolutionary biology. Bioinformatics. 2010; 26:2455–2457. [PubMed: 20671151]

Dóró R, Farkas SL, Martella V, Bányai K. Zoonotic transmission of rotavirus: surveillance and control. Expert Rev Anti Infect Ther. 2015; 13:1337–1350. [PubMed: 26428261]

Duchêne S, Holmes EC, Ho SY. Analyses of evolutionary dynamics in viruses are hindered by a time-dependent bias in rate estimates. Proc Biol Sci. 2014; 281:20140732. [PubMed: 24850916]

Esona MD, Mijatovic-Rustempasic S, Conrardy C, Tong S, Kuzmin IV, Agwanda B, Breiman RF, Banyai K, Niezgoda M, et al. Reassortant group A rotavirus from straw-colored fruit bat (*Eidolon helvum*). Emerg Infect Dis. 2010; 16:1844–1852. [PubMed: 21122212]

Estes, MK.; Kapikian, AZ. Rotaviruses. In: Fields, BN.; Knipe, DM.; Howley, PM., editors. Fields Virology. 5th. Philadelphia, PA: Wolters Kluwer Health/Lippincott Williams & Wilkins; 2007. p. 1917-1974.

Gentsch JR, Glass RI, Woods P, Gouvea V, Gorziglia M, Flores J, Das BK, Bhan MK. Identification of group A rotavirus gene 4 types by polymerase chain reaction. J Clin Microbiol. 1992; 30:1365–1373. [PubMed: 1320625]

Gentsch JR, Laird AR, Bielfelt B, Griffin DD, Banyai K, Ramachandran M, Jain V, Cunliffe NA, Nakagomi O, et al. Serotype diversity and reassortment between human and animal rotavirus strains: implications for rotavirus vaccine programs. J Infect Dis. 2005; 192(Suppl 1):S146–S159. [PubMed: 16088798]

Ghosh S, Kobayashi N. Whole-genomic analysis of rotavirus strains: current status and future prospects. Future Microbiol. 2011; 6:1049–1065. [PubMed: 21958144]

Guindon S, Dufayard JF, Lefort V, Anisimova M, Hordijk W, Gascuel O. New algorithms and methods to estimate maximum-likelihood phylogenies: assessing the performance of PhyML 3.0. Syst Biol. 2010; 59:307–321. [PubMed: 20525638]

Kerin TK, Kane EM, Glass RI, Gentsch JR. Characterization of VP6 genes from rotavirus strains collected in the United States from 1996-2002. Virus Genes. 2007; 35:489–495. [PubMed: 17564821]

Kosakovsky Pond SL, Frost SD. Not so different after all: a comparison of methods for detecting amino acid sites under selection. Mol Biol Evol. 2005; 22:1208–1222. [PubMed: 15703242]

Kosakovsky Pond SL, Posada D, Gravenor MB, Woelk CH, Frost SD. Automated phylogenetic detection of recombination using a genetic algorithm. Mol Biol Evol. 2006; 23:1891–1901. [PubMed: 16818476]

Lole KS, Bollinger RC, Paranjape RS, Gadkari D, Kulkarni SS, Novak NG, Ingersoll R, Sheppard HW, Ray SC. Full-length human immunodeficiency virus type 1 genomes from subtype C-infected seroconverters in India, with evidence of intersubtype recombination. J Virol. 1999; 73:152–160. [PubMed: 9847317]

Maes P, Matthijnssens J, Rahman M, Van Ranst M. RotaC: a web-based tool for the complete genome classification of group A rotaviruses. BMC Microbiol. 2009; 9:238. [PubMed: 19930627]

Matthijnssens J, Ciarlet M, Heiman E, Arijs I, Delbeke T, McDonald SM, Palombo EA, Iturriza-Gómara M, Maes P, et al. Full genome-based classification of rotaviruses reveals a common origin between human Wa-Like and porcine rotavirus strains and human DS-1-like and bovine rotavirus strains. J Virol. 2008; 82:3204–3219. [PubMed: 18216098]

Matthijnssens J, Bilcke J, Ciarlet M, Martella V, Bányai K, Rahman M, Zeller M, Beutels P, Van Damme P, Van Ranst M. Rotavirus disease and vaccination: impact on genotype diversity. Future Microbiol. 2009; 4:1303–1316. [PubMed: 19995190]

Matthijnssens J, Taraporewala ZF, Yang H, Rao S, Yuan L, Cao D, Hoshino Y, Mertens PP, Carner GR, et al. Simian rotaviruses possess divergent gene constellations that originated from interspecies transmission and reassortment. J Virol. 2010; 84:2013–2026. [PubMed: 19939934]

Matthijnssens J, Ciarlet M, McDonald SM, Attoui H, Bányai K, Brister JR, Buesa J, Esona MD, Estes MK, et al. Uniformity of rotavirus strain nomenclature proposed by the Rotavirus Classification Working Group (RCWG). Arch Virol. 2011a; 156:1397–1413. [PubMed: 21597953]

Matthijnssens J, De Grazia S, Piessens J, Heylen E, Zeller M, Giammanco GM, Bányai K, Buonavoglia C, Ciarlet M, et al. Multiple reassortment and interspecies transmission events contribute to the diversity of feline, canine and feline/canine-like human group A rotavirus strains. Infect Genet Evol. 2011b; 11:1396–1406. [PubMed: 21609783]

Matthijnssens J, Miño S, Papp H, Potgieter C, Novo L, Heylen E, Zeller M, Garaicoechea L, Badaracco A, et al. Complete molecular genome analyses of equine rotavirus A strains from different continents reveal several novel genotypes and a largely conserved genotype constellation. J Gen Virol. 2012; 93:866–875. [PubMed: 22190012]

McNeal MM, Sestak K, Choi AH, Basu M, Cole MJ, Aye PP, Bohm RP, Ward RL. Development of a rotavirus-shedding model in rhesus macaques, using a homologous wild-type rotavirus of a new P genotype. J Virol. 2005; 79:944–954. [PubMed: 15613323]

Mijatovic-Rustempasic S, Bányai K, Esona MD, Foytich K, Bowen MD, Gentsch JR. Genome sequence based molecular epidemiology of unusual US Rotavirus A G9 strains isolated from Omaha, USA between 1997 and 2000. Infect Genet Evol. 2011; 11:522–527. [PubMed: 21130184]

Mijatovic-Rustempasic S, Frace MA, Bowen MD. Cost-effective paramagnetic bead technique for purification of cycle sequencing products. Sequencing. 2012; 2012:767959.

Mladenova Z, Papp H, Lengyel G, Kisfali P, Steyer A, Steyer AF, Esona MD, Iturriza-Gómara M, Bányai K. Detection of rare reassortant G5P[6] rotavirus, Bulgaria. Infect Genet Evol. 2012; 12:1676–1684. [PubMed: 22850117]

Patel MM, Steele D, Gentsch JR, Wecker J, Glass RI, Parashar UD. Real-world impact of rotavirus vaccination. Pediatr Infect Dis J. 2011; 30:S1–S5. [PubMed: 21183833]

Philippe H, Brinkmann H, Lavrov DV, Littlewood DT, Manuel M, Wörheide G, Baurain D. Resolving difficult phylogenetic questions: why more sequences are not enough. PLoS Biol. 2011; 9:e1000602. [PubMed: 21423652]

Posada D. jModelTest: phylogenetic model averaging. Mol Biol Evol. 2008; 25:1253–1256. [PubMed: 18397919]

Sanjuán R, Nebot MR, Chirico N, Mansky LM, Belshaw R. Viral mutation rates. J Virol. 2010; 84:9733–9748. [PubMed: 20660197]

Santos N, Hoshino Y. Global distribution of rotavirus serotypes/genotypes and its implication for the development and implementation of an effective rotavirus vaccine. Rev Med Virol. 2005; 15:29–56. [PubMed: 15484186]

Tate JE, Cortese MM, Payne DC, Curns AT, Yen C, Esposito DH, Cortes JE, Lopman BA, Patel MM, et al. Uptake, impact, and effectiveness of rotavirus vaccination in the United States: review of the first 3 years of postlicensure data. Pediatr Infect Dis J. 2011; 30:S56–S60. [PubMed: 21183842]

Weinberg GA, Teel EN, Mijatovic-Rustempasic S, Payne DC, Roy S, Foytich K, Parashar UD, Gentsch JR, Bowen MD. Detection of novel rotavirus strain by vaccine postlicensure surveillance. Emerg Infect Dis. 2013; 19:1321–1323. [PubMed: 23876297]

Xia L, Fan Q, He B, Xu L, Zhang F, Hu T, Wang Y, Li N, Qiu W, et al. The complete genome sequence of a G3P[10] Chinese bat rotavirus suggests multiple bat rotavirus inter-host species transmission events. Infect Genet Evol. 2014; 28:1–4. [PubMed: 25218875]

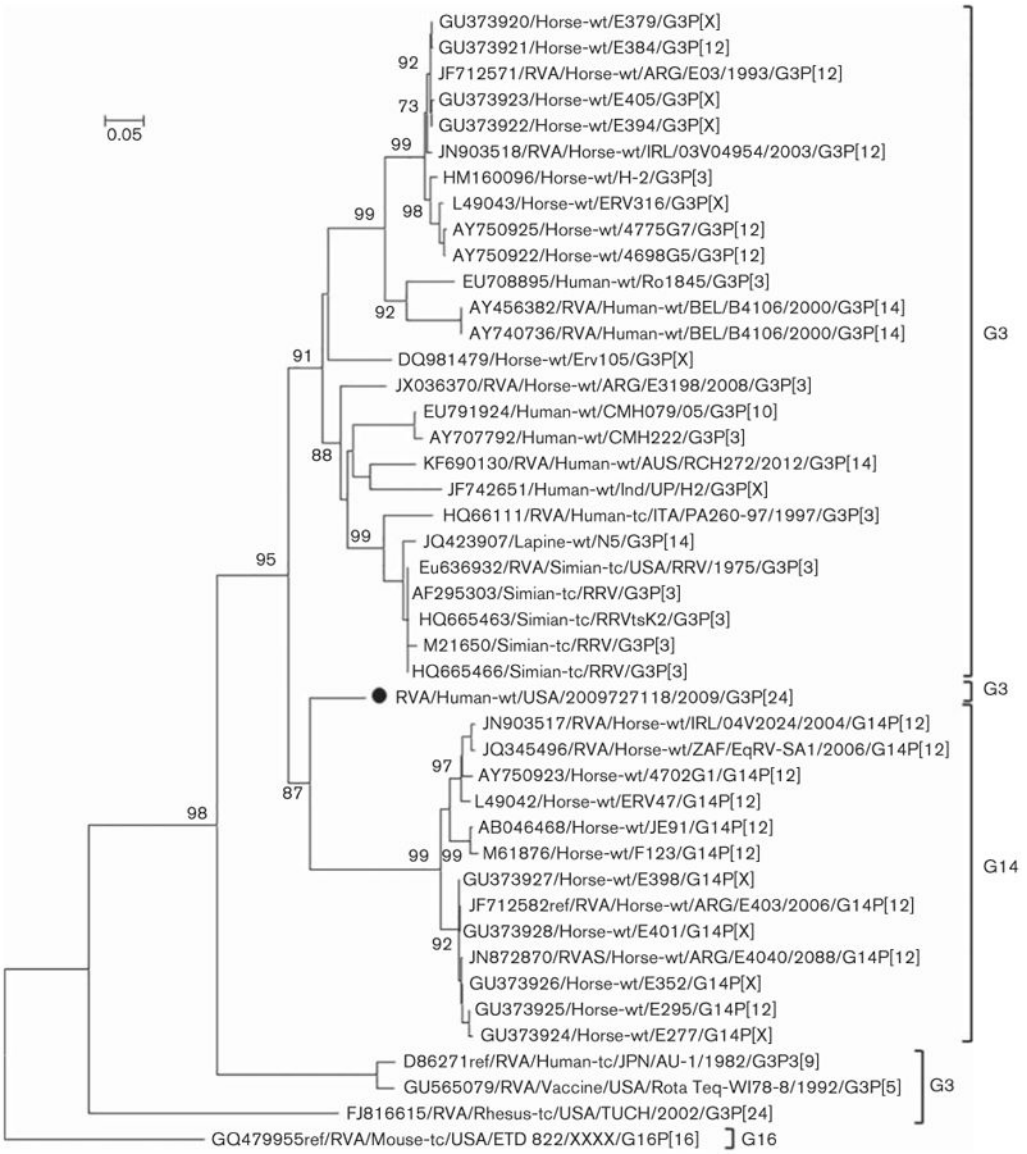


Fig. 1.

Phylogenetic trees based on the full-length nucleotide sequences of the rotavirus VP7 gene.

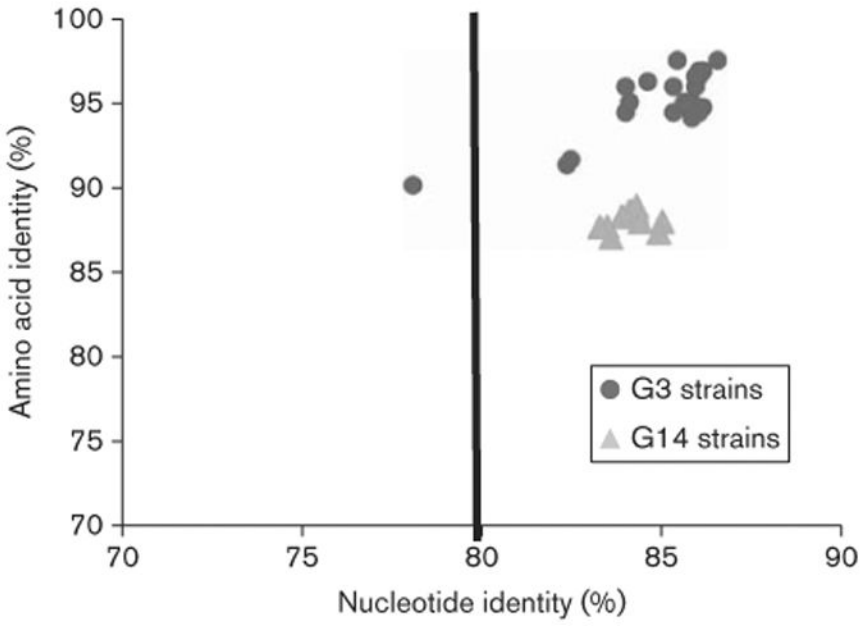


Fig. 2.

Percentage nucleotide and deduced amino acid homologies of the VP7 gene segment of rotavirus strain 2009727118 compared with respective genes deposited in GenBank. The vertical line indicates the nucleotide percentage identity cut-off value defining genotypes for 11 rotavirus gene segments (Matthijnssens *et al.*, 2011a). Dark grey dots and light grey triangles indicate coordinates for each pairwise comparison when percentage nucleotide identity is plotted against percentage amino acid identity for G3 and G14 strains, respectively.

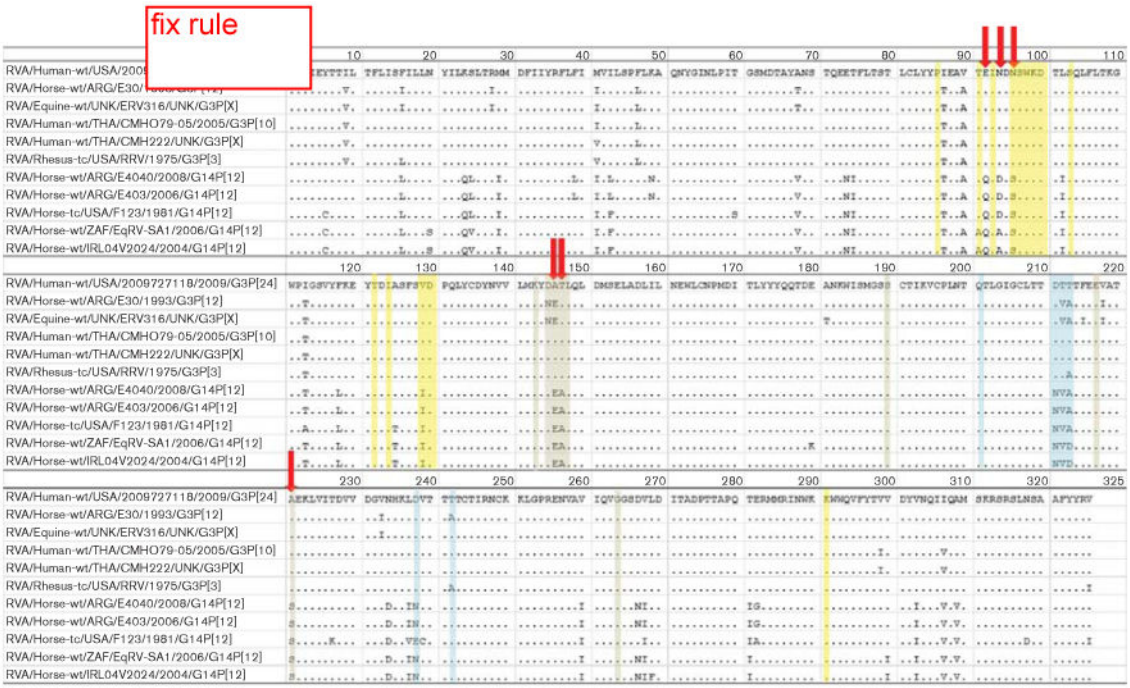
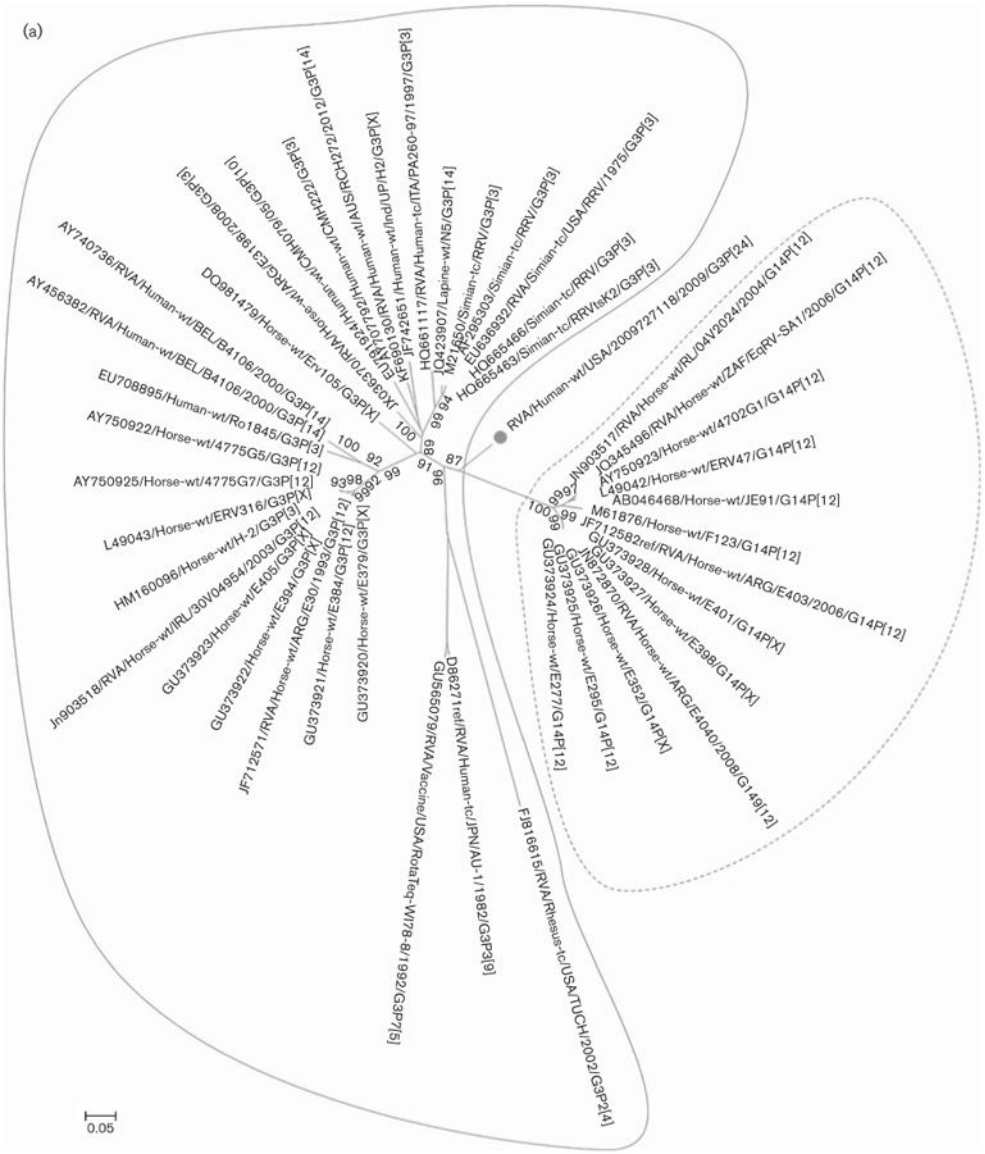
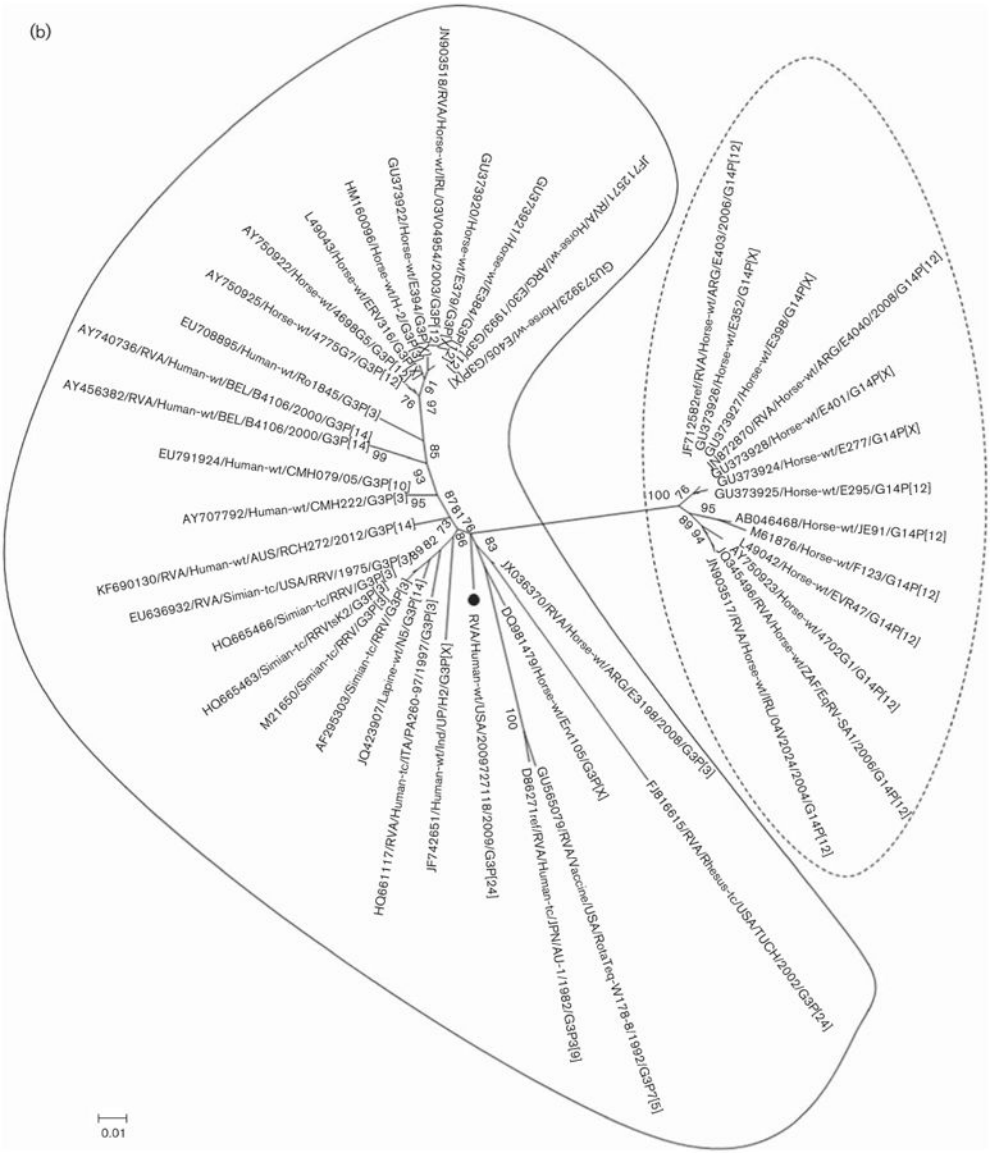


Fig. 3.

Amino acid alignment of the VP7 gene for 2009727118 and four G14 and two G3 strains from horses and three other G3 strains with close identities to this strain. Numbering indicates amino acid position. Red arrows indicate important amino acid positions in the alignment. Antigenic regions 7-1a (yellow), 7-1b (blue) and 7-2 (grey) are highlighted.





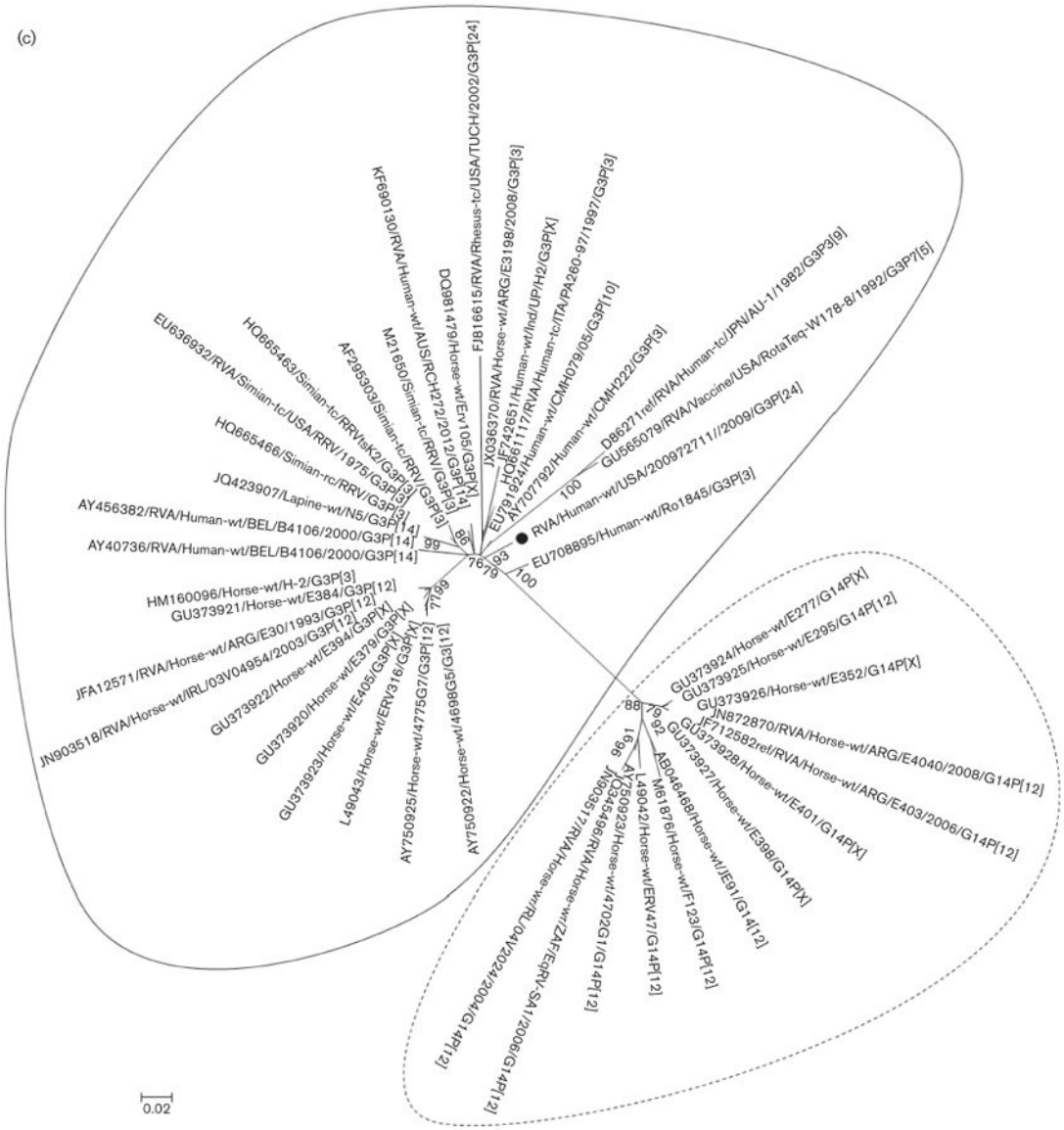


Fig. 4.

Unrooted phylogenetic trees based on the full-length nucleotide sequences of the rotavirus VP7 gene. (a) Phylogenetic alignment including all three bases, (b) phylogenetic tree including first and second bases, and (c) amino acid alignment. G3 and G14 strains are circled with solid and dashed lines, respectively.

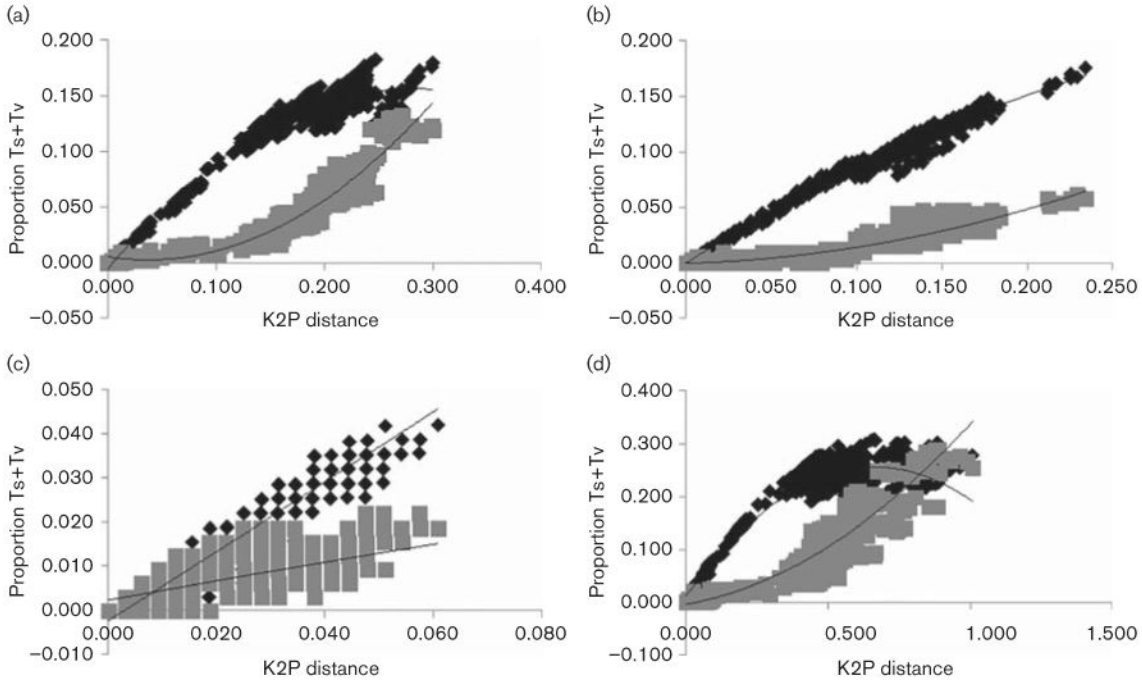


Fig. 5.

Nucleotide substitution saturation plots for the VP7 gene. Sequence divergence using Kimura two-parameter (K2P) distance on the x-axis plotted against the proportion of transitions (Ts, black diamonds) and transversions (Tv, grey squares) on the y-axis for all strains included in the VP7 phylogenetic tree in Fig. 1. (a) All three codon positions, (b) first codon position, (c) second codon position and (d) third codon position.

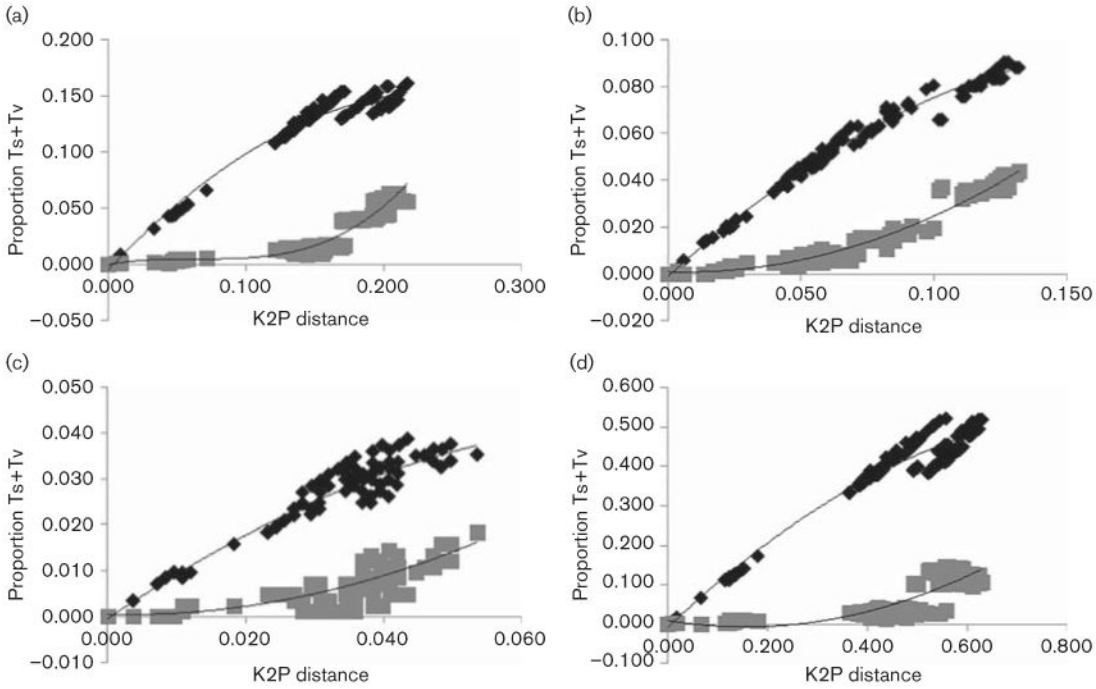
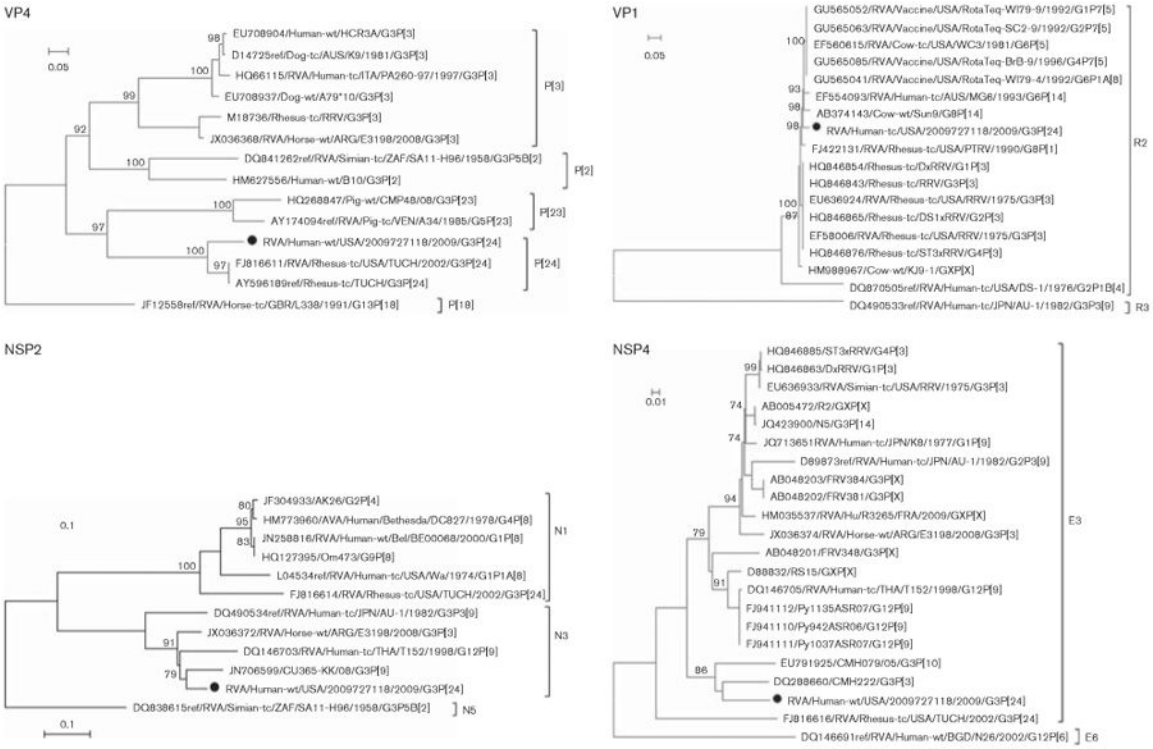


Fig. 6.

Nucleotide substitution saturation plots for VP3 gene. Sequence divergence using Kimura two-parameter (K2P) distance on the x-axis plotted against the proportion of transitions (Ts, black diamonds) and transversions (Tv, grey squares) on the y-axis for all strains included in the VP3 phylogenetic tree in Fig. 2. (a) All three codon positions, (b) first codon position, (c) second codon position and (d) third codon position.



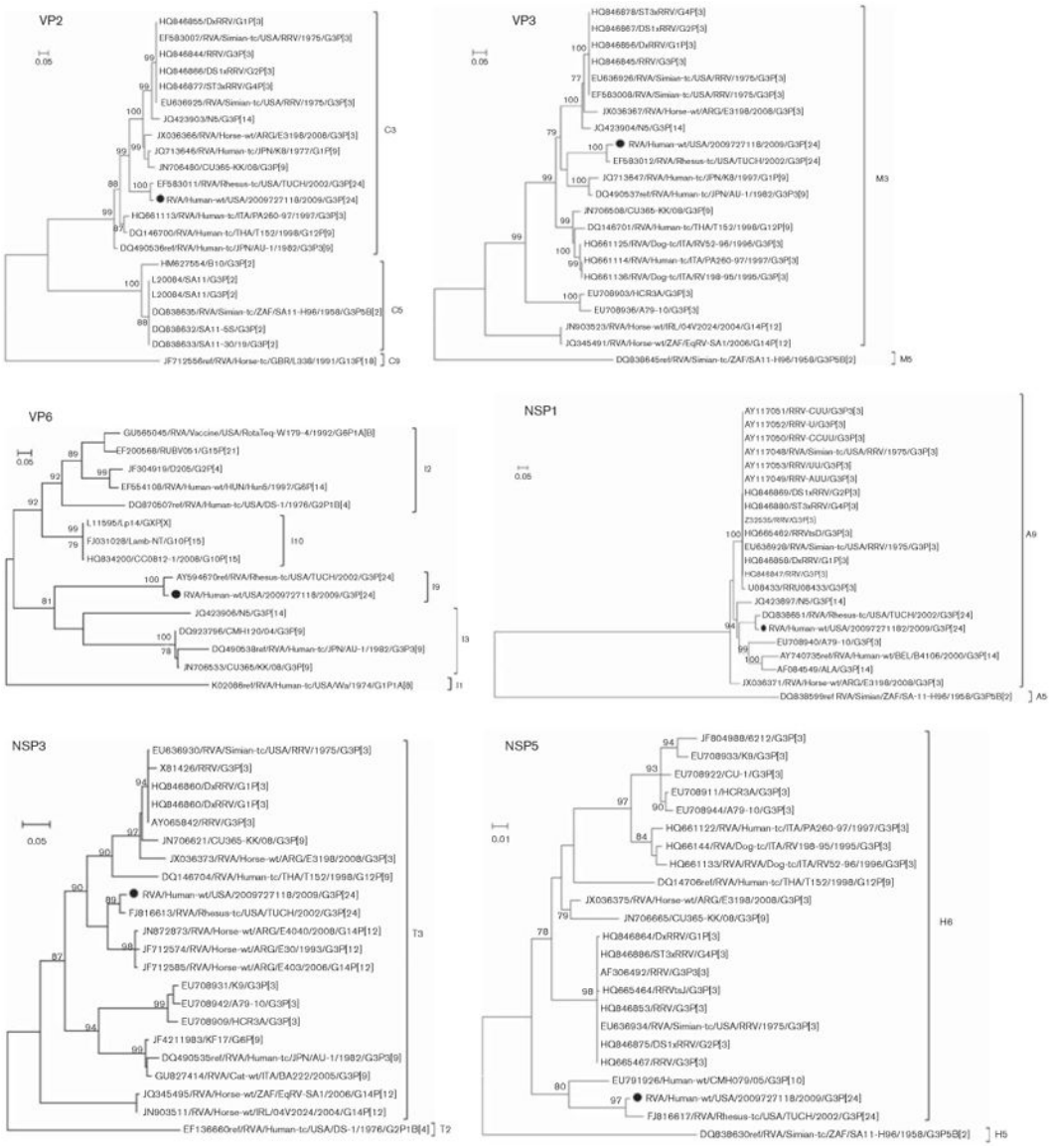


Fig. 7.

Phylogenetic trees based on the full-length nucleotide sequences of the VP and NSP genes.

Table 1
Primers used for sequencing of rotavirus genes designed in this study

Primer	Sequence (5′→3′)	Gene	Nucleotide position, strand	Reference
V1Fint1	TGATTGAYGCAGCATTAGAC	VP1	1034–1054, +	Mijatovic-Rustempasic *et al.* (2011)
V1Rint2	GACTGTAGATACCAATGCTT	VP1	2018–2037, −	Banyai *et al.* (2009)
VP1F 1–23	GGCTATTAAAGCTATACAATGGG	VP1	1–23, +	This study
VP1R 225–205	AGCATTCTCTATAACATCGCC	VP1	225–205, −	This study
VP1F 336–316	CTCATTCACTGTCAAATCTGC	VP1	336–316, +	This study
VP1R 3302–3288	GGTCACATCTAAGCG	VP1	3302–3288, −	This study
VP1F 3039–3061	CGACTTCAATTCACCAGATCTAG	VP1	3039–3061, +	This study
VP1F 2963–2980	CTCAAGATAAATACAGGA	VP1	2963–2980, −	This study
V2Rint1	GGTCTAATATATCGRGCTGTT	VP2	922–942, −	Mijatovic-Rustempasic *et al.* (2011)
V2Rint2	CTGCATTAATGCTTCCATTAA	VP2	1598–1618, −	Banyai *et al.* (2009)
V2Fint1	GCAGATATGAGACAACAAGTT	VP2	716–736, +	Mijatovic-Rustempasic *et al.* (2011)
V2Fint2	GGCTTGTATAACTATGAAYAT	VP2	1720–1740, +	Mijatovic-Rustempasic *et al.* (2011)
VP2F 1–22	GGCTATTAAAGGCTCAATGGCG	VP2	1–22, +	This study
VP2R 197–176	GATTATCTGTGATTATCTCCTC	VP2	197–176, −	This study
VP2R 266–249	CTTCTAATAGTTGTTTGG	VP2	266–249, −	This study
VP2R 361–341	CTTCAGTATAGACTCTTTTGG	VP2	361–341, −	This study
VP2R 471–454	CGCAATTCCTTCTCCCCA	VP2	471–454, −	This study
VP2R 2714–2694	GGTCATATCTCCACARTGGGG	VP2	2714–2694, −	This study
VP2F 2460–2480	CGAATGACTTCTACCTTGTAGC	VP2	2460–2480, +	This study
VP2F 2249–2268	GAAGAACTGATGAGGACAGG	VP2	2249–2268, +	This study
VP2F 2173–2191	GCAGTTAGAAAGAGACGAG	VP2	2173–2191, +	This study
K-VP3F	GGCTATTAAAGCAGTACYAGT	VP3	1–21, +	Mijatovic-Rustempasic *et al.* (2011)
K-VP3R	GGTCACATCYTGACTAGTGT	VP3	2572–2591, −	Mijatovic-Rustempasic *et al.* (2011)
V3Rint1	ATGGATCCCATGTCTCAAATG	VP3	957–977, −	Mijatovic-Rustempasic *et al.* (2011)
V3Fint1	TACATTTGARACATGGGATCC	VP3	955–975, +	Mijatovic-Rustempasic *et al.* (2011)
V3Rint2	GAATATGARGTCGCAGACTCT	VP3	1786–1806, −	Mijatovic-Rustempasic *et al.* (2011)
V3Fint2	TTACATCATATGATGGTTGT	VP3	1614–1633, +	Mijatovic-Rustempasic *et al.* (2011)
V3Fint3	TCAGTTTCAGGACATGTYTA	VP3	1871–1890, +	Mijatovic-Rustempasic *et al.* (2011)
VP3F 487–506	GAACGCCGCAACTGATGACG	VP3	487–506, +	This study
VP3R 1102–1080	CTCCAATCCACATTCCCACGGTC	VP3	1102–1080, −	This study
VP4F 1–9	TGTAAAACGACGGCCAGTGGCT ATAAA	VP4	1–10, +	This study
VP4F 717–737	TCATCAGCCAGCACGCTCAAG	VP4	717–737, +	This study
VP4F 1820–1838	CCATCGCTACGCAAACTGC	VP4	1820–1838, +	This study

Primer	Sequence (5′→3′)	Gene	Nucleotide position, strand	Reference
VP4R 2362–2341	GGTCACATCCTCTGGAAATTGC	VP4	2362–2341, −	This study
VP4R 1424–1403	CATCGTTTGATGGCACAAGGGA	VP4	1424–1403, −	This study
VP4R 341–321	GAAGCGACATTTGGCTCAATC	VP4	341–321, −	This study
JRG7	GGCTTTAAAACGAAGTCTTC	VP6	1–20, +	Kerin *et al.* (2007)
JRG8	GGTCACATCCTCTCACTACAT	VP6	1336–1356, −	Kerin *et al.* (2007)
VP6F 1111–1128	CACTAACTACTCACCTTC	VP6	1111–1128, +	This study
VP6R 549–528	GCGTTTAACCACATAGTACCCA	VP6	549–528, −	This study
K-NSP1F	GGCTTTTTTTWTGAAAAGTCT	NSP1	1–21, +	Mijatovic-Rustempasic *et al.* (2011)
K-NSP1R	GGTCACACTTTTTTATGCTGC	NSP1	1548–1568, −	Mijatovic-Rustempasic *et al.* (2011)
N1Rint1	TCTYAAATCATTCCATACAG	NSP1	1023–1042, −	Mijatovic-Rustempasic *et al.* (2011)
N1Fint1	GTYTCHACTAGATGTAG	NSP1	878–894, +	Mijatovic-Rustempasic *et al.* (2011)
K-NSP2F	GGCTTTTAAAGCGTCTCAGTC	NSP2	1–21, +	Mijatovic-Rustempasic *et al.* (2011)
K-NSP2R	GGTCACATAAGCGCTTTCTAT	NSP2	1039–1059, −	Mijatovic-Rustempasic *et al.* (2011)
N2Rint1	AACTCTATAGTGACCTTTAC	NSP2	711–730, −	Mijatovic-Rustempasic *et al.* (2011)
N2Fint1	ATAGCTATTGGTCAVTCAA	NSP2	458–476, +	Mijatovic-Rustempasic *et al.* (2011)
K-NSP3F	GGCTTTTAATGCTTTTCAGTGG	NSP3	1–22, +	Mijatovic-Rustempasic *et al.* (2011)
K-NSP3R	GGTCACATAACGCCCCTATAG	NSP3	1064–1084, −	Mijatovic-Rustempasic *et al.* (2011)
N3Fint1	GAATCAYTRAARCAAAGAGT	NSP3	569–588, +	Mijatovic-Rustempasic *et al.* (2011)
K-NSP4F	GGCTTTTAAAAGTTCTGTTCC	NSP4	1–21, +	Mijatovic-Rustempasic *et al.* (2011)
K-NSP4R	GGTCACATTAAGACCYTTCCT	NSP4	731–751, −	Mijatovic-Rustempasic *et al.* (2011)
K-NSP5F	GGCTTTWAAAGCGCTACAGTG	NSP5	1–21, +	Mijatovic-Rustempasic *et al.* (2011)
K-NSP5R	GGTCACAAAACGGGAGTGG	NSP5	646–664, −	Mijatovic-Rustempasic *et al.* (2011)