Published in final edited form as:
Virology. 2012 September 15; 431(1-2): 50–57. doi:10.1016/j.virol.2012.05.011.

Rotavirus RNA Polymerases Resolve into Two Phylogenetically Distinct Classes that Differ in Their Mechanism of Template Recognition

Kristen M. Ogden[1], **Reimar Johne**[2], and **John T. Patton**[1,*]

[1]Laboratory of Infectious Diseases, National Institute of Allergy and Infectious Diseases, National Institutes of Health, Bethesda, Maryland, USA

[2]Federal Institute for Risk Assessment, Max-Dohrn-Strasse 8-10, $_{10589}$ Berlin, Germany

Abstract

Rotaviruses (RVs) are segmented double-stranded RNA viruses that cause gastroenteritis in mammals and birds. Within the RV genus, eight species (RVA-RVH) have been proposed. Here, we report the first RVF and RVG sequences for the viral RNA polymerase (VP1)-encoding segments and compare them to those of other RV species. Phylogenetic analyses indicate that the VP1 RNA segments and proteins resolve into two major clades, with RVA, RVC, RVD and RVF in clade A, and RVB, RVG and RVH in clade B. Plus-strand RNA of clade A viruses, and not clade B viruses, contain a 3′-proximal UGUG cassette that serves as the VP1 recognition signal. VP1 structures for a representative of each RV species were predicted using homology modeling. Structural elements involved in interactions with the UGUG cassette were conserved among VP1 of clade A, suggesting a conserved mechanism of viral RNA recognition for these viruses.

Keywords

Rotavirus; RNA polymerase; VP1; polymerase structure; double-stranded RNA virus

Introduction

Rotaviruses (RVs) are members of the *Reoviridae* family and contain genomes consisting of eleven segments of double-stranded (ds)RNA. The International Committee of Taxonomy of Viruses (ICTV) recognizes five species within the RV genus: *Rotavirus A, B, C, D*, and *E* (RVA-RVE). Three additional species have been proposed: *Rotavirus F, G,* and *H* (RVF-RVH) (Attoui et al., 2011; Matthijnssens et al., 2012). Viruses of the RVA species have been the most widely studied, due to their significance as the primary cause of life-threatening diarrhea in infants and young children (Greenberg and Estes, 2009). In addition, RVA is an important cause of diarrheal disease in other mammalian and avian species. RVB and RVH are most noted for their association with outbreaks of diarrheal disease in older children and adults in Asia. RVC has been associated with limited outbreaks of gastroenteritis in institutional settings and in children and families (Attoui et al., 2011). RVE

*Corresponding author. John T. Patton, Laboratory of Infectious Diseases, Building 50, Room 6313, 50 South Dr., MSC8007, NIH, Bethesda, MD 20892. Phone: (301) 594-1615; Fax (301) 480-7941; jpatton@niaid.nih.gov.

was isolated from diseased pigs, while RVD, RVF, and RVG have been recovered only from birds (Trojnar et al., 2010).

The RV virion is a non-enveloped, triple-layered, icosahedral particle (Estes and Kapikian, 2007; Settembre et al., 2011). The predominant component of the outer layer is the glycoprotein, VP7. Projecting from the VP7 layer are spikes formed by the protease-sensitive attachment protein, VP4. The intermediate layer of the virion consists of a single protein, VP6, while the inner layer is assembled from the core shell protein, VP2. Attached to the interior side of the VP2 layer are the minor core components, VP1 and VP3. VP1 represents the viral RNA-dependent RNA polymerase (RdRP), and VP3 represents the viral RNA capping enzyme.

Historically, the species assignment for a RV isolate has been made based on the reactivity of its VP6 component to reference sera or monoclonal antibodies. More recently, sequencing and phylogenetic analysis of the VP6 gene has been used to define RV species (Matthijnssens et al., 2012). Nucleotide sequences have been reported for all eleven genome segments of at least one isolate of RVA, RVB, RVC, RVD and RVH; no sequences are available for any RVE isolate. RVF and RVG are poorly sequenced species, with only the VP6 segment of a single isolate of each of these known (Johne et al., 2011). Based on phylogenetic analysis of VP6 sequences, RVs can be divided into two clades (Attoui et al., 2012; Johne et al., 2011). One includes the RVA, RVC, RVD, and RVF species (herein referred to as clade A), while the other includes the RVB, RVG, and RVH species (clade B).

RV dsRNA genome segments are largely monocistronic, containing a single open reading frame (ORF) that is flanked by a short 5′ untranslated region (UTR) and a short-to-long 3′ UTR. During the RV life cycle, the genome segments are transcribed, producing plus-strand (+) RNA that contain 5′ caps but that lack 3′ poly(A) tails. The 5′-terminal sequence of the viral (+)RNA of all RV species tends begin with the short consensus sequence: 5′-GGC(A/U). For RVA viruses, the consensus 3′-terminal sequence of the RNA is 5′-UGUGACC-3′. The UGUG cassette of this element is important for (+)RNA recognition by the viral RdRP, while the terminal CC residues have a critical role in the initiation of RNA synthesis (Lu et al., 2008; Tortorici et al., 2003). The UGUG cassette and its position, with respect to the 3′ terminus, appear to be conserved not only among RVA viruses, but for all clade A RV (Chen et al., 2002; Johne et al., 2011; Trojnar et al., 2009; Trojnar et al., 2010). The UGUG cassette has not been found at the 3′ termini of any clade B (+)RNA (Ghosh et al., 2010; Jiang et al., 2008; Johne et al., 2011).

In this study, we describe the first known RdRP sequences for RVF and RVG viruses and contrast these to the RdRP sequences of other RV species. Consistent with previous studies of the VP6 capsid protein, our analysis shows that the RdRP of the eight RV species resolve phylogenetically into two genetically distant clades. The RVF RdRP clusters with RVA, RVC, and RVD sequences in clade A, while the RVG RdRP clusters with RVB and RVH sequences in clade B. Like other clade A viruses, the RVF VP1 (+)RNA contains a 3′-proximal UGUG cassette. In contrast, the RVG VP1 (+)RNA lacks this cassette, like other clade B viruses. Protein homology modeling indicates that all clade A viruses encode structurally similar RdRP that rely on recognition of the UGUG cassette to form complexes with viral (+)RNA templates. Based on the absence of UGUG cassettes in the clade B (+)RNA and predicted structural differences in clade A and B RdRPs, the clade B RdRP can be presumed to use an alternative mechanism for recognizing viral (+)RNA.

Results

Sizes of RVF and RVG VP1 RNA and protein

Virus strains 03V0567 (RVG) and 03V0568 (RVF) were derived from the intestinal contents of stunted chicks, as previously described (Johne et al., 2011). The dsRNA genome segments were purified from the viruses and resolved by gel electrophoresis. The largest segment of each virus was gel purified, and sequence-independent amplification methods as well as degenerated primer approaches were used to generate cDNAs of the RNA. The cDNAs were ligated into TOPO TA cloning vectors (Invitrogen) and sequenced. The analysis showed that the largest genome segment of RVF 03V0568 was 3296 nucleotides in length and encoded a VP1 protein of 1086 amino acids (125 kD). These sizes are typical of the VP1 segments of clade A RVs (i.e, RVA, RVC, RVD) (Table 1). Similarly, the sizes of the 5′-UTR (18 nt) and 3′-UTR (17 nt) of the RVF isolate approximate those of several species within clade A, particularly the RVA and RVD viruses. For RVG 03V0567, the largest genome segment was 3526 nucleotides in length and encoded a VP1 protein of 1160 amino acids (133 kD). These sizes are similar to those of other clade B viruses (i.e., RVB, RVH) and are larger than those of clade A viruses (Table 1). In particular, the RVG RdRP contains an additional 70–81 amino acids in comparison to clade A RdRP.

Terminal sequences of VP1 (+)RNA

Sequence comparisons showed that the 5′ termini of the VP1 (+)RNA of the RVF 03V0568 and RVG 03V0567 viruses began with the sequence GGC, followed by several A and U residues (Table 2). The 5′ termini of the RVF and RVG VP1 (+)RNA are like those of the VP6 (+)RNA of these viruses, except that the RVG VP6 (+)RNA initiates with GGA instead of GGC (Johne et al., 2011). While, in most cases, the third residue of RV (+)RNA has been reported to be a C, there are several examples of RVA, RVD, and RVH (+)RNA that, like the RVG VP6 (+)RNA, do not conserve the C residue (Ghosh et al., 2010; Jiang et al., 2008; Trojnar et al., 2010).

Similar to RVA viruses in general, as well as RVD 05V0049, the 3′-terminal sequence of the VP1 (+)RNA of RVF 03V0568 is UGUGACC (Table 2) (Estes and Kapikian, 2007; Trojnar et al., 2009; Trojnar et al., 2010). The VP1 (+)RNA of the RVC Bristol strain terminates with GCU, instead of the ACC sequence that is more typical of clade A RV, but the Bristol VP1 (+)RNA does contain the 3′-proximal UGUG cassette. Despite a high level of conservation of the UGUG cassette among clade A viruses, the 3′-terminal sequence of the VP6-encoding segment of RVF 03V0568, 5′-UAUGACC-3′, indicates that the UGUG cassette is not absolutely conserved (Johne et al., 2011). The 3′-terminal sequence of the VP1 (+)RNA of RVG 03V0567, 5′-AAAGACCC-3′, is similar to those of RVB and RVH. The majority of RVB and RVH genome segments terminate with 5′-ACCC-3′ (Ahmed et al., 2004; Ghosh et al., 2010; Jiang et al., 2008; Nagashima et al., 2008). Upstream residues, however, vary in a species-specific manner. The 3′-terminal sequence of the VP6 (+)RNA of RVG 03V0567 is identical to that of its VP1 (+)RNA, which suggests that this sequence is characteristic for the isolate (Johne et al., 2011).

Phylogenetic analysis

Maximum likelihood phylogenetic trees were generated to compare the relatedness of full-length RVF 03V0568 and RVG 03V0567 VP1 (+)RNA sequences and the deduced VP1 amino acid sequences with those of representative strains from other RV groups (Fig. 1). Included in the trees were both mammalian (SA11, RRV, UK, ETD, Bristol, Cowden, Bang373, DB176, B219, J219) and avian (PO-13, 02V0002G3, 05V0049) RV isolates. Each VP1 tree resolved into two major clades. Just as in VP6 trees, RVA, RVC, RVD, and RVF clustered in clade A and RVB, RVG and RVH clustered in clade B in the VP1 trees (Johne

et al., 2011). Within a clade, representative strains each occupied a distinct branch of the tree, consistent with individual RV species.

Nucleotide and deduced amino acid sequence comparison

To assess identity and similarity, the nucleotide sequences of the VP1 (+)RNA and the deduced VP1 amino acid sequences for RVF 03V0568 and RVG 03V0567 were directly compared to those of representatives from the other RV species. For clade A viruses, VP1 was predicted to be 1079–1088 amino acids in length (Table 1). For clade B viruses, VP1 was predicted to contain 1158–1167 amino acids. In keeping with phylogenetic analyses, RVF 03V0568 had higher nucleotide (55–61%) and amino acid (46–54%) identities with other clade A viruses than with the clade B viruses (Table 3). Similarly, RVG 03V0567 had higher nucleotide (59–64%) and amino acid (58–60%) identity with the clade B viruses than with the clade A viruses. Conversely, VP1 amino acid sequences differed dramatically between RVF 03V0568 and the clade B viruses (20–22% identity) and between RVG 03V0567 and the clade A viruses (19–20% identity), reinforcing the intra-clade relationships of these species.

Predicted VP1 structures

A high level of structure and sequence conservation among RdRP polymerase domains makes these molecules good templates for homology modeling. To predict the structures of viral polymerases for isolates from different RV species, deduced VP1 amino acid sequences were input into the I-TASSER server, and the highest scoring homology model for each isolate was subsequently analyzed. Homology models for the clade A viruses RVC Bristol, RVD 05V0049, and RVF 03V0568 scored extremely well, in terms of confidence and template modeling (Table 4). They also exhibited low overall root mean square deviations (RMSD; ~3.5–4.5 Å) from the RVA SA11 VP1 structures on which they were primarily modeled. These observations suggest that homology models for clade A VP1 are very good approximations of the biological molecular structures. When homology models of clade A VP1 proteins were overlaid on the structure of RVA SA11 VP1, only minor predicted differences in overall architecture were observed (Fig. 2)

In comparison to predicted clade A VP1 structures, homology models of VP1 for clade B viruses RVB Bang373, RVB DB176, RVG 03V0567, RVH B219, and RVH J19 received lower confidence and template modeling scores and exhibited higher RMSD (~7.5–9.5 Å) (Table 4). These findings are not surprising, given the greater phylogenetic distances, reduced amino acid identities, and differences in predicted length between RVA SA11 VP1 and clade B VP1. While lower scores suggest that homology models for clade B viruses might not reflect the biological RdRP structures as accurately as the clade A models, clade B models may provide valuable insight into the locations of sites of variation, such as insertions or deletions, in the RdRP structure.

Clade B VP1 proteins are approximately 70 amino acids longer overall than clade A VP1 (Table 1). Based on the RVG 03V0567 homology model, which was the best scoring clade B VP1 model (Table 4), the majority of the difference in size was accounted for by five insertions of 7–16 amino acids, and a handful of smaller insertions, within the N-terminal domain (Fig. 2). Most insertions were located in loops, but two additional α-helical turns were predicted at the N terminus for RVG 03V0567 VP1. Several small insertions (1–4 amino acids) in loops of the polymerase and C-terminal domains accounted for the remaining difference in VP1 size. Small insertions in the polymerase domain were primarily observed in loops of the fingers and thumb subdomains. Within the catalytic palm subdomain, the location of conserved RdRP motifs A–E, the only predicted insertion for

RVG 03V0567 VP1 was of three amino acids in the priming loop. A small deletion was predicted in the loop between motifs C and D.

Interactions between VP1 and viral RNA

To gain insight into potential mechanisms of viral (+)RNA recognition by VP1 from different RV species, we generated structure-based sequence alignments of VP1 homology models and RVA SA11 VP1 (Fig. S1). Positions in RVA SA11 VP1 occupied by amino acids that interact directly with UGUGACC RNA oligonucleotides are located in the template entry tunnel and are contributed by the fingers, palm, and thumb subdomains of the polymerase domain and by the N-terminal domain (Figs. 2 and 3) (Lu et al., 2008). Some of these amino acids interact with the RNA ribose-phosphate backbone in a sequence-independent manner, while others interact directly with specific bases in the UGUG cassette. The location and identity of amino acids involved in sequence-independent interactions with RNA were primarily conserved among all of the VP1 models (Fig. S1). For example, four lysine side chains in RVA SA11 VP1 that form H-bonds to the RNA ribose-phosphate backbone (amino acids 419, 420, 594, and 597), were conserved in all VP1 homology models (Figs. 3, S1) (Lu et al., 2008). Ser398 and Ser401 are located in a loop and interact with the ribose-phosphate backbone. In each homology model, this loop adopted a slightly different conformation than the same region of RVA SA11 VP1 (Fig. 3). These differences, however, reflect changes induced in the RVA SA11 VP1 entry tunnel due to RNA binding, and the models closely matched the structure of the RVA SA11 VP1 apoenzyme in this region (data not shown). Based on apoenzyme structure, Ser401 was conserved among all RV groups, and Ser398 was conserved among clade A isolates, including RVF, but not among clade B isolates, including RVG (Fig. S1).

Some conservation and some differences were observed among models of VP1 from clade A and clade B viruses at positions in RVA SA11 VP1 involved in sequence-specific interactions with the U7 and G6 bases in the UGUG (+)RNA recognition cassette. RVA SA11 VP1 interacts with the pyrimidine U7 base through two H-bonds between main chain atoms of Phe416, and the Phe416 phenyl ring stacks against the nucleotide base (Fig. 3) (Lu et al., 2008). The clade A RVF 03V0568 VP1 model contained a phenylalanine at this position, but it was an arginine in other clade A VP1 models and a lysine in the clade B VP1 models (Fig. S1). While base-specific main chain interactions would not likely be affected by changes at this position, stacking could be altered. The side chains of Asn186 and Lys188 in RVA SA11 VP1 form H-bonds with the G6 base, and the Arg190 side chain stacks against the G6 purine ring (Fig. 3) (Lu et al., 2008). The location and identity of Asn186 was conserved in all VP1 models, with the exception of those for RVF and RVB, which contained aspartic acid at this position (Fig. S1). Lys188 was conserved in all models, except that of RVC Bristol, in which it was replaced by an arginine. Arg190 was conserved among clade A viruses, but this position was occupied by serine for the majority of clade B viruses.

Clade-specific variation was observed for residues involved in interactions with the U5 and G4 RNA bases in RVA SA11 VP1. These bases interact via H-bonds with main chain atoms in Arg701 and Gly702 (Fig. 3) (Lu et al., 2008). Thus, the identity of the amino acids at these positions is not likely as important as local polymerase structure. In VP1 homology models, length and sequence in the vicinity Arg701 and Gly702 were well conserved among clade A viruses and among clade B viruses, but they differed significantly between the two clades (Figs. 3 and S1). For the clade B viruses, an insertion of three amino acids was predicted to extend the length of the loop in which these positions reside. Such an extension is anticipated to alter the shape of the template entry tunnel, such that it would be unable to engage the 3′ terminus of (+)RNA in the same conformation adopted by UGUGACC. Rather, the loop would appear to clash with the U5 and G4 bases. Together, clade-specific

differences in homology models suggest that clade A and B VP1 interact differently with the 3′ terminus of viral (+)RNA.

Discussion

RVF and RVG are distinct species

Analyses presented herein support the idea that RVF 03V0568 and RVG 03V0567 represent distinct RV species. Previous studies showed that RVF 03V0568 and RVG 03V0567 dsRNA genome segments have unique electrophoretic migration patterns, that the viruses failed to cross-react with antisera raised against isolates from other RV species, and that VP6 (+)RNA and amino acid sequences occupied distinct branches from those of other RV species in phylogenetic trees (Johne et al., 2011; McNulty et al., 1984). Here, we demonstrate that RVF 03V0568 and RVG 03V0567 VP1 (+)RNA and amino acid sequences also occupy distinct branches of phylogenetic trees, with respect to isolates from RVA, RVB, RVC, RVD, RVH, and one another (Fig. 1) (Johne et al., 2011). Due to a high level of conservation, the sequences of RdRPs, such as VP1, tend to be used to distinguish evolutionary relationships among *Reoviridae* viruses (Attoui et al., 2012). The variability in VP1 RNA sequence (>35%) and amino acid identity (>40%) between RVF or RVG isolates and isolates from other RV species (Table 3) is consistent with accepted rotavirus species demarcation criteria.

RVs resolve into two distinct clades

RV isolates historically have been classified based on several properties. When one considers the number of dsRNA genome segments, overall virion morphology, disease associated with infection, and host range, isolates from both clade A and clade B have much in common. However, several differences distinguish these two groups. It is clear that clade A and clade B viruses are phylogenetically distinct from one another, as exemplified by the nucleotide and amino acid sequences of their VP1 and VP6 segments (Table 3) (Attoui et al., 2012; Johne et al., 2011). The 3′-terminal UGUG (+)RNA recognition cassette is conserved among all known clade A viruses to date, but not for the clade B viruses (Table 2). Furthermore, predicted RdRP structures appear to be fairly conserved within RV clades, but to differ between them (Figs. 2–4 and S1). Finally, clade B VP1 is longer overall than clade A VP1, and it is predicted to contain several insertions, primarily located in the N-terminal domain, in comparison to RVA SA11 VP1.

According to the International Committee on Taxonomy of Viruses, *Reoviridae* viruses from different genera usually have <26% RdRP amino acid identity, whereas within a single genus RdRP identities tend to be >33% (Attoui et al., 2012). Clade A and clade B RV isolates share <22% RdRP amino acid identity (Table 3). Together with the aforementioned distinguishing features, these observations raise the possibility that clade A and clade B might reasonably be classified as distinct virus genera. Completion of genome sequencing for at least a single representative of each RV group, in combination with studies directed at characterizing in greater detail the virion structures, genomes, functions of encoded proteins, and disease pathogenesis might inform decisions regarding classification of viruses currently considered as members of the RV genus.

Clade A and B rotaviruses differ in the mechanism viral (+)RNA recognition

Conservation of sequences near the 3′ terminus of (+)RNA and similarities in predicted VP1 structures suggest that clade A viruses employ a conserved mechanism of +RNA recognition. Clade A viruses share a high level of conservation of a UGUG cassette at positions four through seven from 3′ terminus of viral (+)RNA (Table 2) (Attoui et al., 2012; Estes and Kapikian, 2007; Johne et al., 2011). The UGUG cassette is not completely

conserved, as evidenced by identification of a UAUGACC sequence at the 3′ terminus of the VP6-encoding segment of RVF 03V0568 (Johne et al., 2011). However, there tends to be a preference for UGUG, and a purine is maintained at the 6th-to-terminal position in this altered sequence. For RVA SA11 VP1, the UGUG cassette is important for (+)RNA binding by VP1, and the UGUG bases are specifically recognized by amino acids in the VP1 template entry tunnel (Lu et al., 2008; Tortorici et al., 2003). Conservation, in clade A VP1 homology models, of the identity and location of nearly all of the amino acids involved in (+)RNA recognition by RVA SA11 VP1 suggests that the UGUG cassette is recognized using a conserved mechanism (Figs. 3 and S1) (Lu et al., 2008). The extraordinarily high quality of clade A VP1 homology models (Table 4) suggests that direct structural comparisons with RVA SA11 VP1 are biologically meaningful for these RdRP. In the few cases where amino acid identity differed between RVA SA11 VP1 and another clade A virus at a position involved in (+)RNA recognition, the non-conserved residues were generally anticipated to have little effect on (+)RNA binding (Figs. 3 and S1). For example, replacement of Lys188 with arginine for RVC Bristol is a conservative change that is unlikely to prevent interactions with the G6 base. For Gly702, which was replaced by asparagine or threonine in some clade A isolates, it is the main chain, rather than the side chain, that is involved in interactions with (+)RNA. Similarly, while replacement of Phe416 with arginine might alter the capacity of this residue to stack with the U7 base, H-bonds with the main chain atoms would unlikely be affected. In contrast to these situations, it is likely that replacement of Asn186 with aspartic acid for RVF 03V0568 would alter interactions with the G6 base. It is possible that a neighboring amino acid performs the Asn186 function for RVF. Future biochemical studies might bolster the hypothesis that all clade A viruses use a similar mechanism of (+)RNA recognition.

Clade B RVs appear to employ a different mechanism of (+)RNA recognition than clade A RVs. The extreme 3′-terminal sequence, ACCC, is conserved among clade B viruses (Table 2). However, upstream sequences vary between isolates from different species within clade B. A UGUG sequence is not observed near the 3′ termini of any of the clade B segments, and there does not appear to be a specific pattern of purine or pyrimidine conservation in the region just upstream of the 3′ terminus. Similar to the 3′-terminal CC residues of clade A (+)RNA, the terminal ACCC sequence might be important for the initiation of RNA synthesis by VP1. It is possible, however, that for clade B viruses the RdRP also specifically recognizes the extreme 3′ (+)RNA terminus.

The observation that nearly all of the amino acids involved in sequence-independent RNA binding by RVA SA11 VP1 are structurally conserved within VP1 homology models from representatives of each RV group (Figs. 3 and S1) suggests that all RV VP1 proteins engage the RNA ribose-phosphate backbone in the template entry tunnel using a similar mechanism. A single clade-specific difference in amino acid identity at RVA SA11 VP1 position Ser398 suggests only a minor difference in the mechanism of sequence-independent interaction with RNA between clade A and clade B viruses. Conservation of residues involved in sequence-independent interactions is not surprising, considering that the majority of the interactions are based on charge and that each RdRP must guide RNA molecules of varying sequence through the entry tunnel to the catalytic site. In terms of sequence-dependent interactions, some potentially more significant differences are predicted between clade A and clade B VP1. Perhaps the most significant of these differences is an insertion of three amino acids in a loop analogous to the one containing Arg701 and Gly702 in RVA SA11 VP1. This insertion is predicted for all homology models of clade B VP1 (Fig. 3 and S1). Such a change would narrow and alter the shape of the entry tunnel in a region that is important for specific interaction with the G4 and U5 bases in the RVA RdRP. Thus, clade B viruses are unlikely to bind RNA in precisely the same orientation as the clade A viruses. Future

structural and biochemical studies might reveal the mechanisms by which clade B VP1 specifically recognizes cognate viral (+)RNA.

Materials and methods

Sequence determination of RVF and RVG genome segments encoding VP1

Virus strains 03V0567 (RVG) and 03V0568 (RVF) were derived from the intestinal contents of stunted chicks (Johne et al., 2011). Isolation of genomic dsRNA, amplification by a full-length amplification of cDNA (FLAC) technique and by degenerated primers, and sequencing of the VP1-encoding genome segment were performed essentially as described (Johne et al., 2011). Briefly, the isolated dsRNA was subjected to ligation with the specific DNA adaptor iSP9 (Johne et al., 2011) to the dsRNA ends. After electrophoretic separation of dsRNA genome segments on an ethidium bromide-stained 1% agarose gel, the slowest-migrating dsRNA bands, with positions approximately corresponding to 3.5 kbp, were excised. Only products shorter than 1 kbp could be generated by FLAC amplification of these segments, which were subsequently cloned into vector pCR4-TOPO (Invitrogen, The Netherlands) and sequenced. Based on the generated sequences and on an alignment of sequences of the VP1-encoding segments of several rotavirus groups (not shown), specific and degenerated primers were constructed: primers 5′-TAA TAA AAT CAA AGC TAG AAC TA-3′, 5′-AAA TAT GCA TGT TAT GTC TGA C-3′, 5′-GCA TTA ACT GCT TCC TAT ATC C-3′ for RVF 03V0568, and primers 5′-ATC AGC TAT TCG CAC CCC TAA C-3′, 5′-AAG GCT CTA GCA TCA TGG ACG G-3′, 5′-TAT TGC TCC TCT CTC AAA GA-3′, 5′-AGA ATA GAC AAT SCW CAA GAT GC-3′ and 5′-CTG TAT TAT GTT GCG ACG CAT CCC-3′ for RVG-03V0567. The primers were used in combination with primer 5-15-1 (complementary to the ligated adaptor iSP9) under variable RT-PCR conditions employing the QIAGEN LongRange 2Step RT-PCR Kit (Qiagen, Hilden, Germany). PCR products with the expected lengths were cloned and sequenced as described (Johne et al., 2011).

Sequence analysis

The sequences of the VP1-encoding genome segments for RVF 03V0568 and RVG 03V0567 were assembled using the SeqBuilder module of the DNASTAR software package (Lasergene, Madison, USA) and submitted to the GenBank database with accession numbers GenBank ID: JN596591 (RVF 03V0568) and GenBank ID: JN596592 (RVG 03V0567). The VP1 amino acid sequences were deduced from the nucleotide sequences using the same module of the above mentioned software package. Maximum likelihood phylogenetic trees were constructed using PhyML (Guindon and Gascuel, 2003) employing the Hasegawa-Kishino-Yano substitution model (HKY85) and gamma-distributed rate variation among sites. Bootstrap analysis was performed based on 1000 replicates, and trees were visualized using Geneious, version 5.4.4 (Drummond et al., 2010). Pairwise and multiple alignments for nucleotide and amino acid identity matrices were made using the ClustalW function in MacVector, version 12.0.6, with default settings. Accession numbers for RVA, RVB, RVC, RVD, and RVH sequences used in the analyses are available upon request.

VP1 homology modeling

Homology models of group B, C, D, F, G, and H VP1 were generated using the I-TASSER server (Roy et al., 2010; Zhang, 2007). Accession numbers for amino acid sequences of representative strains from RVB, RVC, RVD, and RVH that were input into I-TASSER are NC_007547 (Bristol), NC_014511 (05V0049), EU490415 (Bang373), GQ358719 (DB176), EF453355 (B219), and NC_007548 (J19). C-scores, TM-scores, and RMSDs were calculated by the I-TASSER server.

VP1 models were structurally aligned using the MatchMaker function of UCSF Chimera (Petterson, 2004), alpha version 1.6, with default settings, PDB file 2R7Q (RVA SA11 VP1 apoenzyme) or 2R7W (RVA SA11 VP1 complex with UGUGACC oligo) (Lu et al., 2008), and PDB files for models generated by I-TASSER. Structure-based sequence alignments of VP1 homology models and RVA SA11 VP1 were generated using UCSF Chimera's Match–>Align function, with a residue-residue distance cutoff of 5Å and iteration until convergence in stretches of at least three consecutive columns. Consensus sequences for Clade A and Clade B VP1 were made by comparison of the structure-based sequence alignments.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

K.M.O. and J.T.P were supported by the Intramural Research Program of the National Institute of Allergy and Infectious Diseases, National Institutes of Health (USA). R.J. was supported by a grant from the Deutsche Forschungsgemeinschaft (JO 369/4-1).

References

Ahmed MU, Kobayashi N, Wakuda M, Sanekata T, Taniguchi K, Kader A, Naik TN, Ishino M, Alam MM, Kojima K, Mise K, Sumi A. Genetic analysis of group B human rotaviruses detected in Bangladesh in 2000 and 2001. J Med Virol. 2004; 72:149–155. [PubMed: 14635024]

Attoui, H.; Becnel, J.; Belaganahalli, S.; Bergoin, M.; Brussaard, CP.; Chappell, JD.; Ciarlet, M.; del Vas, M.; Dermody, TS.; Dormitzer, PR.; Duncan, R.; Fcang, Q.; Graham, R.; Guglielmi, KM.; Harding, RM.; Hillman, B.; Makkay, A.; MarzachiÃ, C.; Matthijnssens, J.; Mertens, PPC.; Milne, RG.; Mohd Jaafar, F.; Mori, H.; Noordeloos, AA.; Omura, T.; Patton, JT.; Rao, S.; Maan, M.; Stoltz, D.; Suzuki, N.; Upadhyaya, NM.; Wei, C.; Zhou, H. Part II: The Viruses – The Double Stranded RNA Viruses - Family Reoviridae. In: King, AMQ.; Adams, MJ.; Carstens, EB.; Lefkowitz, EJ., editors. Virus taxonomy: classification and nomenclature: Ninth Report of the International Committee on Taxonomy of Viruses. Elsevier Academic Press; San Diego: 2012. p. 541-637.

Chen Z, Lambden PR, Lau J, Caul EO, Clarke IN. Human group C rotavirus: completion of the genome sequence and gene coding assignments of a non-cultivatable rotavirus. Virus Res. 2002; 83:179–187. [PubMed: 11864750]

Drummond, AJ.; Ashton, B.; Buxton, S.; Cheung, M.; Cooper, A.; Duran, C.; Field, M.; Heled, J.; Kearse, M.; Markowitz, S.; Moir, R.; Stones-Havas, S.; Sturrock, S.; Thierer, T.; Wilson, A. Geneious v5.4.4. 2010. Available from http://www.geneious.com

Estes, MK.; Kapikian, AZ. Rotaviruses. In: Howley, PM., editor. Fields Virology. 5. Lippincott, Williams and Wilkins; Philadelphia: 2007. p. 1918-1974.

Ghosh S, Kobayashi N, Nagashima S, Chawla-Sarkar M, Krishnan T, Ganesh B, Naik TN. Molecular characterization of the VP1, VP2, VP4, VP6, NSP1 and NSP2 genes of bovine group B rotaviruses: identification of a novel VP4 genotype. Arch Virol. 2010; 155:159–167. [PubMed: 19936611]

Guindon S, Gascuel O. A simple, fast, and accurate algorithm to estimate large phylogenies by maximum likelihood. Syst Biol. 2003; 52:696–704. [PubMed: 14530136]

Jiang S, Ji S, Tang Q, Cui X, Yang H, Kan B, Gao S. Molecular characterization of a novel adult diarrhoea rotavirus strain J19 isolated in China and its significance for the evolution and origin of group B rotaviruses. J Gen Virol. 2008; 89:2622–2629. [PubMed: 18796732]

Johne R, Otto P, Roth B, Lohren U, Belnap D, Reetz J, Trojnar E. Sequence analysis of the VP6-encoding genome segment of avian group F and G rotaviruses. Virology. 2011; 412:384–391. [PubMed: 21329955]

Lu X, McDonald SM, Tortorici MA, Tao YJ, Vasquez-Del Carpio R, Nibert ML, Patton JT, Harrison SC. Mechanism for coordinated RNA packaging and genome replication by rotavirus polymerase VP1. Structure. 2008; 16:1678–1688. [PubMed: 19000820]

Matthijnssens J, Otto PH, Ciarlet M, Desselberger U, Van Ranst M, Johne R. VP6 sequence-based cut-off values as a criterion for rotavirus species demarcation. Arch Virol. 2012 in press.

McNulty MS, Todd D, Allan GM, McFerran JB, Greene JA. Epidemiology of rotavirus infection in broiler chickens: recognition of four serogroups. Arch Virol. 1984; 81:113–121. [PubMed: 6331344]

Nagashima S, Kobayashi N, Ishino M, Alam MM, Ahmed MU, Paul SK, Ganesh B, Chawla-Sarkar M, Krishnan T, Naik TN, Wang YH. Whole genomic characterization of a human rotavirus strain B219 belonging to a novel group of the genus Rotavirus. J Med Virol. 2008; 80:2023–2033. [PubMed: 18814255]

Petterson EF, Goddard TD, Huang CC, Couch GS, Greenblatt DM, Meng EC, Ferrin TE. UCSF chimera-A visualization system for exploratory research and analysis. J Computational Chem. 2004; 25:1605–1612.

Roy A, Kucukural A, Zhang Y. I-TASSER: a unified platform for automated protein structure and function prediction. Nat Protoc. 2010; 5:725–738. [PubMed: 20360767]

Settembre EC, Chen JZ, Dormitzer PR, Grigorieff N, Harrison SC. Atomic model of an infectious rotavirus particle. EMBO J. 2011; 30:408–416. [PubMed: 21157433]

Tortorici MA, Broering TJ, Nibert ML, Patton JT. Template recognition and formation of initiation complexes by the replicase of a segmented double-stranded RNA virus. J Biol Chem. 2003; 278:32673–32682. [PubMed: 12788926]

Trojnar E, Otto P, Johne R. The first complete genome sequence of a chicken group A rotavirus indicates independent evolution of mammalian and avian strains. Virology. 2009; 386:325–333. [PubMed: 19246068]

Trojnar E, Otto P, Roth B, Reetz J, Johne R. The genome segments of a group D rotavirus possess group A-like conserved termini but encode group-specific proteins. J Virol. 2010; 84:10254–10265. [PubMed: 20631147]

Zhang Y. Template-based modeling and free modeling by I-TASSER in CASP7. Proteins. 2007; 69(Suppl 8):108–117. [PubMed: 17894355]

HIGHLIGHTS

- VP1, the rotavirus RNA polymerase, specifically recognizes viral (+)RNA
- We report and analyze VP1-encoding sequences for species F and G rotaviruses
- Rotavirus species cluster phylogenetically into two distinct clades
- Clade A VP1 structures and homology models differ from those of clade B
- The mechanism of viral (+)RNA recognition appears conserved for clade A VP1

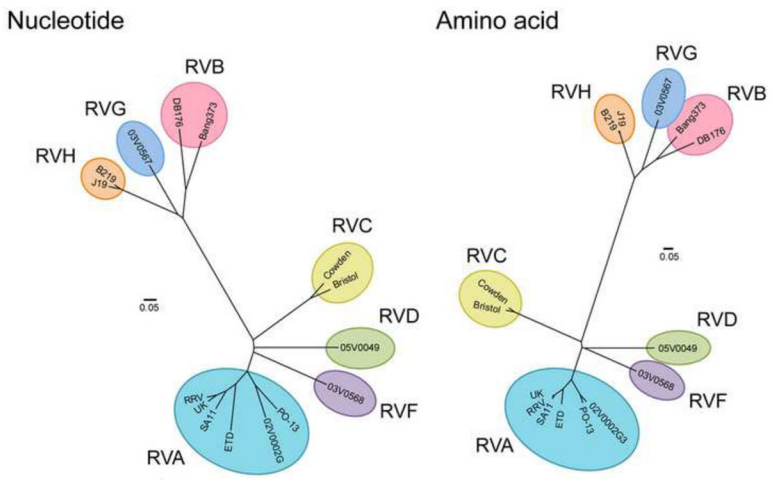


Figure 1.

Maximum likelihood phylogenetic trees comparing the nucleotide sequences of VP1 (+)RNA (left) or the deduced amino acid sequences of RdRP VP1 (right) for several RV isolates. The species of isolates are indicated. Scale bar represents 0.05 substitutions per site.

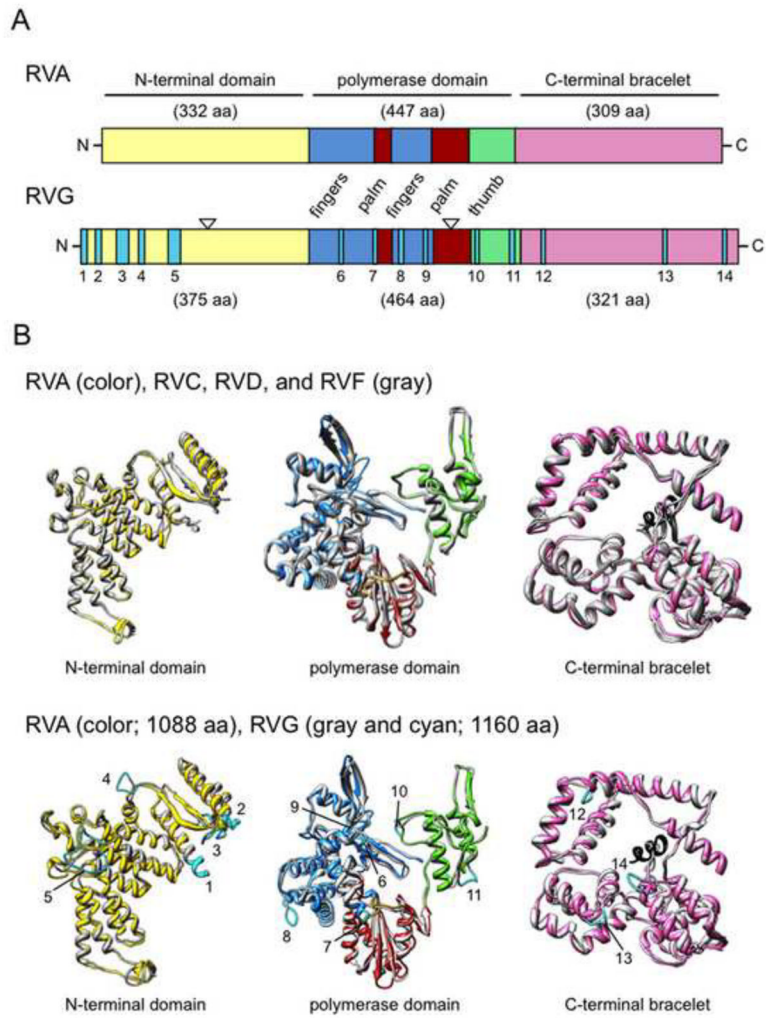
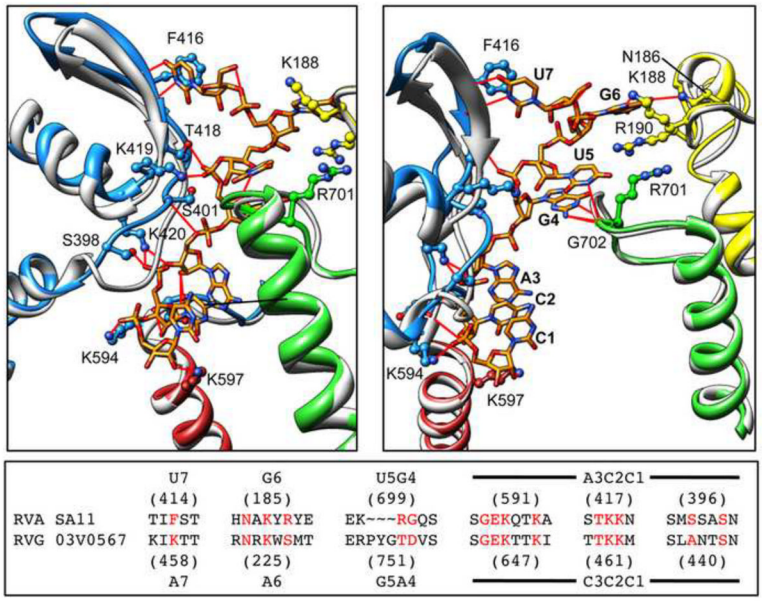


Figure 2.

Comparison of predicted VP1 structures. (A) Schematic bar representation of VP1 colored by domain (N-terminal, yellow; C-terminal, pink) and subdomain (fingers, blue; palm, brick red; thumb, green). The length of each domain for RVA SA11 VP1 (upper) and each predicted domain for RVG 03V0567 VP1 (lower) is indicated. Locations of predicted insertions in RVG VP1 that are greater than three amino acids in length, based on homology modeling, are indicated with cyan bars and numbered. Open triangles indicate deletions of three or more amino acids. (B) Ribbon drawings comparing structures of the N-terminal (left), polymerase (center), and C-terminal (right) domains of RVA SA11 VP1 to those of VP1 homology models. The models included in the upper panel are of Bristol (RVC), 05V0049 (RVD), and 03V0568 (RVF) VP1, and in the lower panel is the model of 03V0567 (RVG) VP1. RVA SA11 VP1 is colored as in panel (A). Homology models are colored gray. The locations of predicted insertions in RVG VP1 greater than three amino acids in length are colored cyan and numbered as in panel (A).



	U7	G6	U5G4		A3C2C1	
	(414)	(185)	(699)	(591)	(417)	(396)
RVA SA11	TIFST	HNAKYRYE	EK~~~RGQS	SGEKQTKA	STKKN	SMSSASN
RVG 03V0567	KIKTT	RNRKWSMT	ERPYGTDVS	SGEKTTKI	TTKKM	SLANTSN
	(458)	(225)	(751)	(647)	(461)	(440)
	A7	A6	G5A4		C3C2C1	

Figure 3.

Viral (+)RNA recognition by clade A and clade B VP1. (Upper) Close up of the template entry tunnel of RVA SA11 VP1 (colored ribbon) with bound RNA oligonucleotide (orange sticks) in two slightly different views. PDB 2R7W was used to make the images. RNA bases are labeled and numbered in the 3′ to 5′ direction. RVA SA11 VP1 subdomains are colored as in Figure 2. Side chains of VP1 residues that interact with the RNA are shown and labeled. The predicted structure of RVG 03V0567 VP1 (gray ribbon) is overlaid on the RVA structure. Parts of the structure have been hidden for clarity. (Lower) Structure-based sequence alignment of specific regions of VP1, made using PDB 2R7Q and the RVG 03V0567 homology model. RVA SA11 residues that interact with RNA, and the aligning residues in the RVG VP1 model, are colored red. The initial amino acid number in the series and the RNA nucleotide with which the RVA SA11 VP1 residues interact are indicated above the alignment. The identities of the aligning amino acids and aligning nucleotides for RVG 03V0567 are indicated below the alignment.

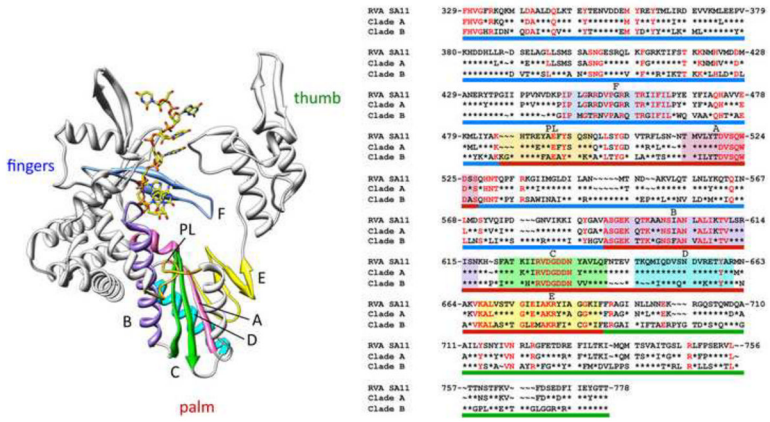


Figure 4.

Consensus alignments for the polymerase domain of clade A and clade B VP1. (Left) Ribbon drawing of the polymerase domain of RVA SA11 VP1 (PDB 2R7W). Conserved RdRP motifs are colored light blue (motif F), pink (motif A), lavender (motif B), green (motif C), teal (motif D), and yellow (motif E), and the 'priming loop' is colored gold. A bound UGUGACC RNA oligonucleotide (yellow sticks) is shown for orientation. (Right) Alignments of the RVA SA11 VP1 polymerase domain sequence and consensus sequences for clade A (SA11, Bristol, 05V0049, and 03V0568) and clade B (Bang373, DB176, 03V0567, B219, and J19) polymerase domains. Alignments are based primarily on predicted VP1 structures and PDB 2R7Q, with regions of discrepancy aligned by hand. Tildas indicate insertions. Asterisks indicate amino acids that differ within a clade. Single-letter codes for amino acids that are conserved within a clade are in black text, and amino acids that are conserved between both clades are in red text. Fingers (blue), palm (brick red), and thumb (green) subdomains are indicated by colored bars. Amino acid sequences of RdRP motifs are highlighted with colors similar to those in the left panel and labeled.

Table 1

Characteristics of rotavirus RNA polymerases.

			Protein		RNA			
Clade	**Species**	**Strain**	**MW (kd)**	**AA**	**NT**	**5′-UTR**	**3′-UTR**	**ORF**
A	RVA	SA11	125.3	1088	3302	18	17	3267
A	RVA	RRV	125.1	1088	3302	18	17	3267
A	RVA	UK	125	1088	3302	18	17	3267
A	RVA	ETD	124.5	1088	3302	18	17	3267
A	RVA	PO-13	125.1	1088	3302	18	17	3267
A	RVA	02V0002G3	125.1	1089	3305	18	17	3270
A	RVC	Bristol	125.8	1090	3309	11	25	3273
A	RVD	05V0049	125.4	1079	3274	18	16	3240
A	RVF	03V0568	125.4	1086	3296	18	17	3261
B	RVB	Bang373	132.1	1160	3511	8	20	3483
B	RVB	DB176	131.3	1158	3504	7	20	3477
B	RVG	03V0567	133	1160	3526	14	29	3483
B	RVH	B219	132.9	1167	3538	6	28	3504
B	RVH	J19	132.9	1167	3538	6	28	3504

Table 2

Terminal sequences of RV (+)RNA encoding VP1

	% Nucleotide Identity														
Strain	**SA11**	**RRV**	**UK**	**ETD**	**PO-13**	**02V0[a]**	**Brist**	**05V0**	**03V0**	**Bang**	**DB17**	**03V0**	**B219**	**J19**	
SA11		86	85	75	72	71	57	59	60	41	40	41	42	41	RVA
RRV	97		92	76	72	71	56	60	60	41	40	41	41	41	RVA
UK	97	99		75	71	70	56	59	60	40	40	40	41	40	RVA
ETD	86	87	87		68	68	53	56	59	40	39	39	39	39	RVA
PO-13	76	77	77	76		79	57	59	61	40	40	41	42	41	RVA
02V0002G3	76	76	77	76	92		56	59	61	41	41	41	42	41	RVA
Bristol	46	47	47	47	47	47		56	56	39	40	42	41	40	RVC
05V0049	50	50	50	50	50	50	45		59	39	39	41	41	40	RVD
03V0568	54	54	54	54	54	54	46	52		39	38	40	40	40	RVF
Bang373	20	20	20	21	20	20	19	19	20		68	59	59	58	RVB
DB176	19	20	19	20	20	21	20	18	22	69		60	58	58	RVB
03V0567	19	19	19	20	20	20	20	19	21	60	58		64	63	RVG
B219	21	21	21	20	21	21	20	20	21	56	53	58		93	RVH
J19	21	21	21	20	21	21	20	20	21	56	53	58	97		RVH
	RVA	RVA	RVA	RVA	RVA	RVA	RVC	RVD	RVF	RVB	RVB	RVG	RVH	RVH	Species
	% Amino Acid Identity														

[a]Only the first 4–5 characters were used to identify some strains.

Table 3

Sequence identity among VP1 proteins of different rotavirus species.

Clade	Species	Strain	5′ End[a]	3′ End[a]
A	RVA	SA11	GGCUAUUAAA	AGAUGUGACC
A	RVA	RRV	GGCUAUUAAA	AGAUGUGACC
A	RVA	UK	GGCUAUUAAA	AGAUGUGACC
A	RVA	ETD	GGCUAUUAAA	CGAUGUGACC
A	RVA	PO-13	GGCUAUUAAA	AGAUAUGACC
A	RVA	02V0002G3	GGCUAUUAAA	CGAUGUGACC
A	RVC	Bristol	GGCUAAAAAA	UAUUGUGGCU
A	RVD	05V0049	GGCAAUUAAA	ACUUGUGACC
A	RVF	03V0568	GGCAAUUAAU	UAAUGUGACC
B	RVB	Bang373	GGCACAAUAU	AUAAAAACCC
B	RVB	DB176	GGCAAUAAUG	GUAAAAACCC
B	RVG	03V0567	GGCAUAUAUU	AUAAAGACCC
B	RVH	B219	GGCACUAUGG	UAAUAUACCC
B	RVH	J19	GGCACUAUGG	UAAUAUACCC

[a]Sequences are written from 5′ to 3′.

Table 4

VP1 structure predictions by the I-TASSER server.

Species	Strain	C-score[a]	TM-score[b]	RMSD[c]
RVC	Bristol	2.1	0.99 ± 0.04	4.5 ± 3.0
RVD	05V0049	2.64	0.99 ± 0.03	3.4 ± 2.4
RVF	03V0568	2.63	0.99 ± 0.03	3.5 ± 2.4
RVB	Bang373	0.14	0.73 ± 0.11	8.9 ± 4.6
RVB	DB176	0.53	0.78 ± 0.11	8.0 ± 4.4
RVG	03V0567	0.77	0.82 ± 0.09	7.5 ± 4.3
RVH	B219	−0.09	0.70 ± 0.12	9.5 ± 4.6
RVH	J19	0.23	0.74 ± 0.11	8.7 ± 4.6

[a]Confidence score (C-score) estimates the quality of the predicted model.

Scores typically range from [−5, 2], with higher C-scores indicating greater confidence.

[b]Template modeling ™ scores range from [0, 1], with 1 indicating a perfect match.

[c]Root mean square deviation (RMSD) is the average distance (in Å) between backbone atoms.