Published in final edited form as:
Virology. 2014 April ; 0: 25–33. doi:10.1016/j.virol.2014.01.018.

Discovery of rosavirus 2, a novel variant of a rodent-associated picornavirus, in children from The Gambia

Efrem S. Lim[1], **Song Cao**[1], **Lori R. Holtz**[2], **Martin Antonio**[3], **O. Colin Stine**[4], and **David Wang**[1,#]

[1]Departments of Molecular Microbiology and Pathology & Immunology, Washington University School of Medicine, St. Louis, Missouri, USA

[2]Department of Pediatrics, Washington University School of Medicine, St. Louis, Missouri, USA

[3]Medical Research Council Unit, Banjul, The Gambia

[4]Department of Epidemiology and Public Health, University of Maryland School of Medicine, Baltimore, Maryland, USA

Abstract

We describe the identification of a novel picornavirus recovered from the fecal specimen of a child in The Gambia, provisionally named rosavirus 2. Comparison of the rosavirus 2 complete genome demonstrated 71.9% nucleotide identity to its closest relative rosavirus M-7, an unclassified picornavirus identified from rodent fecal material. A unique RNA structure was predicted in the 3′ UTR of rosavirus 2 that was conserved with rosavirus M-7 and cadiciviruses. We detected rosavirus 2 in four pediatric fecal specimens (0.55% prevalence) in a Gambian diarrheal case-control cohort, but we did not detect it in a panel of 634 pediatric diarrheal stool specimens from USA. There was no statistical evidence that rosavirus 2 was associated with diarrheal cases. This study broadens our understanding of unknown viruses present in children in developing country settings.

Keywords

Picornavirus; virus discovery; emerging viruses; developing country

Introduction

Diarrhea is the leading cause of morbidity and mortality among children less than 5 years old in developing countries. In Africa, more than 25% of infant mortality is due to diarrhea (Walker et al., 2012). The viral aetiology can be linked to rotaviruses, caliciviruses,

#Corresponding author. Tel.: +131 428 611 23; fax: +131 436 212 32. Mailing address: Washington University School of Medicine, Campus Box 8230, 660 S. Euclid Ave., St. Louis, MO 63110, USA, davewang@borcim.wustl.edu.

astroviruses and enteric adenoviruses. However, approximately 40% of diarrhea cases have an unknown etiology (Chikhi-Brachet et al., 2002; Denno et al., 2007; Kapikian, 1993). To address this, large prospective studies such as the Global Enteric Multi-center Study (GEMS) have been conducted to provide multi-year clinical and epidemiological insight into diarrheal diseases in sub-Saharan Africa and South Asia (Kotloff et al., 2012). In this study, we analyzed fecal samples collected from children in The Gambia as part of the GEMS study and report the identification of a novel picornavirus.

Picornaviruses are a family of single stranded, positive sense RNA viruses. Recent studies suggest that picornaviruses, such as human kobuvirus, cardiovirus, and salivirus, might also be associated with acute gastroenteritis (Ambert-Balay et al., 2008; Holtz et al., 2009; Li et al., 2009; Pham et al., 2007; Ren et al., 2009; Shan et al., 2010). Other picornaviruses, such as enteroviruses and parechoviruses are frequently detected in the gastrointestinal tract, but are not thought to be enteric pathogens. On the other hand, poliovirus is shed in feces for extended periods of time but causes a neurologic disease rather than gastrointestinal disease (Hird and Grassly, 2012). There are 17 genera in the *Picornaviridae* family: *Aphthovirus, Aquamavirus, Avihepatovirus, Cardiovirus, Cosavirus, Dicipivirus, Enterovirus, Erbovirus, Hepatovirus, Kobuvirus, Megrivirus, Parechovirus, Salivirus, Sapelovirus, Senecavirus, Teschovirus* and *Tremovirus* (Adams et al., 2013; Knowles, 2012). Picornaviruses typically encode a single polyprotein that is proteolytically cleaved by viral-encoded proteases. However, this ‘single polyprotein’ paradigm was challenged with the identification of cadiciviruses (a picornavirus-like virus, formerly known as canine picodicistrovirus) that encodes two polyproteins separated by an internal ribosome entry site (IRES) element (Woo et al., 2012). As such, cadicivirus was proposed to be the evolutionary ‘missing link’ between the *Picornaviridae* and *Dicistroviridae* families (Woo et al., 2012). However, the debate over the evolutionary origin and diversification of viruses in the *Picornavirales* order remains unresolved particularly due to the limited roster of known picorna-like viruses (Koonin et al., 2008; Le Gall et al., 2008). In this regard, the identification and characterization of novel picornaviruses around the evolutionary space of the ‘missing link’ might clarify the evolutionary history of the *Picornavirales* order.

A hallmark of picornaviruses is the presence of extensive RNA secondary structures in the genome critical to viral replication. Secondary structures in the 5′ UTR regions typically form an IRES element required for the recruitment of the ribosomal translation initiation complex to allow cap-independent translation initiation (reviewed in (Martinez-Salas, 2008)). Similarly, the secondary structures formed in the 3′ UTR region are essential for picornavirus replication. For example, poliovirus replication is dependent on binding of host proteins to the 3′ UTR region for circularization and genome replication (Herold and Andino, 2001). The 3′ UTR of Kobuviruses also share a ‘barbell’ structure that is conserved in Avihepatoviruses that is thought to be essential for viral replication (Boros et al., 2012). As a result, RNA secondary structures in UTR regions of picornaviruses might be structurally well conserved between picornavirus members despite their high sequence diversity.

Here, we describe the identification of a novel picornavirus, provisionally named rosavirus 2, through the deep sequencing of a fecal specimen from a child in The Gambia. The

complete genome of rosavirus 2 shared 71.9% nucleotide identity to rosavirus M-7, a picornavirus whose partial genome was identified in rodent stool (Phan et al., 2011). We found that cadicivirus, rosavirus 2 and rosavirus M-7 form a monophyletic clade within the *Picornaviridae* family. We developed an RT-PCR assay to detect rosavirus 2 and screened fecal specimens from a pediatric cohort of primarily diarrheal cases in Saint Louis, USA and a pediatric diarrheal case-control cohort from The Gambia. We detected rosavirus 2 in 4 out of 722 specimens from The Gambia (0.55% prevalence) but none of the Saint Louis diarrhea samples were positive. There was no statistically significant evidence of association between rosavirus 2 with diarrheal cases. These results underscore the diversity of unknown viruses that remain to be discovered in children.

Materials and Methods

Clinical Specimens

The study was approved by the Human Research Protection Office (HRPO) of Washington University in St. Louis, Missouri, USA; the Institutional Review Board of the University of Maryland Baltimore, Baltimore, Maryland, USA, and the Joint Medical Research Council/ Gambia Government Ethics Committee, Fajara, The Gambia.

The index stool specimen was obtained in October 29 2008 from a healthy, 16 month-old female living in The Upper River Region in The Gambia as part of a Global Enteric Multi-center Study (GEMS) (Kotloff et al., 2012). 722 fecal specimens (332 cases and 390 controls) were randomly selected from a total of 2598 fecal samples collected from children aged 0 to 5 years old living in The Gambia as part of the GEMS study were available for this study (Kotloff et al., 2012). A panel of 634 stool specimens from St. Louis Childrens' hospital was collected from children age 0 to 18 years old, primarily with diarrheal diseases from July, 2009 through June, 2010 as previously described (Lim et al., 2013).

Unbiased pyrosequencing

From the index case, the stool sample was diluted in 6:1 in PBS and filtered through a 0.45 um membrane to minimize recovery of intact bacteria. Total nucleic acid was extracted from the filtrate, subjected to random-priming cDNA synthesis and amplification, and sequenced by FLX Titanium pyrosequencing as previously described (Holtz et al., 2009). High quality reads with no detectable similarity to the reference human genome or NCBI nt database by BLASTn were analyzed by BLASTx alignment against the NCBI non-redundant (nr) protein database (Zhao et al., 2013) in order to identify divergent viral sequences.

Amplification of complete genome

The complete genome of the rosavirus 2 GA7403 strain was amplified by RT-PCR in eight overlapping fragments, cloned and sequenced as previously described (Lim et al., 2013). The following primers were used: RV2-P1F (5′– GTAGCGATCATCCAGAGCTAGCGG–3′) with RV2-P1r (5′–CGC TGC TCA TTA GAT GAT GGC GAG– 3′); RV2-P2F (5′–GTCACTGATGCCATCGCC TCTG–3′) with RV2-P2r (5′–GCATCAATAGCTGCTGCCTGCAG–3′); RV2-P3F (5′–CTGCAGGCAGCAGCTATTGATGC–3′) with RV2-P3r (5′–

CAGGCGATGTGTGGTTGCAC–3′); RV2-P4F (5′–TCTGCTCCCCTGTCTCCCGC–3′) with RV2-P4r (5′–GTGGGTGCCTCATATGCTGC–3′); RV2-P5F (5′–GAAGCTCCTGAAGCCAGGTTC–3′) with RV2-P5r (5′–GTGGCGATCACCACACGAGAC–3′); RV2-P6F (5′–CATGGCTGACCTCGAGCAGAAG–3′) with RV2-P6r (5′–GTCATTGGAGATGAGTGGTGC–3′); RV2-P7F (5′–CTCTCCTGCAGTCATGGCTG–3′) with RV2-P7r (5′–AGCACACATGGTCTGGAACC–3′); RV2-P8F (5′–GTTGTGGAAGCTGGATCTCTGC– 3′) with RV2-P8r (5′–CGAGACACTAGAGCACCAGCG–3′). 5′ RACE was performed with RV2-5RACE1r (5′–GGGGAGATCCGCTAGCTCTGGATG–3′) and RV2-5RACE2r (5′–AGTGCGCCTACTACTCCACCCCTG–3′); 3′ RACE was performed with RV2-3RACE1F (5′– TCCACCGAG GGCCCAAG ACTTATGG–3′).

Diversity analyses and phylogenetic methods

Amino acid sequences of the full-length polyprotein from rosavirus 2, rosavirus M-7 (JF973687) and cadicivirus 244U (JN819204), 209 (JN819202) and 236 (JN819203) strains were aligned by MUSCLE (Edgar, 2004). Diversity plots were generated with Simplot (Lole et al., 1999), employing sliding windows of 200 amino acids in length and a step size of 20 amino acids, with Kimura (2-parameter) correction.

Phylogenetic trees were constructed from alignments of the concatenated 2C3CD and P1 (VP4231) regions from the following picornaviruses: enterovirus A (NC_001612), enterovirus B (NC_001472), enterovirus C (NC_001428), enterovirus D (NC_001430), enterovirus E (NC_001859), enterovirus F (DQ092770), enterovirus G (NC_004441), enterovirus H (NC_003988), enterovirus J (NC_010415), rhinovirus A (FJ445111), rhinovirus B (DQ473485), rhinovirus C (EF077280), simian sapelovirus (NC_004451), porcine sapelovirus (NC_003987), avian sapelovirus (NC_006553), bovine rhinitis A virus (JN936206), bovine rhinitis B virus (NC_010354), food-and-mouth disease virus (NC_004004), equine rhinitis A virus (NC_003982), equine rhinitis B virus (NC_003983), theilovirus (NC_001366), encephalomyocarditis virus (NC_001479), seneca valley virus (NC_011349), porcine teschovirus (NC_003985), porcine kobuvirus (EU787450), bovine kobuvirus (AB084788), aichi virus (NC_001918), salivirus (GQ253930), turdivirus 2 (NC_014412), turkey hepatitis virus 1 (HQ189775), rosavirus M-7 (JF973687), cadicivirus 209 (JN819202), cadicivirus 236 (JN819203), cadicivirus 244U (JN819204), human parechovirus (FM178558), Ljungan virus (EF202833), duck hepatitis A virus (NC_008250), hepatitis A virus (NC_001489), avian encephalomyelitis virus (NC_003990). The 3Dpol/RdRp region from the following representative members of the *Picornavirales* order were analyzed: infectious flacherie virus (AB000906), perina nuda virus (AF323747), varroa destructor virus-1 (AY251269), broad bean wilt virus 1 (AB084450), bean pod mottle virus (NC_003496), cowpea mosaic virus (NC_003549), aphid lethal paralysis virus (NC_004365), Kashmir bee virus (NC_004807), acute bee paralysis virus (NC_002548), cricket paralysis virus (NC_003924), drosophila C virus (NC_001834), chaetoceros tenuissimus RNA virus 01 (AB375474), heterosigma akashiwo RNA virus (AY337486), and Schizochytrium single-stranded RNA virus (NC_007522). Alignments were performed with a probabilistic, multiple sequence alignment algorithm Fast Statistical Alignment

(FSA) (Bradley et al., 2009). Phylogenies were constructed with PhyML v3.0 (Guindon and Gascuel, 2003) by the maximum likelihood (ML) method. Analyses were performed at least twice. Support for ML trees (LG+I+G+F) was assessed by 1,000 nonparametric bootstraps. Bayesian MCMC inference was performed with MrBayes v3.2.1 (Huelsenbeck and Ronquist, 2001). MrBayes analyses (RtREV+I+G+F) were run for 10,000,000 steps with a sample frequency set to 500 and a 25% burn-in period. Convergence and mixing were assessed with Tracer v1.5 and AWTY (Drummond and Andrew, 2009; Nylander et al., 2008). The two methods yielded trees with similar topologies.

For the phylogenetic analysis of four rosavirus 2 sequences obtained through the screening assay, nucleotide sequences were aligned by Muscle (Edgar, 2004) and primer sequences were trimmed from the alignment. A phylogeny was constructed by the neighbor-joining method using the Jukes-Cantor method of correction. Maximum likelihood method yielded a similar phylogeny.

RNA secondary structure analysis

RNA secondary structures from the 5′ and 3′ UTR structures were predicted using the Mfold web server version 2.3 (Zuker, 2003). 4612 picornavirus genomes (1903 complete genomes and 2709 > 4kb partial genomes) were obtained from the NIAID Virus Pathogen Database and Analysis Resource (ViPR) online through the web site at http://www.viprbrc.org (Pickett et al., 2012) and compiled into a custom database. Blast search was performed to query the ‘q motif’ against the custom picornavirus database.

Diagnostic RT-PCR amplification

Standard precautions to avoid end product contamination were taken for all PCR assays, including the use of PCR hoods and maintaining separate areas for PCR set up and analysis. For every 88 samples tested, seven no-template negative controls were interspersed between the actual samples. OneStep RT-PCR (Qiagen) was used to amplify 5 μl of extracted samples using the following PCR program: 50°C for 35 min, 95°C for 15 min, 40 cycles of 95°C for 30 sec, 55°C for 30 sec, 72°C for 21 sec, followed by 72°C for 10 min. The presence of rosavirus 2 was detected with the “forward primer” RV2ScreenF (5′–CAGAGYGATGAGCGYTTGTGTGCAG –3′) in combination with the “reverse primer” RV2Screenr (5′– GGTCASCACTGACCTGGGCAATGTC –3′) that together generated a 334 bp amplicon from the 3Dpol region. Products were visualized following electrophoresis on 1.25% agarose gels. Amplicons were cloned and sequence verified.

Accession numbers

The sequences of the complete genome of rosavirus 2 and amplicons from rosavirus 2 strains have been entered into the GenBank database under accession numbers: KJ158169 – KJ158172.

Results

Discovery of a novel picornavirus (rosavirus 2)

As part of a broader effort to identify novel viruses associated with childhood diarrhea in developing countries, we performed shotgun 454 pyrosequencing of total nucleic acid extracted from a fecal specimen from a child in The Gambia. We identified 2,199 out of 23,137 reads from the specimen that shared limited sequence identity to known picornaviruses. De novo assembly of the sequence reads generated an 8,713-nucleotide contig encoding a single predicted open reading frame (Figure 1A). To validate the viral genome, we designed primer pairs that generate 7 overlapping amplicons. Additionally, the 5′ and 3′ ends of the genome were defined by 5′- and 3′- rapid amplification of cDNA ends (RACE). Using RACE methods, we extended the initial contig by 186 nt in the 5′ end and 32 nt in the 3′ end that ended with a poly(A) tail. The resulting complete genome of 8,931 nt, excluding the 3′ poly(A) tail, was Sanger sequenced to more than 3× coverage.

Genome analysis of rosavirus 2

The genome encoded a single open reading frame of 7,404 nt with predicted typical picornavirus genomic organization and molecular features characteristic of picornaviruses (Figure 1B). The putative P1 region encoded a GXXXT/S myristoylation motif (G_3RKET). The putative 2C protease region had the GXXGXGKS NTP binding motif (G_{1729}GPGCGKS) and DDLXQ helicase activity motif (D_{1780}DLGQ). The GXCG cysteine active site is conserved in the putative 3C protease region (G_{2214}YCG). Finally, the putative 3D region maintains the $_{2591}$YGDD active site motif, and K_{2420}DELR, F_{2640}LKR, G_{2549}AMPSG motifs. Similar to cadicivirus and rosavirus, the (PS)ALXAXETG motif and RNA-binding domain KFRDI motif were absent from the putative VP1 and 3C protease region respectively.

Whole genome sequence analyses demonstrated that rosavirus 2 was most similar to the partial genome of rosavirus M-7 (*r*odent *s*tool *a*ssociated picorna*virus*), a picornavirus identified in a metagenomic survey of rodent stool specimens (Phan et al., 2011). While the manuscript was in preparation, the NCBI entry for the rodent rosavirus was updated with a near-complete genome sequence of rosavirus M-7 (Phan et al., 2013). Rosavirus M-7 had 71.9% nucleotide identity to rosavirus 2. Sequence analyses showed that next most similar virus sequences to rosavirus 2 were cadiciviruses (previously named as canine picodicistrovirus) (Woo et al., 2012). Cadicivirus shared 34.1, 24.2 and 39.4% nucleotide identity to rosavirus 2 in the P1, P2 and P3 region. The overall pairwise amino acid identity between rosavirus 2 and cadiciviruses was 40 – 43%. However there were regions of limited identity particularly in the VP2, VP1-2A junction and 3C regions (Figure 1C). The pairwise amino acid identity of rosavirus 2 compared to rosavirus in the P1, P2 and P3 region was 76.5, 80.7 and 83.8%, with an overall amino acid identity of 80.1%. Additionally, rosavirus 2 and rosavirus shared 85.6% amino acid identity in the 2C and 3CD regions. According to ICTV guidelines, picornavirus members of a species share >70% amino acid identities in the P1 and 70% amino acid identity in the 2C and 3CD regions (Fauquet CM, 2005; Knowles, 2012). By these criteria, rosavirus 2 belongs in the same species as rosavirus

Characterization of the 5′ and 3′ untranslated regions

Picornaviruses encode extensive RNA stem-and-loop structures in the 5′ and 3′ UTR regions that are critical for viral replication. The 829 nt long 5′ UTR of rosavirus 2 was predicted to form a type II IRES element (Figure 2A), similar to cardioviruses and apthoviruses (Martinez-Salas, 2008). The predicted central domain I had a typical cruciform structure including the conserved purine-rich GNRA and RAAA motifs, C-rich poly(rC) binding protein (PCBP) loop, and putative RNase P cleavage site (Rz P) (Correll and Swinger, 2003; Fernandez-Miragall et al., 2006; Serrano et al., 2007; Toyoda et al., 2007). Likewise, the two polypyrimidine tracts were predicted to be encoded in domain H and L. However, while the typical type II IRES such as EMCV and FMDV begin with a 5′ hairpin structure and poly(C) tract, the 5′ end of rosavirus 2 was predicted to form a cloverleaf stem-loop structure (Figure 2A, domain A – C). The predicted structure of rosavirus M-7 domains I – L (Phan et al., 2013) were similar to rosavirus 2, however we were unable to compare the 5′ cloverleaf stem-loop as the 5′ end of rosavirus M-7 genome was incomplete.

The 696 nt 3′ UTR of rosavirus 2 was shorter than the 795 nt 3′ UTR of rosavirus, and only shared a pairwise nucleotide identity of 59.9%. The 3′ UTR was predicted to form multiple stem-loop structures. The gallivirus/kobuvirus/avihepatovirus ‘barbell-like’ structure (Boros et al., 2012) was not found in rosavirus 2. In the process of searching for the ‘barbell-like’ structure, multiple sequence alignment between rosavirus 2, rosavirus M-7 and cadicivirus strains highlighted two stretches of highly conserved sequences flanking a variable region (31 – 62 nt) (Figure 2B). While the overall predicted structure of each 3′ UTR was different, the conserved sequences were predicted to fold into a similar stem-loop structure that we describe as a ‘q-shaped’ motif (Figure 2C). We queried the ‘q motif’ sequences by performing BLAST searches against sequences from 4612 picornavirus genomes (1903 complete genomes and 2709 > 4kb partial genomes) and only found the ‘q motif’ sequence in rosavirus and cadicivirus strains. This suggests that the ‘q motif’ might be a uniquely conserved feature of rosavirus 2, rosavirus M-7 and cadiciviruses.

Phylogenetic analysis indicates rosavirus 2 is a *bona fide* picornavirus—To investigate the evolutionary relationship between rosavirus 2 to other members of the *Picornaviridae* family, we performed maximum likelihood and Bayesian phylogenetic analyses using concatenated 2C3CD and P1 (VP4231) regions. Both methods yielded trees with similar topologies. Phylogenies of the 2C3CD and P1 regions strongly supported that rosavirus 2 was most closely-related to rosavirus, and they both shared a common ancestor with cadiciviruses. The phylogeny based on the 2C3CD region suggested that the rosavirus 2/rosavirus M-7/cadicivirus clade was most closely-related to the kobuvirus/salivirus/megrivirus clade (Figure 3A). However, in the P1 region, the rosavirus 2/rosavirus M-7/cadicivirus clade placed sister group to the enterovirus/cardiovirus/teschovirus clade (Figure 3B).

Cadiciviruses have been proposed as the transitional virus or ‘missing link’ between dicistroviruses and picornaviruses (Woo et al., 2012) as they encode a unique dicistronic opening reading frame, compared to the single open reading frame paradigm of other known picornaviruses. Since rosavirus 2 and the newly-available rosavirus M-7 sequences were

most closely-related to cadiciviruses (Figure 3), the proposed 'missing link' scenario would imply that rosavirus 2 and rosavirus M-7 were the most basal members of the *Picornaviridae* family. However, in order to accurately interpret the 'missing link' scenario between *Dicistroviridae* and *Picornaviridae*, we needed to determine the phylogenetic relationships of cadicivirus and rosavirus 2 in the broader context of the *Picornavirales* order. To investigate this, we performed phylogenetic analyses of the 3Dpol/RdRp region from representative members of the *Picornavirales* order *(Picornaviridae, Iflavivirade, Secoviridae, Dicistrovirdae, Bacillarnavirus, Marnaviridae* and *Labyrnavirus).* We found that *Picornaviridae* and *Dicistroviridae* were paraphyletic (Figure 4A), consistent with other studies (Le Gall et al., 2008; Sanfacon et al., 2009). Cadicivirus, rosavirus 2 and rosavirus M-7 formed a monophyletic clade within the diversity of the *Picornaviridae* family with strong support. Instead, there was strong statistical support for the common ancestor of parechovirus and hepatovirus being basal to other picornaviruses. To verify this, we next performed a phylogenetic analysis on the conserved 2C/S3H region. While there were differences in the branching order of virus families, we found strong support for the monophyly of cadicivirus, rosavirus 2 and rosavirus M-7 branching deeper within the *Picornaviridae* family (data not shown). Taken together, these results demonstrated that cadicivirus, rosavirus 2 and rosavirus M-7 are *bona fide* members of the *Picornaviridae* family, but they do not support a role for rosavirus 2 or cadiciviruses in the 'missing link' hypothesis between *Picornaviridae* and *Dicistroviridae*.

Prevalence of rosavirus 2 in The Gambia and USA—To investigate the prevalence of rosavirus 2, we developed an RT-PCR assay using primers designed from a 3Dpol region conserved between rosavirus 2 and rosavirus M-7. The assay amplified a specific 334 bp product (Figure 5A). Since rosavirus 2 was identified from a fecal specimen from a child in The Gambia, we examined whether rosavirus 2 might be associated with childhood diarrhea in The Gambia. We screened fecal samples collected from The Gambia as part of the Global Enteric Multi-center Study (GEMS) (Kotloff et al., 2012). For each enrolled case with moderate-to-severe diarrhea, one healthy control without diarrhea (matched for age, gender and time of presentation) was randomly selected from the community. We screened 332 cases and 390 controls for the presence of rosavirus 2, and found that the overall prevalence of rosavirus 2 in The Gambia was 0.55% (4 out of 772). Of the four positive specimens, one specimen was a moderate-to-severe diarrheal case and three specimens from the control group (Table 1). We RT-PCR amplified and sequenced a 2,428 bp region from the 3′ end of the genome (partial 3C, 3D and 3′ UTR) for 2 of the positive specimens (GA7242 and GA7400) and found they shared 99.7 % and 99.6% nucleotide identity to the index strain (Data not shown). There was no statistically significant association between rosavirus 2 with diarrheal cases (Figure 5B). We next screened 634 pediatric fecal specimens from patients primarily with diarrhea sent for bacterial culture to the clinical microbiology laboratory at the Saint Louis Children's Hospital. None of the specimens tested positive for rosavirus 2 (Figure 5C). We sequenced verified the amplicons and performed phylogenetic analyses using the neighbor-joining method. All four sequences formed a distinct monophyletic clade from rosavirus M-7 with high confidence, indicating that these were rosavirus 2 sequences (Figure 5D). These results indicate that rosavirus 2 can be detected at low prevalence in fecal specimens from children in The Gambia.

Discussion

A picornavirus (rosavirus 2) was identified from a fecal specimen from a child in The Gambia. The amino acid identity between rosavirus 2 and rosavirus M-7 was 76.5% and 85.6% in the P1 and 2C3CD regions respectively. Based on ICTV guidelines, rosavirus 2 would be classified as a same species with rosavirus M-7, a picornavirus isolated from rodent stool (Phan et al., 2011). In terms of nomenclature, rosavirus stands for *r*odent *s*tool *a*ssociated picorna*virus* (Phan et al., 2011); however in this instance we identified closely related sequences in stool samples from multiple human individuals (Figure 5), suggesting that members of this clade of viruses are not limited to association with rodents. Nonetheless, to be consistent with existing nomenclature, we propose to name this new variant as rosavirus 2.

The prevalence of rosavirus 2 was 0.55% (4 out of 772) in The Gambia. While we did not find evidence that rosavirus 2 was associated with diarrheal disease in the case-control cohort from The Gambia, the low overall prevalence in this study precludes a definitive conclusion about potential association with diarrhea. Furthermore, whether rosavirus 2 might cause other infectious diseases remains an open question. For example, poliovirus, a member of the *Picornaviridae*, causes neurological disease but has been well documented to be shed in stool (Hird and Grassly, 2012). In comparison to The Gambia cohort, we did not detect rosavirus 2 in a cohort of 634 pediatric specimens from Saint Louis, USA. The difference in prevalence between geographic regions was not statistically significant (Fisher's exact test $p > 0.13$). This finding contrasts the prevalence of human cosavirus, a picornavirus found at very high prevalence in South Asia but rarely detected in the United Kingdom (Kapoor et al., 2008). Taken together, the identification of rosavirus 2 underscores the importance of understanding unique factors that might influence public health of resource-poor countries.

The identification of novel viruses is also important for understanding the evolutionary trajectory of viruses (Koonin et al., 2008). Cadicivirus was previously proposed to be the ‘missing link’ between *Picornaviridae* and *Dicistroviridae* (Woo et al., 2012). While all picornaviruses encode a single open reading frame (IRES-P1-P2-P3) and dicistroviruses encode a dicistronic (IRES-P2-P3-IRES-P1) configuration, cadicivirus encodes a unique dicistronic (IRES-P1-IRES-P2-P3) picornavirus-like genome. Since rosavirus 2, and the newly-available rosavirus M-7 sequence, are most closely-related to cadicivirus, we revisited this hypothesis. The hypothesis suggests that rosavirus 2/rosavirus M-7 would be the most basal members of the *Picornaviridae* family or that their phylogenetic relationships might be difficult to resolve between *Picornaviridae* and *Dicistroviridae* families when examined in the broader context of the *Picornavirales* order. We found strong support that cadicivirus, rosavirus 2 and rosavirus M-7 formed a monophyletic clade within the *Picornaviridae* family (Figure 3 and 4), which does not support the ‘missing link’ hypothesis. This suggests that the dicistronic cadicivirus is not the ‘missing link’ between *Picornaviridae* and *Dicistroviridae* families. Instead, we propose a more parsimonious explanation that cadiciviruses acquired an additional IRES element between P1 and P2 regions through an independent event after it diverged from the rosavirus 2/rosavirus M-7

ancestor. Thus, the identification of novel viruses advances our understanding of the evolutionary history and relationships of broad viral lineages.

Acknowledgments

This work was supported in part by NIH grant U54 AI057160 to the Midwest Regional Center of Excellence for Biodefense and Emerging Infectious Disease Research, National Institutes of Health/National Center for Research Resources Washington University-ICTS (KL2 RR024994), Children's Discovery Institute (MD-FR-2013-292), and the Bill & Melinda Gates Foundation (OPP1016839). DW holds an investigator in the pathogenesis of infectious disease award from the Burroughs Wellcome Fund. ESL is an Eli & Edythe Broad Fellow of the Life Sciences Research Foundation.

References

Adams MJ, King AM, Carstens EB. Ratification vote on taxonomic proposals to the International Committee on Taxonomy of Viruses (2013). Arch Virol. 2013; 158:2023–2030. [PubMed: 23580178]

Ambert-Balay K, Lorrot M, Bon F, Giraudon H, Kaplon J, Wolfer M, Lebon P, Gendrel D, Pothier P. Prevalence and genetic diversity of Aichi virus strains in stool samples from community and hospitalized patients. J Clin Microbiol. 2008; 46:1252–1258. [PubMed: 18256215]

Boros A, Nemes C, Pankovics P, Kapusinszky B, Delwart E, Reuter G. Identification and complete genome characterization of a novel picornavirus in turkey (Meleagris gallopavo). J Gen Virol. 2012; 93:2171–2182. [PubMed: 22875254]

Bradley RK, Roberts A, Smoot M, Juvekar S, Do J, Dewey C, Holmes I, Pachter L. Fast statistical alignment. PLoS Comput Biol. 2009; 5:e1000392. [PubMed: 19478997]

Chikhi-Brachet R, Bon F, Toubiana L, Pothier P, Nicolas JC, Flahault A, Kohli E. Virus diversity in a winter epidemic of acute diarrhea in France. J Clin Microbiol. 2002; 40:4266–4272. [PubMed: 12409408]

Correll CC, Swinger K. Common and distinctive features of GNRA tetraloops based on a GUAA tetraloop structure at 1.4 A resolution. Rna. 2003; 9:355–363. [PubMed: 12592009]

Denno DM, Klein EJ, Young VB, Fox JG, Wang D, Tarr PI. Explaining unexplained diarrhea and associating risks and infections. Animal health research reviews / Conference of Research Workers in Animal Diseases. 2007; 8:69–80. [PubMed: 17692144]

Drummond AJ, Andrew R. Tracer v1.5. 2009

Edgar RC. MUSCLE: multiple sequence alignment with high accuracy and high throughput. Nucleic Acids Res. 2004; 32:1792–1797. [PubMed: 15034147]

Fauquet CM, MM.; Maniloff, J.; Desselberger, U.; Ball, LA. Virus taxonomy: Eighth report of the International Committee on Taxonomy of Viruses. San Diego, CA: Elsevier Academic Press; 2005.

Fernandez-Miragall O, Ramos R, Ramajo J, Martinez-Salas E. Evidence of reciprocal tertiary interactions between conserved motifs involved in organizing RNA structure essential for internal initiation of translation. Rna-a Publication of the Rna Society. 2006; 12:223–234.

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]

Herold J, Andino R. Poliovirus RNA replication requires genome circularization through a protein-protein bridge. Molecular cell. 2001; 7:581–591. [PubMed: 11463383]

Hird TR, Grassly NC. Systematic review of mucosal immunity induced by oral and inactivated poliovirus vaccines against virus shedding following oral poliovirus challenge. PLoS Pathog. 2012; 8:e1002599. [PubMed: 22532797]

Holtz LR, Finkbeiner SR, Zhao G, Kirkwood CD, Girones R, Pipas JM, Wang D. Klassevirus 1, a previously undescribed member of the family Picornaviridae, is globally widespread. Virology journal. 2009; 6:86. [PubMed: 19552824]

Huelsenbeck JP, Ronquist F. MRBAYES: Bayesian inference of phylogenetic trees. Bioinformatics. 2001; 17:754–755. [PubMed: 11524383]

Kapikian AZ. Viral gastroenteritis. JAMA: the journal of the American Medical Association. 1993; 269:627–630. [PubMed: 8380619]

Kapoor A, Victoria J, Simmonds P, Slikas E, Chieochansin T, Naeem A, Shaukat S, Sharif S, Alam MM, Angez M, Wang C, Shafer RW, Zaidi S, Delwart E. A highly prevalent and genetically diversified Picornaviridae genus in South Asian children. Proc Natl Acad Sci U S A. 2008; 105:20482–20487. [PubMed: 19033469]

Knowles NJ, Hovi T, Hyypiä T, King AMQ, Lindberg AM, Pallansch MA, Palmenberg AC, Simmonds P, Skern T, Stanway G, Yamashita T, Zell R. Virus Taxonomy: Classification and Nomenclature of Viruses: Ninth Report of the International Committee on Taxonomy of Viruses. Picornaviridae. 2012:855–880.

Koonin EV, Wolf YI, Nagasaki K, Dolja VV. The Big Bang of picorna-like virus evolution antedates the radiation of eukaryotic supergroups. Nature reviews Microbiology. 2008; 6:925–939.

Kotloff KL, Blackwelder WC, Nasrin D, Nataro JP, Farag TH, van Eijk A, Adegbola RA, Alonso PL, Breiman RF, Faruque AS, Saha D, Sow SO, Sur D, Zaidi AK, Biswas K, Panchalingam S, Clemens JD, Cohen D, Glass RI, Mintz ED, Sommerfelt H, Levine MM. The Global Enteric Multicenter Study (GEMS) of diarrheal disease in infants and young children in developing countries: epidemiologic and clinical methods of the case/control study. Clin Infect Dis. 2012; 55(Suppl 4):S232–245. [PubMed: 23169936]

Le Gall O, Christian P, Fauquet CM, King AM, Knowles NJ, Nakashima N, Stanway G, Gorbalenya AE. Picornavirales, a proposed order of positive-sense single-stranded RNA viruses with a pseudo-T = 3 virion architecture. Arch Virol. 2008; 153:715–727. [PubMed: 18293057]

Li LL, Victoria J, Kapoor A, Blinkova O, Wang CL, Babrzadeh F, Mason CJ, Pandey P, Triki H, Bahri O, Oderinde BS, Baba MM, Bukbuk DN, Besser JM, Bartkus JM, Delwart EL. A Novel Picornavirus Associated with Gastroenteritis. Journal of Virology. 2009; 83:12002–12006. [PubMed: 19759142]

Lim ES, Reyes A, Antonio M, Saha D, Ikumapayi UN, Adeyemi M, Stine OC, Skelton R, Brennan DC, Mkakosya RS, Manary MJ, Gordon JI, Wang D. Discovery of STL polyomavirus, a polyomavirus of ancestral recombinant origin that encodes a unique T antigen by alternative splicing. Virology. 2013; 436:295–303. [PubMed: 23276405]

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]

Martinez-Salas E. The impact of RNA structure on picornavirus IRES activity. Trends Microbiol. 2008; 16:230–237. [PubMed: 18420413]

Nylander JA, Wilgenbusch JC, Warren DL, Swofford DL. AWTY (are we there yet?): a system for graphical exploration of MCMC convergence in Bayesian phylogenetics. Bioinformatics. 2008; 24:581–583. [PubMed: 17766271]

Pham NT, Khamrin P, Nguyen TA, Kanti DS, Phan TG, Okitsu S, Ushijima H. Isolation and molecular characterization of Aichi viruses from fecal specimens collected in Japan, Bangladesh, Thailand, and Vietnam. J Clin Microbiol. 2007; 45:2287–2288. [PubMed: 17522267]

Phan TG, Kapusinszky B, Wang C, Rose RK, Lipton HL, Delwart EL. The fecal viral flora of wild rodents. PLoS Pathog. 2011; 7:e1002218. [PubMed: 21909269]

Phan TG, Vo NP, Simmonds P, Samayoa E, Naccache S, Chiu CY, Delwart E. Rosavirus: the prototype of a proposed new genus of the Picornaviridae family. Virus genes. 2013

Pickett BE, Sadat EL, Zhang Y, Noronha JM, Squires RB, Hunt V, Liu M, Kumar S, Zaremba S, Gu Z, Zhou L, Larson CN, Dietrich J, Klem EB, Scheuermann RH. ViPR: an open bioinformatics database and analysis resource for virology research. Nucleic Acids Res. 2012; 40:D593–598. [PubMed: 22006842]

Ren L, Gonzalez R, Xiao Y, Xu X, Chen L, Vernet G, Paranhos-Baccala G, Jin Q, Wang J. Saffold cardiovirus in children with acute gastroenteritis, Beijing, China. Emerg Infect Dis. 2009; 15:1509–1511. [PubMed: 19788828]

Sanfacon H, Wellink J, Le Gall O, Karasev A, van der Vlugt R, Wetzel T. Secoviridae: a proposed family of plant viruses within the order Picornavirales that combines the families Sequiviridae and

Comoviridae, the unassigned genera Cheravirus and Sadwavirus, and the proposed genus Torradovirus. Arch Virol. 2009; 154:899–907. [PubMed: 19350366]

Serrano P, Gomez J, Martinez-Salas E. Characterization of a cyanobacterial RNase P ribozyme recognition motif in the IRES of foot-and-mouth disease virus reveals a unique structural element. Rna. 2007; 13:849–859. [PubMed: 17449727]

Shan T, Wang C, Cui L, Yu Y, Delwart E, Zhao W, Zhu C, Lan D, Dai X, Hua X. Picornavirus salivirus/klassevirus in children with diarrhea, China. Emerg Infect Dis. 2010; 16:1303–1305. [PubMed: 20678331]

Toyoda H, Franco D, Fujita K, Paul AV, Wimmer E. Replication of poliovirus requires binding of the poly(rC) binding protein to the cloverleaf as well as to the adjacent C-rich spacer sequence between the cloverleaf and the internal ribosomal entry site. J Virol. 2007; 81:10017–10028. [PubMed: 17609276]

Walker CL, Aryee MJ, Boschi-Pinto C, Black RE. Estimating diarrhea mortality among young children in low and middle income countries. PLoS One. 2012; 7:e29151. [PubMed: 22235266]

Woo PC, Lau SK, Choi GK, Huang Y, Teng JL, Tsoi HW, Tse H, Yeung ML, Chan KH, Jin DY, Yuen KY. Natural occurrence and characterization of two internal ribosome entry site elements in a novel virus, canine picodicistrovirus, in the picornavirus-like superfamily. J Virol. 2012; 86:2797–2808. [PubMed: 22205729]

Zhao G, Krishnamurthy S, Cai Z, Popov VL, Travassos da Rosa AP, Guzman H, Cao S, Virgin HW, Tesh RB, Wang D. Identification of Novel Viruses Using VirusHunter -- an Automated Data Analysis Pipeline. PLoS One. 2013; 8:e78470. [PubMed: 24167629]

Zuker M. Mfold web server for nucleic acid folding and hybridization prediction. Nucleic Acids Res. 2003; 31:3406–3415. [PubMed: 12824337]

Highlights

- Identification of a new picornavirus, Rosavirus 2, from a child in The Gambia
- We detected rosavirus 2 in multiple pediatric fecal specimens from The Gambia
- We found a ‘q-motif’ RNA stem-loop structure in the 3′ UTR uniquely conserved in rosavirus and cadiciviruses

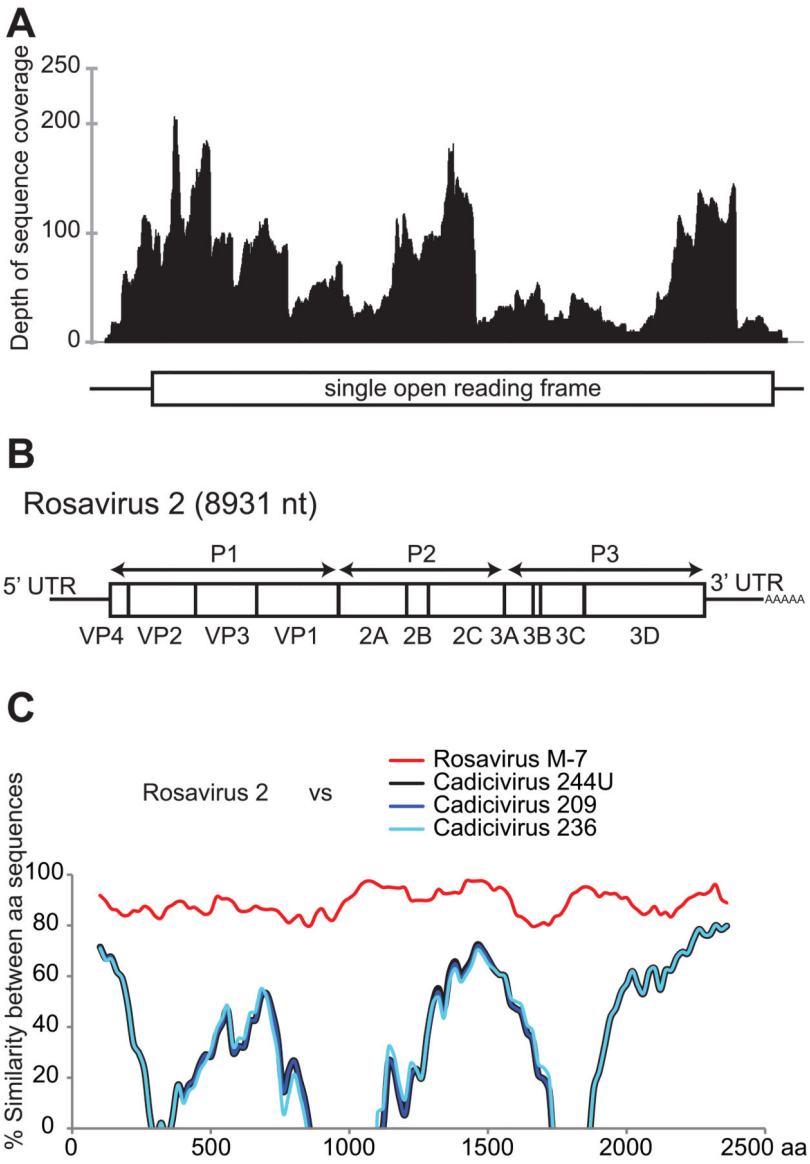


Figure 1. Identification of a novel picornavirus

(A) Coverage map of 454 pyrosequencing reads mapping to the initial assembled 8713 nt contig. (B) Schematic shows the complete rosavirus 2 genome. (C) Diversity plots of amino acid sequences are shown comparing rosavirus 2 to rosavirus M-7 (red) and cadicivirus strains (blue and black).

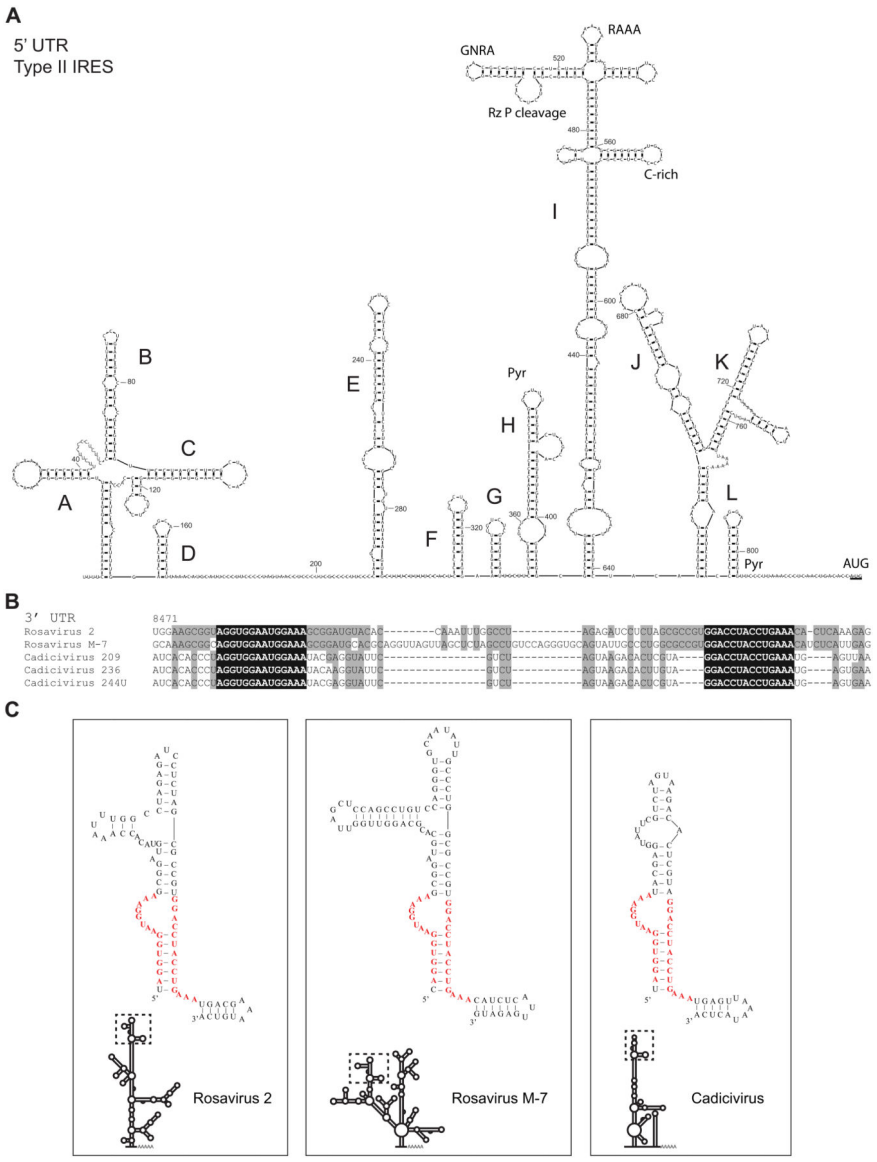


Figure 2. 5′ UTR of rosavirus 2 predicted to form a type II IRES

(A) Diagram shows the predicted RNA structure of the 5′ UTR of rosavirus 2. Conserved motifs are indicated: purine-rich GNRA and RAAA motifs, cysteine-rich poly(rC) binding protein loop (C-rich), ribozyme P cleavage site (Rz P cleavage), and polypyrimidine tracts (Pyr). IRES domains are labeled (A – L) and the start codon is underlined (AUG). (B) Nucleotide alignment of the 3′ UTR of rosavirus 2, rosavirus M-7 and 3 cadicivirus strains is shown, starting at nucleotide position 8471 in reference to the rosavirus 2 genome. The ‘q motif’ nucleotide sequences conserved across the viruses are highlighted in black, nucleotides identical to the rosavirus 2 are highlighted in gray. (C) Panels show the predicted 3′ UTR structure of rosavirus 2 (left), rosavirus M-7 (middle) and cadicivirus (right). The ‘q motif’ is bolded in red, and outlined within a dashed box in reference to the predicted 3′ UTR structure.

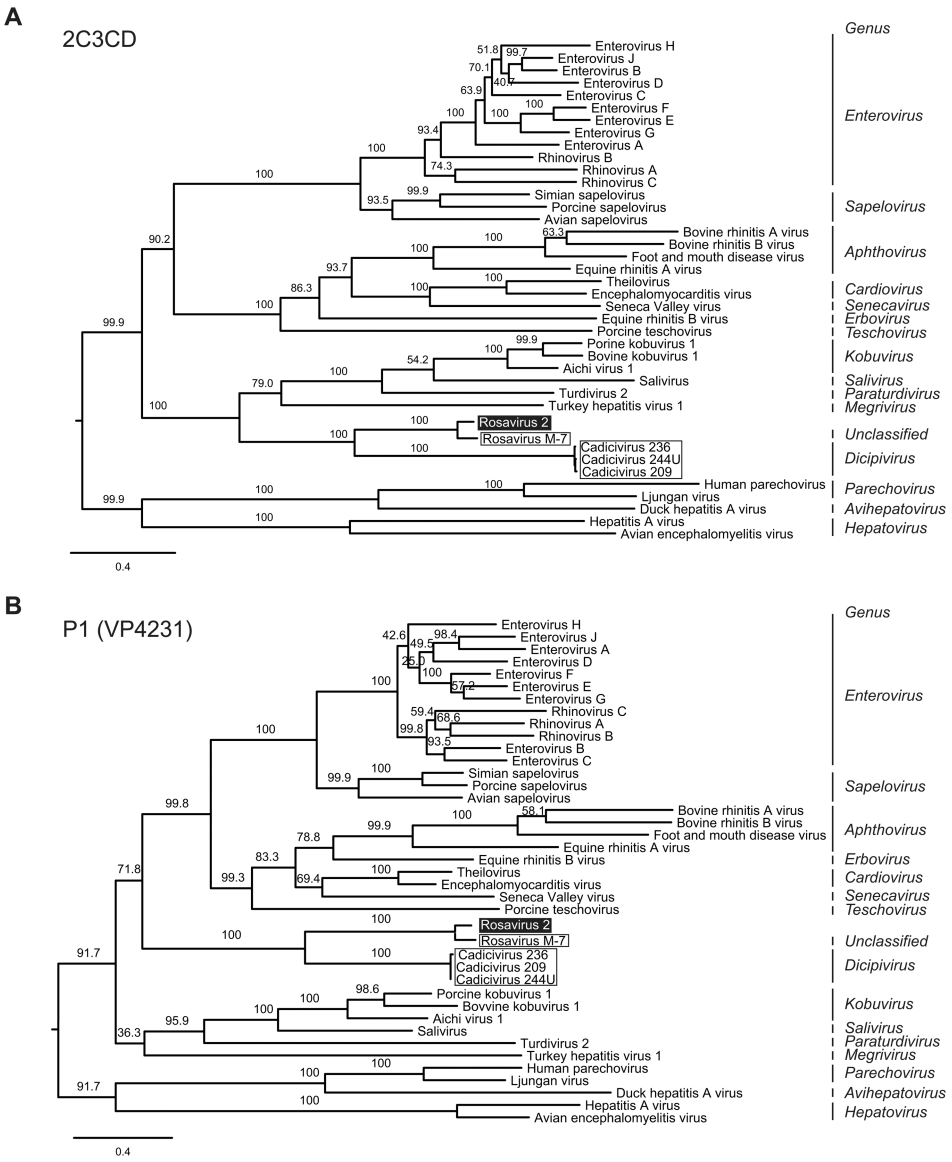


Figure 3. Rosavirus 2 is most closely-related to rosavirus M-7

(A) Phylogenetic relationships of representative members of the *Picornaviridae* family were inferred from the concatenated 2C3CD amino acid alignment, generated by the maximum likelihood method. The rosavirus 2 is highlighted in black, rosavirus M-7 and cadicivirus strains are outlined. (B) Phylogenetic tree was generated from the amino acid alignment of the P1 (VP4231) region. Internal branch labels indicate the bootstrap values. The Bayesian inference method yielded trees with similar topologies.

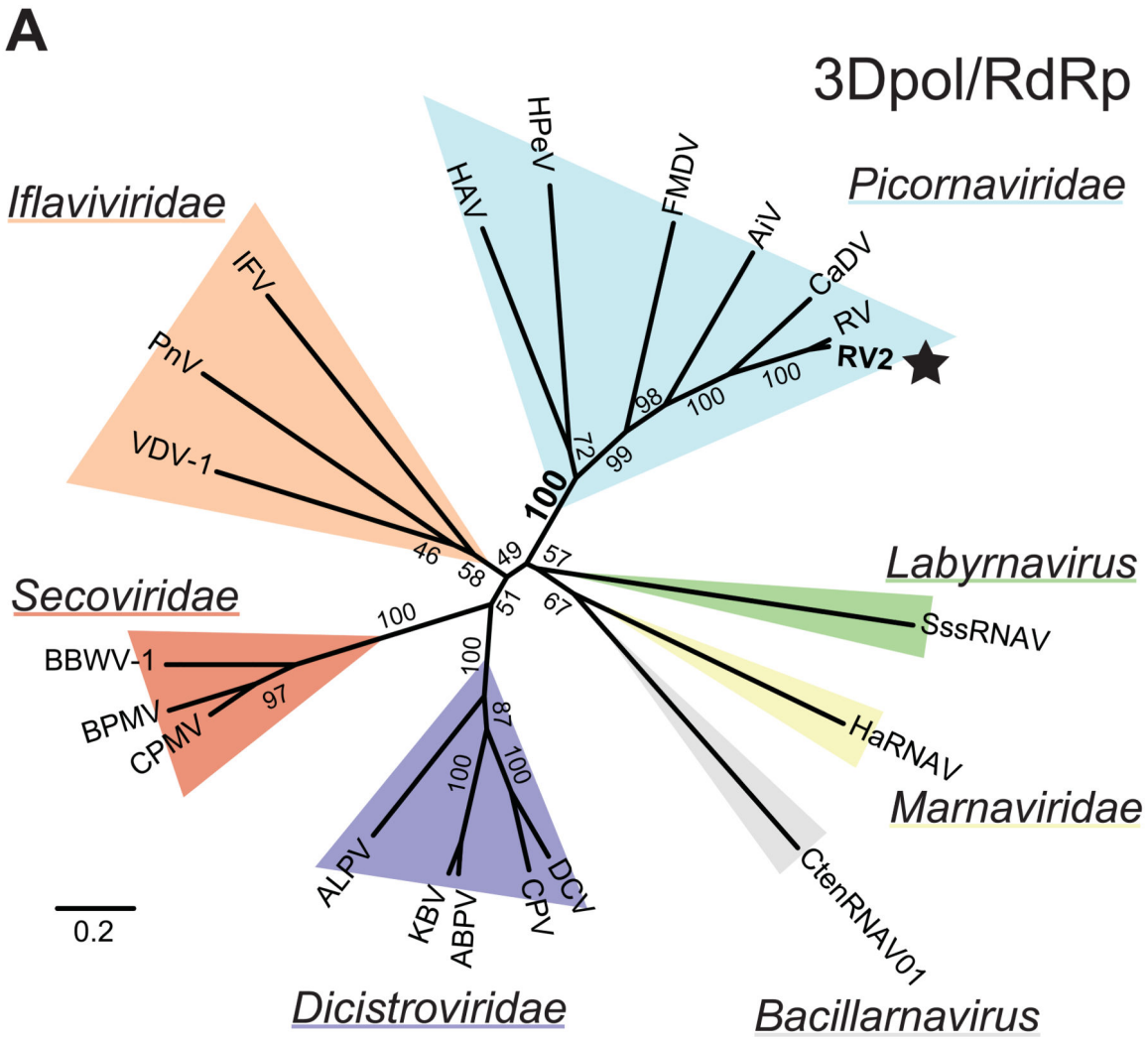
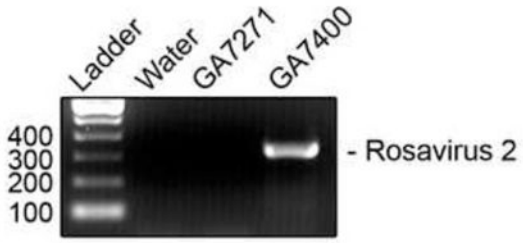


Figure 4. Cadicivirus, rosavirus 2 and rosavirus M-7 form a monophyletic clade in the *Picornaviridae* family

(A) Unrooted phylogenetic tree of 3Dpol/RdRp sequences from representative members of the *Picornavirales* order are shown. The major families are shaded in color as indicated. Representative members of each family were used: hepatitis A virus (HAV), human parechovirus (HPeV), foot-and-mouth disease virus (FMDV), aichi virus (AiV), cadicivirus (CaDV), rosavirus M-7 (RV), rosavirus 2 (RV2), infectious flacherie virus (IFV), perina nuda virus (PnV), varroa destructor virus-1 (VDV-1), broad bean wilt virus 1 (BBWV-1), bean pod mottle virus (BPMV), cowpea mosaic virus (CPMV), aphid lethal paralysis virus (ALPV), Kashmir bee virus (KBV), acute bee paralysis virus (ABPV), cricket paralysis virus (CPV), drosophila C virus (DCV), chaetoceros tenuissimus RNA virus 01 (CtenRNAV01), heterosigma akashiwo RNA virus (HaRNAV), and Schizochytrium single-stranded RNA virus (SssRNAV). Internal branch labels indicate the bootstrap values. Rosavirus 2 is indicated with a star.

A



B

Gambia Diarrheal Case-Control Study

	Case	Control	
Rosavirus 2 present	1	3	OR = 0 (NS)
Rosavirus 2 absent	331	387	Fisher exact test p = 1.0 (NS)

C

Saint Louis Children's Study

	Rosavirus 2
Number of positive samples	0 / 634

D

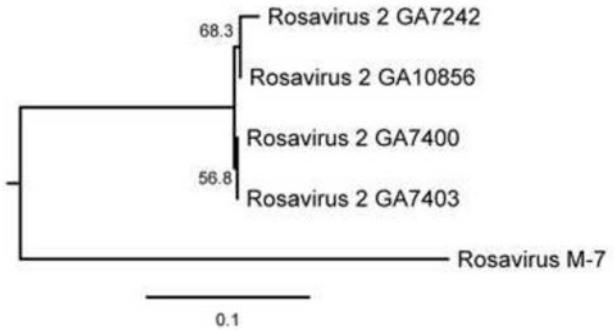


Figure 5. Prevalance of rosavirus 2

(A) RT-PCR analysis of rosavirus 2 is shown for water (control), or representative specimens found to be negative (GA7271), and positive (GA7400) for rosavirus 2. Band corresponds to a 334 bp PCR product. (B) Prevalance of rosavirus 2 in 722 fecal specimens from the Gambian diarrheal case-control study is shown. Odds ratio (OR), 95% confidence interval (CI) and Fisher exact test indicate that there is no statistically significant evidence for rosavirus 2 association with diarrheal cases (NS = not significant). (C) Results of RT-PCR screening for rosavirus 2 in a panel of 634 pediatric specimens from the Saint Louis Children's study are shown. (D) Neighbor-joining phylogeny inferred from the nucleotide sequences of the four rosavirus 2 strains screened positive from (B) and rosavirus M-7 strain are shown. Internal branch labels indicate the bootstrap values.

Table 1

Pediatric fecal specimens (aged 0 to 5 years) from The Gambia diarrheal case-control cohort detected positive for rosavirus 2.

ID	Status	Date of collection	Gender
GA7242	Control	July 16 2008	Male
GA7400	Diarrheal case	October 27 2008	Male
GA7403	Control	October 29 2008	Female
GA10856	Control	February 4 2009	Male