Virology 454-455 (2014) 25–33

Contents lists available at ScienceDirect

Virology

journal homepage: www.elsevier.com/locate/yviro

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

Efrem S. Lim [a], Song Cao [a], Lori R. Holtz [b], Martin Antonio [c], O. Colin Stine [d], David Wang [a,*]

[a] Departments of Molecular Microbiology and Pathology & Immunology, Washington University School of Medicine, St. Louis, MO, USA
[b] Department of Pediatrics, Washington University School of Medicine, St. Louis, MO, USA
[c] Medical Research Council Unit, Banjul, The Gambia
[d] Department of Epidemiology and Public Health, University of Maryland School of Medicine, Baltimore, MD, USA

ARTICLE INFO

Article history:
Received 20 December 2013
Returned to author for revisions
16 January 2014
Accepted 22 January 2014
Available online 20 February 2014

Keywords:
Picornavirus
Virus discovery
Emerging viruses
Developing country

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 the 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.

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 etiology can be linked to rotaviruses, caliciviruses, 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 et al., 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

* Correspondence to: Washington University School of Medicine, Campus Box 8230, 660 S. Euclid Ave., St. Louis, MO 63110, USA.
E-mail address: davewang@borcim.wustl.edu (D. Wang).

0042-6822/$ - see front matter © 2014 Elsevier Inc. All rights reserved.
http://dx.doi.org/10.1016/j.virol.2014.01.018

(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.

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 2199 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 8713-nucleotide contig encoding a single predicted open reading frame (Fig. 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 8931 nt, excluding the 3′ poly (A) tail, was Sanger sequenced to more than 3X coverage.

Genome analysis of rosavirus 2

The genome encoded a single open reading frame of 7404 nt with predicted typical picornavirus genomic organization and molecular features characteristic of picornaviruses (Fig. 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 (rodent stool associated 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 (Fig. 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 et al., 2005; Knowles et al., 2012). By these criteria, rosavirus 2 belongs in the same species as rosavirus.

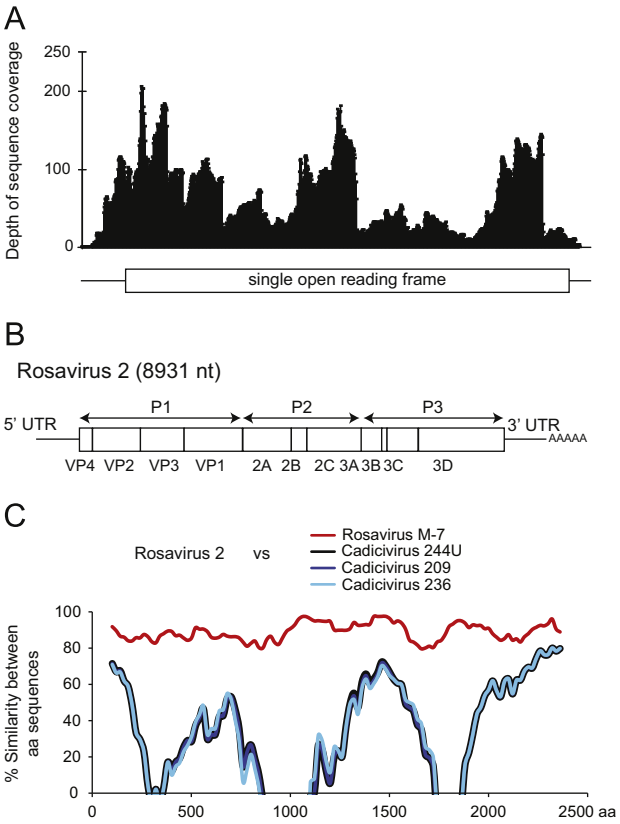

Fig. 1. Identification of a novel picornavirus. (A) Coverage map of 454 pyrosequencing reads mapping to the initial assembled 8713 nt contig. (B) Schematic shows the complete rosavirus 2 genome. (C) Diversity plots of amino acid sequences are shown comparing rosavirus 2 to rosavirus M-7 (red) and cadicivirus strains (blue and black).

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 (Fig. 2A), similar to cardioviruses and apthoviruses (Martinez-Salas, 2008). The predicted central domain I had a typical cruciform structure including the conserved purine-rich