

Published in final edited form as:

*Virology*. 2014 November ; 0: 556–564. doi:10.1016/j.virol.2014.09.012.

## Geographic variation in the eukaryotic virome of human diarrhea

Lori R. Holtz<sup>a</sup>, Song Cao<sup>b</sup>, Guoyan Zhao<sup>b</sup>, Irma K. Bauer<sup>c</sup>, Donna M. Denno<sup>d,e</sup>, Eileen J. Klein<sup>d</sup>, Martin Antonio<sup>f</sup>, O. Colin Stine<sup>g</sup>, Thomas L. Snelling<sup>h,i</sup>, Carl D. Kirkwood<sup>j</sup>, and David Wang<sup>b,c,#</sup>

Lori R. Holtz: Holtz\_L@kids.wustl.edu; Song Cao: scao@path.wustl.edu; Guoyan Zhao: GZhao@pathology.wustl.edu; Irma K. Bauer: Bauer\_I@kids.wustl.edu; Donna M. Denno: ddenno@u.washington.edu; Eileen J. Klein: eileen.klein@seattlechildrens.org; Martin Antonio: mantonio@mrc.gm; O. Colin Stine: ostin001@umaryland.edu; Thomas L. Snelling: tsnelling@ichr.uwa.edu.au; Carl D. Kirkwood: carl.kirkwood@mcri.edu.au

<sup>a</sup>Washington University School of Medicine, Department of Pediatrics, St. Louis, MO, USA

<sup>b</sup>Washington University School of Medicine, Department of Pathology and Immunology, St. Louis, MO, USA

<sup>c</sup>Washington University School of Medicine, Department of Molecular Microbiology, St. Louis, MO, USA

<sup>d</sup>University of Washington, Department of Pediatrics, Seattle, WA, USA

<sup>e</sup>University of Washington, Department of Global Health, Seattle, WA, USA

<sup>f</sup>Medical Research Council Unit, The Gambia

<sup>g</sup>Department of Epidemiology and Public Health, University of Maryland School of Medicine, Baltimore, MD, USA

<sup>h</sup>Menzies School of Health Research, Charles Darwin University, Northern Territory, Australia

<sup>i</sup>Telethon Institute for Child Health Research, Centre for Child Health Research, University of Western Australia

<sup>j</sup>Enteric Virus Research Group, Murdoch Childrens Research Institute, Royal Children's Hospital, Victoria, Australia

### Abstract

Little is known about the population of eukaryotic viruses in the human gut (“virome”) or the potential role it may play in disease. We used a metagenomic approach to define and compare the eukaryotic viromes in pediatric diarrhea cohorts from two locations (Melbourne and Northern

© 2014 Elsevier Inc. All rights reserved.

Corresponding author information: David Wang, PhD, Campus Box 8230, Department of Molecular Microbiology, 660 S. Euclid Ave., St. Louis, MO 63110, Tel: 314-286-1123 Fax: 314-362-1232, davewang@wustl.edu.

**Reviewer link to data:** [http://vchip.wustl.edu/download/totreads\\_NT\\_Melbourne.zip](http://vchip.wustl.edu/download/totreads_NT_Melbourne.zip)

Availability of Supporting Data: Sequences (human reads removed) have been uploaded to the MG-RAST <http://metagenomics.anl.gov/> server under the project name Australian Pediatric Diarrhea [MG-Rast ID 4529846.3- 4529932.3].

**Competing Interests:** The authors have no competing interests.

**Publisher's Disclaimer:** This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final citable form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

Territory, Australia). We detected viruses known to cause diarrhea, non-pathogenic enteric viruses, viruses not associated with an enteric reservoir, viruses of plants, and novel viruses. Viromes from Northern Territory children contained more viral families per sample than viromes from Melbourne, which could be attributed largely to an increased number of sequences from the families *Adenoviridae* and *Picornaviridae* (genus enterovirus). qRT-PCR/PCR confirmed the increased prevalence of adenoviruses and enteroviruses. Testing of additional diarrhea cohorts by qRT-PCR/PCR demonstrated statistically different prevalences in different geographic sites. These findings raise the question of whether the virome plays a role in enteric diseases and conditions that vary with geography.

## Keywords

Diarrhea; virome; metagenomics

## Background

It is well established that disease prevalence varies in different geographical regions of the world. Improvements in hygiene and decreased microbial exposure in childhood have been hypothesized to be responsible for the increased occurrence of allergies, autoimmune disorders, and inflammatory bowel disease in the westernized world (Strachan, 1989). On the other hand, conditions such as environmental enteropathy and decreased oral vaccine efficacy are seen in the developing world. Environmental enteropathy is a diffuse villous atrophy of the small bowel, which is ubiquitous in children in the developing world (Campbell et al., 2003; Menzies et al., 1999). It has been observed that environmental enteropathy reverses on transfer to an environment with improved hygiene and sanitation (Lindenbaum, 1968; Lindenbaum et al., 1971). Clearly, the environment is an important factor in the development of human disease.

The human gut contains a diverse microbial community, and it is postulated that many disorders of digestion and growth including diarrhea, inflammatory bowel disease, environmental enteropathy, and malnutrition could be related to perturbations in this biomass. Significant effort has been made to understand the bacterial component of the human stool microbiome and the role it may play in human disease. For example, a dysbiosis or shift in the relative abundances of the bacterial taxa has been associated with obesity (Turnbaugh et al., 2009), inflammatory bowel disease (Frank et al., 2007; Sartor, 2008), diabetes (Larsen et al., 2010), and necrotizing enterocolitis (Mai et al., 2011). The bacterial microbiome can be influenced by several factors including diet, geography, and phage populations. Studies comparing the bacterial gut communities from healthy populations in different parts of the world show that the bacterial microbiome varies with geography (De Filippo et al., 2010; Yatsunenko et al., 2012).

A few studies have examined the human bacteriophage virome in stool. All of the studies reported to date have been done in healthy humans. These findings have shown that the phage virome changes rapidly in the first week of life (Breitbart et al., 2008), interpersonal variation is high while intrapersonal diversity is low (Reyes et al., 2010), diet plays an

important role (Minot et al., 2011), and that the phage virome evolves quickly (Minot et al., 2013).

By contrast, little is known about the enteric eukaryotic virome or the potential role it, and variation within it, may play in human disease. In monkeys, pathogenic infection with simian immunodeficiency virus is associated with expansion of the enteric virome, including both RNA viruses and DNA viruses (Handley et al., 2012). A limited number of studies have begun to define the eukaryotic stool viromes of people with diarrhea (Finkbeiner et al., 2008; Nakamura et al., 2009; Phan et al., 2012; Smits et al., 2014; van Leeuwen et al., 2010), non-polio acute flaccid paralysis (Victoria et al., 2009), and healthy adults (Zhang et al., 2006). However, a limitation of the previous human studies is that none have explicitly compared eukaryotic viromes between geographic sites, disease states or age groups, and thus it is not known what factors might influence the composition of the human virome.

Acute diarrhea is one of the leading causes of mortality worldwide, and viruses are known to play a major etiologic role. As a significant fraction of cases are of unexplained etiology, many metagenomic studies have focused upon identifying novel candidate microbial agents. Because previous metagenomic studies demonstrated significant virus diversity in patients with diarrhea (Finkbeiner et al., 2008) we focused this study on comparing the eukaryotic virus populations in stools of children with diarrhea collected from two different locations, Melbourne, Australia and the Northern Territory, Australia.

## Results

### Sequencing statistics and cohort demographics

Metagenomic sequencing was performed on 43 stool samples from Melbourne and 44 stool samples from Northern Territory. The Melbourne and Northern Territory cohorts were 53% and 50% female, respectively ( $p$ =non-significant). The average age of the Melbourne cohort was 25.4 months, while the average age of the Northern Territory cohort was 15.7 months ( $p$  = 0.002 (t-test)). 454 sequencing of the Melbourne and Northern Territory samples yielded 646,630 and 406,252 total reads and 76,347 and 118,420 unique reads respectively. The average sequence length of the Melbourne and the Northern Territory samples was 339 and 329 respectively. 83.6% percent of the unique reads possessed detectable similarity to sequences in GenBank while the remaining 16.4% of reads had no significant similarity to sequences in GenBank. Of the 162,784 classifiable sequences 2.5% were viral, 75.8% bacteria, 4.4% phage, 8.9 % fungi, 7.0% human, and 1.4% other (plant, fish, etc). Further breakdown of these statistics by cohort is shown in Table 1.

### Viruses detected by metagenomic analysis

Because all of the stool samples sequenced were from children with acute diarrhea, it was anticipated that a known diarrhea virus would be detected in many of the samples. Twenty-four (28%) of the 87 samples contained sequences derived from the families *Reoviridae*, *Caliciviridae*, *Adenoviridae* or *Astroviridae*. Strikingly, sequences from many additional virus families were detected in these samples as well as in samples that did not contain sequences from the four canonical diarrheagenic virus families. The most commonly

detected virus family was the *Anelloviridae*, which was present in 37 samples. In total, viral sequences were identified in 57 (66%) of the 87 Australian samples while 35/87 (40%) of the samples had sequences from 2 or more viral families (supplemental table 1). Thirty-five of the 44 (80%) samples from the Northern Territory had one or more viral families detected which was more than the 22 of the 43 (51%) samples from Melbourne that had virus detected. One sample from Northern Territory contained sequences from 8 different viral families (Figure 1a). Overall, sequences from 22 different viral families were detected in these 87 samples. These included viruses from families known to reside in the gastrointestinal tract such as *Picornaviridae* (Pallansch, 2007), *Anelloviridae* (Okamoto et al., 1998), *Circoviridae* (Li et al., 2010), and *Orthomyxoviridae* (Wootton et al., 2006) and known plant viruses, including members of the families *Betaflexiviridae*, *Bromoviridae*, *Endornaviridae*, and *Virgaviridae*, consistent with previous studies that reported detection of plant viruses (Finkbeiner et al., 2008; Victoria et al., 2009; Zhang et al., 2006). In addition, viruses not previously thought to have an enteric reservoir were detected. For example, human parainfluenza 3 was detected in one sample. The most common viral families detected were *Anelloviridae* (37 samples), *Picornaviridae* (26 samples) and *Adenoviridae* (10 samples). As with all sequencing based studies, our findings are limited by the depth of sequencing achieved. It is possible that deeper sequencing may detect additional viruses present at low abundance.

A number of samples contained what are likely to be novel viruses that shared only very limited sequence identity with known viruses. For example, one sample contained 110 unique sequences with limited similarity to viruses in the order of *Picornavirales*. Assembly of these reads and sequences that did not share detectable sequence similarity with anything in the database yielded 9 contigs that shared highest sequence similarity with viruses in the order *Picornavirales*. The longest contig of 9898nt (601 reads, 18× coverage) shared only 28% amino acid identity to the Israel acute paralysis virus of bees and is likely to be almost the complete genome. A second contig of 7748 nt (322 reads, 12× coverage) shared 28% amino acid identity to the Kashmir bee virus. The two novel viruses shared 82% nt identity with each other. These contigs have been deposited in GenBank [accession numbers KJ420969 - KJ420970]. In other samples, sequences were also identified from divergent DNA and RNA viruses, including small DNA viruses from the family *Anelloviridae* and RNA virus families including *Endornaviridae*, *Betaflexiviridae*, *Partitiviridae*, and *Virgaviridae*. It is unknown if these newly described viruses are capable of infecting humans or if they are dietary passengers that infect food ingested by the individual.

### Comparison of viromes

For each stool specimen, it would be ideal to define the number of distinct virus species present and the relative abundance of each species. However, a number of technical challenges and limitations precluded this form of analysis. First, highly variable coverage levels of the viruses ranging from a single read to hundreds of reads of a virus were obtained, presumably due to variations in viral load as well as the relative abundance of the virus compared to other non-viral nucleic acid molecules in the sample. For samples with low coverage levels, it is impossible to determine whether multiple reads that align to different regions of the same reference virus are derived from the same or related but distinct

viruses; therefore, it is challenging to accurately define the number of virus species present. Furthermore estimation of the viral load or copy number from the number of reads obtained is complicated by the sequence independent PCR step, which may introduce amplification biases.

Due to these considerations, we conservatively chose to define viral diversity based on the presence or absence of sequences derived from each of the 95 virus families defined by the ICTV. Therefore, the virome of an individual sample was defined as the number and type of viral taxa with at least one representative sequence present in the sample. These conservative criteria will likely underestimate the true viral diversity of the sample. Similar analyses were also performed by quantifying the presence or absence of sequences from each virus genus.

### Virome comparison by gender

To assess whether gender influences the virome, we compared the number of viral families detected per sample by gender in the Northern Territory and Melbourne cohorts and as an aggregate cohort. There was no statistically significant difference (Wilcoxon) in the number of viral families per sample by gender ( $p=0.912$  (regardless of location),  $p=0.444$  (Melbourne),  $p=0.952$  (Northern Territory)). We then assessed if there were gender differences at the individual virus family level. No individual virus family had a statistically significant difference.

### Virome comparison by geographic location

To determine if the virome differs by geographic location we compared the number of viral families and genera per sample in each location (Figure 1). Diarrhea samples from the Northern Territory contained more viral families per sample than diarrhea from Melbourne ( $p=0.0002$  (Wilcoxon)). This was also true at the genus level ( $p=0.0003$  (Wilcoxon)). To control for differences for the larger number of unique reads generated in the Northern Territory samples compared to the Melbourne samples (Table 1), we randomly selected unique reads from the Northern Territory samples to achieve the same sampling depth of the Melbourne samples 100 times. All 100 iterations showed that samples from the Northern Territory contained more viral families (and viral genera) per sample than the samples from Melbourne ( $p<0.05$  (Wilcoxon)). This result demonstrates that the significant difference between Northern Territory and Melbourne is not due to the higher number of unique reads obtained for Northern Territory samples. To determine if a particular viral family drove this difference we compared the prevalence of each individual virus family in the two cohorts. The highly abundant anelloviruses were detected with similar frequencies in the two cohorts (Northern Territory, 52.3%, Melbourne, 32.6%;  $p=0.083$  (Fisher's exact)). Two families, the *Picornaviridae* ( $p=0.0003$  (Fisher's exact)) and the *Adenoviridae* ( $p=0.015$  (Fisher's exact)) were statistically different between the two locations. As there are many picornaviruses known to infect humans, we then determined which picornavirus genera were present. Sequences from the following genera in the family *Picornaviridae* were detected: *Cardiovirus*, *Enterovirus*, *Klassevirus*, *Kobuvirus*, and *Parechovirus* (Table 2). The genus *Enterovirus* was the most frequently detected and was the only one that was statistically different between the two locations ( $p<0.001$  (Fisher's exact)).

As the average age of the two cohorts was statistically significantly different, (25.4 months (Melbourne) vs. 15.7 months (Northern Territory)), we evaluated if the difference in age was confounding the difference seen by location. We assessed if there was a correlation between age and the total number of viral families present. No correlation was detected in either Northern Territory (Spearman, rho -0.271 (p=0.075)) or Melbourne (Spearman, rho -0.113 (p=0.472)). Furthermore, to assess if the relationship between location and total number of viral families per sample was stable regardless of age, we performed a semiparametric analysis of covariance in which the number of total viral families per sample was rank transformed. After adjusting for age, both the number of viral families (p= 0.001) and the number of viral genera (p=0.001) between the two sites were still statistically different. Therefore, age differences between the Northern Territory and Melbourne cohorts do not contribute significantly to the virome differences between the two regions.

### PCR of Adenovirus, Astrovirus, Coronavirus, Enterovirus, Norovirus, and Rotavirus

In order to independently confirm the sequencing results, we used PCR to define the prevalence of the most frequently detected viruses for which pan-family or pan-genus primers could be used including: adenovirus, astrovirus, enterovirus, norovirus, and rotavirus. We additionally screened the samples for a family not detected by sequencing, the *Coronaviridae*.

There was good concordance between the sequencing results and the qPCR for adenovirus. All ten of the positive samples by sequencing were confirmed by our qPCR. Additionally, there were two samples that were negative by sequencing, but positive by qPCR. The adenovirus qPCR was capable of detecting five copies per reaction (Supplemental Figure 1a). The two samples that sequencing failed to identify as adenovirus positive were the two with the lowest copy number by qPCR ( $6.9 \times 10^2$  and  $1.2 \times 10^3$  copies/g stool).

There were three samples sequencing positive for astrovirus. One of these samples was PCR positive. No samples were PCR positive and sequencing negative. One of the samples positive by sequencing and negative by PCR had 165 reads whose top blast hits were astroviruses. Close inspection of the consensus sequence of these reads showed 3 and 4 mismatches in the forward and reverse primer regions, respectively, which is a likely explanation for why it was not detected by PCR. The other sample positive by sequencing had only a single read, which was not in the region targeted by the primers, so it is unknown if sequence variation in the primer binding sites could be an explanation for this sample as well.

Nineteen samples were positive by sequencing for the enterovirus genus, including 3 that contained rhinovirus species and 16 that contained enterovirus species. We utilized a pan-enterovirus assay (Verstrepen et al., 2001) that does not detect rhinovirus species to evaluate concordance. Twelve of the 16 sequencing positive samples were qRT-PCR positive. The sensitivity limit of the enterovirus qRT-PCR assay was 500 copies as defined by its standard curve (Supplemental Figure 1b), so it is possible that these four discordant samples were below the limit of detection. In addition, there was one sample that was qRT-PCR positive, but sequencing negative.



Six samples were positive by sequencing for norovirus GII. All of these samples were detected by qRT-PCR. There was one additional sample that was qRT-PCR positive but sequencing negative. No samples were sequencing or qRT-PCR positive for norovirus GI.

There were eight samples sequencing positive for rotavirus. All eight of these samples were RT-PCR positive as well. Additionally, there were four samples from the Northern Territory that were positive for rotavirus by RT-PCR, but negative by sequencing. The sequencing based analysis of rotaviruses was not statistically different between the two sites ( $p=0.2658$ ). However, because the RT-PCR assay detected four additional positive samples, rotavirus prevalence as measured by RT-PCR was statistically significant ( $p=0.0264$ ). One factor that complicates interpretation of this difference is that fact that different vaccines were used in the two different regions. In Melbourne RotaTeq (Merck), a live human-bovine pentavalent reassortant vaccine is used, while Rotarix (GlaxoSmithKline), a live attenuated human vaccine strain, is used in the Northern Territory. The stool viral load of Rotarix has been found to be 100 fold higher than children receiving RotaTeq (Hsieh et al., 2014) which may have impacted the differences seen.

No samples were positive by sequencing for coronaviruses. One sample was positive by RT-PCR. This amplicon was cloned and Sanger sequenced and confirmed to be coronavirus 229E, which has been reported in diarrhea samples previously (Risku et al., 2010). Overall there was good concordance between the sequencing data and the PCR screening. The level of concordance observed was similar to that observed in other studies comparing next generation sequencing to PCR (Wylie et al., 2012).

### PCR of Adenovirus and Enterovirus in additional samples

To explore whether the differences seen in adenovirus and enterovirus, could be seen when examining a larger number of samples, we tested additional samples from Melbourne and Northern Territory. Further testing for rotavirus was not pursued as the vaccination status of each subject was unknown and could be a potential confounder. We tested samples from Melbourne ( $n=159$ ) and Northern Territory ( $n=165$ ), including all of the samples that were analyzed by metagenomic sequencing. Enteroviruses were more prevalent in the Northern Territory (18.0%) compared to Melbourne (0.6%) by qRT-PCR and adenoviruses were more abundant in the Northern Territory (18.8%) than in Melbourne (6.3%) by PCR confirming the metagenomic results.

To explore the generalizability of these results, we assessed the prevalence of enteroviruses and adenoviruses in additional diarrhea cohorts from Seattle ( $n=80$ ) and The Gambia ( $n=160$ ) (Table 3a-c), which yielded enterovirus positivity rates of 0% and 28.8%, respectively. Adenovirus positivity was even higher in The Gambia (41.9%), while comparable adenovirus frequencies were detected in Seattle (21.3%), and Northern Territory (18.8%).

## Discussion

We examined the diversity of viral communities in stools from 87 children with diarrhea from two different geographic locations collected in the same time period using standardized

inclusion criteria at both sites. This study design enabled us to minimize variables while defining and comparing the viromes of patients collected from the two sites.

Many virus families were equally prevalent in the two cohorts, including common diarrhea causing viruses such as norovirus (*Caliciviridae*) and astrovirus (*Astroviridae*). In addition, some viruses that have no known pathogenic properties, such as anelloviruses, were frequently detected in both cohorts. Viruses with limited identity to non-human viruses (ex. Israeli acute paralysis virus) were detected. It is unknown whether these represent novel human pathogens or are dietary passengers. Other studies of reclaimed water (Rosario et al., 2009), raw sewage (Cantalupo et al., 2011; Ng et al., 2012), and stool (Victoria et al., 2009) have also found invertebrate viruses. Moreover, it is striking that many samples had multiple viral families detected beyond the viruses typically associated with diarrhea. Most notable was the sample with 8 different viral families detected which raises the question of whether there may be some synergy between viruses (i.e. do some infections make the gut more permissive to infection by additional viruses?). Furthermore, these observations suggest the possibility that viruses are present in the human gut in a fashion analogous to the well-established bacterial microbiome. Longitudinal studies are needed to shed further light on the nature of the eukaryotic virome.

We demonstrated that the diarrhea samples from the Northern Territory had more viral families per sample than the samples from Melbourne. Further analysis demonstrated that the families *Adenoviridae*, *Picornaviridae*, and *Reoviridae* were more common in the Northern Territory than in Melbourne. Furthermore, we found that within the picornaviruses the genus *Enterovirus* was the most commonly detected and was preferentially found in the Northern Territory. These results clearly demonstrate that enteric viromes in patients with diarrhea can differ between two different geographic sites.

There are many possible explanations for the observed population-specific differences in the human diarrhea stool viromes. It is plausible that environmental factors such as diet, living conditions, water quality, hygiene and/or socioeconomic status could dictate the composition of the stool virome. The samples from Melbourne mainly represent patients living in a westernized and urban setting, while the samples from the Northern Territory represent patients largely from remote communities scattered across a large area. Health indicators among children in the Northern Territory, including for a range of infectious diseases, are worse than for children elsewhere in Australia (Gracey and King, 2009; Ruben and Walker, 1995). The paucity of enteroviruses in the Seattle cohort and their comparative abundance in children from The Gambia is consistent with the model that these viruses are associated with children living in impoverished conditions. Culture-based studies from several decades ago of healthy children also demonstrated a greater frequency of excretion of enteroviruses among persons of lower socioeconomic status (Honig et al., 1956; Otatume and Addy, 1975). Additionally, a recent study in Seattle showed no enterovirus detection in stools from children with diarrhea (Braun et al., 2012).

From the sequencing data there was not a statistically significant difference in the prevalence of rotavirus between Melbourne and the Northern Territory. RT-PCR confirmed the presence of rotavirus in these samples and identified four additional positive samples in



the Northern Territory. These additional positive samples then made rotavirus statically different between Melbourne and the Northern Territory. It is possible that these differences are related to the different rotavirus vaccinations given, the proportion of children that were vaccinated, or differences in vaccine efficacy.

While we observed differences in the prevalence of adenoviruses between Melbourne and the Northern Territory, broader analysis of additional diarrhea cohorts (Seattle and The Gambia) yielded a more complex picture. As with the enteroviruses, the highest rate of adenovirus positivity was in The Gambia (42%) and the lowest was in Melbourne (6%). The 42% adenovirus prevalence in The Gambia is similar to previously described reports from some other locations (France and Kenya) (Berciaud et al., 2012; Magwalivha et al., 2010). Unlike the situation with enteroviruses, adenoviruses were detected in approximately equal frequencies in Northern Territory and Seattle. Further studies are needed in larger populations, over multiple years, in these sites to ascertain whether these observations truly reflect geographic differences, or whether additional factors such as seasonality, year-to-year variation, or community outbreaks contribute to these findings.

## Conclusions

This study demonstrates major inter-population differences in the human stool eukaryotic virome. The potential etiologic role of the human stool virome in disorders with profound geographic differences remains to be seen. In one hypothesis, these disorders, which include environmental enteropathy, oral vaccine failures, autoimmune diseases, inflammatory bowel disease, and allergic conditions, could be consequences of perturbations in the stool virome. To evaluate this and other hypotheses, it will be important to define the childhood stool virome in other geographic locations as well in health and other states of disease.

## Methods

All stool samples were obtained with parental/guardian consent under protocols approved by Human Studies Committees from Seattle Children's Hospital, University of Maryland Baltimore, the Central Australian Human Research Ethics Committee, the Human Research Ethics Committee of the Menzies School of Health Research, the Northern Territory Department of Health and Families, the Royal Children's Hospital, and the Joint Ethics Committee of the Gambia Government/Medical Research Council Unit. The use of deidentified samples from the previous studies was approved by the Human Studies Committee of Washington University in St. Louis.

### Australian stool samples

Stool samples were collected from children less than 5 years old admitted to The Royal Children's Hospital, Melbourne, Victoria, Australia and Alice Springs Hospital or Royal Darwin Hospital Northern Territory, Australia, with acute diarrhea between December 2009 and September 2010. These samples were collected under the same inclusion criteria and were archived at -70°C in Melbourne. Rotavirus vaccine was part of the routine childhood vaccinations throughout the collection period. In Melbourne RotaTeq (Merck), a live human-bovine pentavalent reassortant vaccine is used, while Rotarix (GlaxoSmithKline), a

live attenuated human vaccine strain is used in the Northern Territory. 165 samples from the Northern Territory and 159 samples from Melbourne were randomly selected for this study. Of these, 44 samples from Northern Territory and 43 samples from Melbourne were sequenced.

### Seattle stool samples

Stool samples were collected from children less than 5 years old who were evaluated for self-defined diarrhea at the Seattle Children's Hospital Emergency Department from 2003-2005 as part of a prospective study (Denno et al., 2012). Eighty samples were randomly selected from this cohort for the adenovirus and enterovirus qPCR prevalence phase of this study.

### Gambian stool samples

Stool samples were collected from children with moderate-to-severe diarrhea less than 5 years old from 2008-2009 as part of the Global Enteric Multi-center Study (GEMS) (Kotloff et al., 2012). Diarrhea was defined as new (onset after > 7 diarrhea free days), acute (onset within the previous 7 days) and having one of the following criteria: sunken eyes, loss of skin turgor, intravenous hydration administered or prescribed, visible blood in loose stools, or admission to hospital with diarrhea (Kotloff et al., 2013). Samples were collected from children located in the Upper River Division (URD), a rural region in The Gambia. 160 diarrhea stools were randomly selected from this cohort for the adenovirus and enterovirus qPCR prevalence phase of this study.

### Nucleic acid preparation and sequencing

Frozen stool was resuspended in 6 volumes of sterile PBS and centrifuged. Resulting supernatants were passed through a 0.45 um filter. Total nucleic acids (RNA and DNA) were isolated from the filtrate using the Ampliprep DNA extraction machine (Roche). In order to evaluate samples for both RNA and DNA viruses the total nucleic acids were randomly amplified. Specifically, RNA present in the sample was converted to cDNA by reverse transcription. Following second strand synthesis, the double-stranded cDNA and any single-stranded or double-stranded DNA were then PCR amplified using unique sequence primers which were used as a barcode to assign sequences to their corresponding sample (Wang et al., 2003). The products were then pooled, adaptor ligated and sequenced using the 454 GS FLX Titanium platform (454 Life Science).

### Analysis of Sequencing Reads

Sequences were analyzed using a custom bioinformatics pipeline (VirusHunter) (Zhao et al., 2013). Briefly, sequences were deconvoluted based on their barcode (Supplemental table 2). To reduce the risk of cross contamination between samples barcodes were required to be present at both ends of the read with a perfect match and was trimmed before any further analysis. To reduce the number of near identical sequences introduced from PCR bias, sequences were clustered using CD-HIT (Li and Godzik, 2006) requiring 95% identity over 95% of the sequence. The longest sequence from each cluster was chosen as the representative unique sequence. Sequences were then masked by RepeatMasker (<http://>

[www.repeatmasker.org](http://www.repeatmasker.org)). If a sequence did not contain a stretch of at least 50 consecutive non-“N” nucleotides or if greater than 40% of the total length of the sequence was masked, it was removed from further analysis. The remaining high quality unique sequences were then sequentially compared to (1) human genome using BLASTn, (2) nt database by BLASTn and (3) nr database using BLASTx. Minimum E-value cutoffs of  $1 \times 10^{-10}$  for BLASTn and  $1 \times 10^{-5}$  for BLASTx were used. Sequences were then binned based on the top BLAST hit to human, mouse, fungal, bacterial, phage, viral, and other. Sequences aligning to both a virus and another kingdom with the same E-value were classified as ambiguous and not subjected to further analysis. Viral sequences were binned into families based on the taxonomical identity of the best BLAST hit as defined by the NCBI Taxonomy database. Results from this approach have been previously compared directly to other similar BLAST-based taxonomic classifiers such as MEGAN (Handley et al., 2012). Samples were considered positive for a family (or genus) if at least one unique read was present. From previous experience with stool virome analysis, alignments to particular viral families in the NCBI Taxonomy database (herpesviridae, iridoviridae, mimiviridae, phycodnaviridae, poxviridae, unclassified dsDNA viruses, and environmental samples) have a high frequency of false positives, usually due to low complexity and/or repetitive sequence motifs yielding artefactual alignments. Thus, all sequences aligning to these families were manually evaluated and those deemed to be false positives were removed from further analysis.

### Assembly of viral genome

Assembly was performed on one sample, which was noted to contain 110 unique sequences with limited similarity to viruses in the order of *Picornavirales*. Sequences identified as viral as well as sequences that had no significant hit to any sequence in the NR and NT databases were assembled using Newbler (454 Life Sciences, Branford, CT) with default parameters.

### Real-time RT-PCR (qRT-PCR) screening for enterovirus

A previously described pan-enterovirus TaqMan assay targeting conserved sequences in the 5'UTR region of the genome (Verstrepen et al., 2001) was used to assess enterovirus prevalence in all cohorts. The qRT-PCR was performed using the One-Step RT-PCR TaqMan kit (Applied Biosystems). The 25  $\mu$ L reaction included 5  $\mu$ L of extracted sample, 12.5 pmol of each primer, and 6.25 pmol of probe. The following cycling conditions were used: 48°C for 30 min, 95°C for 10 min, 40 cycles of 95°C for 15 sec and 60°C for 1 min. To generate a standard curve for this assay, *in vitro* transcribed RNA was generated from a plasmid containing the region of interest using MAXIscript (Ambion) per the manufacturer's protocol. Serial dilutions of this *in vitro* transcribed RNA from  $5 \times 10^6$  to 5 copies were used to generate a standard curve and a limit of detection of 500 copies was defined (Supplemental Figure 1b). Samples were tested in a 96-well plate format with 8 negative controls and 1 positive control per plate. The threshold of all plates was set at a standard value, and samples were counted as positive if their cycle threshold was <38.00.

### qPCR screening for adenovirus

A pan-adenovirus TaqMan assay that detects all adenovirus species was used to assess adenovirus prevalence in all cohorts (Jothikumar et al., 2005). qPCR was performed with the

TaqMan Universal PCR Master Mix (Applied Biosystems). The 25  $\mu$ L reaction included 5  $\mu$ L of extracted sample, 12.5 pmol of each primer, and 4 pmol of probe. The following cycling conditions were used: 50°C for 2 min, 95°C for 10 min, 45 cycles of 95°C for 15 sec and 60°C for 1 min. To establish a standard curve for this assay, a plasmid containing the region of interest was used in serial dilutions of  $5 \times 10^6$  to 5 copies and a limit of detection of 5 copies was defined (Supplemental Figure 1a). Samples were tested in a 96-well plate format with 8 negative controls and 1 positive control per plate. The threshold of all plates was set at a standard value, and samples were counted as positive if their cycle threshold was <38.00.

### Real-time RT-PCR (qRT-PCR) screening for norovirus

We modified a previously described assay to detect norovirus GI and norovirus GII separately (Rolfe et al., 2007). This was used to independently confirm our sequencing results. The qRT-PCR was performed using the One-Step RT-PCR TaqMan kit (Applied Biosystems). The primers and probes were as previously described (Rolfe et al., 2007) except RING2P was labeled with VIC instead of JOE. The 25  $\mu$ L reaction included 5  $\mu$ L of extracted sample, 12.5 pmol of each primer, and 6.25 pmol of probe. The following cycling conditions were used: 50°C for 30 min, 95°C for 10 min, 50 cycles of 95°C for 15 sec and 56°C for 1 min. To generate a standard curve for this assay, *in vitro* transcribed RNA was generated from plasmids containing the regions of interest using MEGAscript (Ambion) per the manufacturer's protocol. Serial dilutions of these *in vitro* transcribed RNA from  $5 \times 10^6$  to 5 copies were used to generate a standard curve and a limit of detection of 5 copies was defined (Supplemental Figure 1c and d) for both assays. Samples were tested in a 96-well plate format with 8 negative controls and 1 positive control per plate. The threshold of all plates was set at a standard value, and samples were counted as positive if their cycle threshold was <38.00.

### Reverse transcription PCR (RT-PCR) for astrovirus

A previously described astrovirus consensus RT-PCR was used to independently confirm our sequencing results (Finkbeiner et al., 2009). Qiagen OneStep RT-PCR kit was used to screen 3 $\mu$ L of extracted material from each sample using the consensus primers SF0073 (5'-GATTGGACTCGATTTGATGG-3') and SF0076 (5'-CTGGCTTAACCCACATTCC-3') which target the ORF 1b (RNA polymerase) of astroviruses under the following cycling conditions: 30 min RT step, 94°C hold for 10 min, followed by 40 cycles of 94°C for 30s, 52°C for 30s, and 72°C for 50s.

### RT-PCR for rotavirus

To independently confirm our sequencing results a previously described RT-PCR to detect group A rotavirus was used (Rao et al., 1995). RT-PCR screening was done using the Qiagen OneStep RT-PCR kit. Each 25 $\mu$ L reaction used 3 $\mu$ L of extracted material from each sample and forward (5'-ATGCTCAAGATGGAGTCT-3') and reverse (5'-GGTCACATAACGCCCTAT-3') primers which target nonstructural protein (NSP) 3 under the following cycling conditions: 30 min RT step, 94°C hold for 10 min, followed by 40 cycles of 94°C for 30s, 56°C for 30s, and 72°C for 1 min.

## RT-PCR for coronavirus

A previously described pan-coronavirus RT-PCR was used to independently confirm our sequencing results (Vijgen et al., 2008). Qiagen OneStep RT-PCR kit was used to screen 3 $\mu$ L of extracted material from each sample using the primers Cor-FW (5'-ACWCARHTVAAYYTNAARTAYGC -3') and Cor-RV (5'-TCRCAYTTDGGRTARTCCCA -3') which target a 251 bp region of the RNA polymerase of coronaviruses under the following cycling conditions: 30 min RT step, 94°C hold for 10 min, followed by 50 cycles of 94°C for 30s, 48°C for 30s, and 72°C for 1min. PCR amplicons were cloned into pCR4 (Invitrogen) and sequenced using standard Sanger sequencing technology.

## Statistical Analysis

Two-tailed P values were derived using Wilcoxon rank-sum, chi-square, Fisher's exact, or t-tests, as indicated. Wilcoxon rank-sum test, a non-parametric test, was used to compare ordinal data. Fisher's exact test was used for categorical data (ex. presence or absence). Assessment of correlation was done using Spearman Rank. For the site comparisons of qPCR results, chi-square with Sidak adjustment for multiple comparisons was used. Statistical analyses were performed using SAS (Cary, NC) software version 9.2.

## Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

## Acknowledgments

This work was funded in part by National Institutes of Health [grant U54 AI057160] to the Midwest Regional Center of Excellence for Biodefense and Emerging Infectious Disease Research, US Department of Agriculture [grant 0202238], Bill & Melinda Gates Foundation [OPP1016839], Children's Discovery Institute (MD-FR-2013-292), the Morphology and Biobank Cores of the Washington University School of Medicine Digestive Diseases Research Core Center [P30DK052574], and National Institutes of Health-National Center for Advancing Translational Sciences [UL1 TR000448 sub award KL2 TR000450 to L.R.H]. D.W. holds an Investigator in the Pathogenesis of Infectious Disease Award from the Burroughs Wellcome Fund. C.D.K. is supported by a Career Development Award fellowship from the National Health and Medical Research Council, Australia (607347).

## References

- Berciaud S, Rayne F, Kassab S, Jubert C, Faure-Della Corte M, Salin F, Wodrich H, Lafon ME. Adenovirus infections in Bordeaux University Hospital 2008-2010: clinical and virological features. *J Clin Virol.* 2012; 54:302–307. [PubMed: 22608365]
- Braun L, Renaud C, Fairchok M, Kuypers J, Englund J, Martin E. Human Parechovirus and Other Enteric Viruses in Childcare Attendees in the Era of Rotavirus Vaccines. *Journal of Pediatric Infectious Diseases Society.* 2012; 1:136–143.
- Breitbart M, Haynes M, Kelley S, Angly F, Edwards RA, Felts B, Mahaffy JM, Mueller J, Nulton J, Rayhawk S, Rodriguez-Brito B, Salamon P, Rohwer F. Viral diversity and dynamics in an infant gut. *Res Microbiol.* 2008; 159:367–373. [PubMed: 18541415]
- Campbell DI, Murch SH, Elia M, Sullivan PB, Sanyang MS, Jobarteh B, Lunn PG. Chronic T cell-mediated enteropathy in rural west African children: relationship with nutritional status and small bowel function. *Pediatr Res.* 2003; 54:306–311. [PubMed: 12788978]
- Cantalupo PG, Calgua B, Zhao G, Hundesa A, Wier AD, Katz JP, Grabe M, Hendrix RW, Girones R, Wang D, Pipas JM. Raw sewage harbors diverse viral populations. *MBio.* 2011; 2

- De Filippo C, Cavalieri D, Di Paola M, Ramazzotti M, Poullet JB, Massart S, Collini S, Pieraccini G, Lionetti P. Impact of diet in shaping gut microbiota revealed by a comparative study in children from Europe and rural Africa. *Proc Natl Acad Sci U S A*. 2010; 107:14691–14696. [PubMed: 20679230]
- Denno DM, Shaikh N, Stapp JR, Qin X, Hutter CM, Hoffman V, Mooney JC, Wood KM, Stevens HJ, Jones R, Tarr PI, Klein EJ. Diarrhea etiology in a pediatric emergency department: a case control study. *Clinical infectious diseases : an official publication of the Infectious Diseases Society of America*. 2012; 55:897–904. [PubMed: 22700832]
- Finkbeiner SR, Allred AF, Tarr PI, Klein EJ, Kirkwood CD, Wang D. Metagenomic analysis of human diarrhea: viral detection and discovery. *PLoS pathogens*. 2008; 4:e1000011. [PubMed: 18398449]
- Finkbeiner SR, Le BM, Holtz LR, Storch GA, Wang D. Detection of newly described astrovirus MLB1 in stool samples from children. *Emerg Infect Dis*. 2009; 15:441–444. [PubMed: 19239759]
- Frank DN, St Amand AL, Feldman RA, Boedeker EC, Harpaz N, Pace NR. Molecular-phylogenetic characterization of microbial community imbalances in human inflammatory bowel diseases. *Proc Natl Acad Sci U S A*. 2007; 104:13780–13785. [PubMed: 17699621]
- Gracey M, King M. Indigenous health part 1: determinants and disease patterns. *Lancet*. 2009; 374:65–75. [PubMed: 19577695]
- Handley SA, Thackray LB, Zhao G, Presti R, Miller AD, Droit L, Abbink P, Maxfield LF, Kambal A, Duan E, Stanley K, Kramer J, Macri SC, Permar SR, Schmitz JE, Mansfield K, Brenchley JM, Veazey RS, Stappenbeck TS, Wang D, Barouch DH, Virgin HW. Pathogenic simian immunodeficiency virus infection is associated with expansion of the enteric virome. *Cell*. 2012; 151:253–266. [PubMed: 23063120]
- Honig EI, Melnick JL, Isacson P, Parr R, Myers IL, Walton M. An epidemiological study of enteric virus infections: poliomyelitis, coxsackie, and orphan (ECHO) viruses isolated from normal children in two socioeconomic groups. *The Journal of experimental medicine*. 1956; 103:247–262. [PubMed: 13286430]
- Hsieh YC, Wu FT, Hsiung CA, Wu HS, Chang KY, Huang YC. Comparison of virus shedding after lived attenuated and pentavalent reassortant rotavirus vaccine. *Vaccine*. 2014; 32:1199–1204. [PubMed: 24076325]
- Jothikumar N, Cromeans TL, Hill VR, Lu X, Sobsey MD, Erdman DD. Quantitative real-time PCR assays for detection of human adenoviruses and identification of serotypes 40 and 41. *Applied and environmental microbiology*. 2005; 71:3131–3136. [PubMed: 15933012]
- 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. *Clinical infectious diseases : an official publication of the Infectious Diseases Society of America*. 2012; 55(Suppl 4):S232–245. [PubMed: 23169936]
- Kotloff KL, Nataro JP, Blackwelder WC, Nasrin D, Farag TH, Panchalingam S, Wu Y, Sow SO, Sur D, Breiman RF, Faruque AS, Zaidi AK, Saha D, Alonso PL, Tamboura B, Sanogo D, Onwuchekwa U, Manna B, Ramamurthy T, Kanungo S, Ochieng JB, Omere R, Oundo JO, Hossain A, Das SK, Ahmed S, Qureshi S, Quadri F, Adegbola RA, Antonio M, Hossain MJ, Akinsola A, Mandomando I, Nhampossa T, Acacio S, Biswas K, O'Reilly CE, Mintz ED, Berkeley LY, Muhsen K, Sommerfelt H, Robins-Browne RM, Levine MM. Burden and aetiology of diarrhoeal disease in infants and young children in developing countries (the Global Enteric Multicenter Study, GEMS): a prospective, case-control study. *Lancet*. 2013
- Larsen N, Vogensen FK, van den Berg FW, Nielsen DS, Andreasen AS, Pedersen BK, Al-Soud WA, Sorensen SJ, Hansen LH, Jakobsen M. Gut microbiota in human adults with type 2 diabetes differs from non-diabetic adults. *PLoS One*. 2010; 5:e9085. [PubMed: 20140211]
- Li L, Kapoor A, Slikas B, Bamidele OS, Wang C, Shaukat S, Masroor MA, Wilson ML, Ndjanga JB, Peeters M, Gross-Camp ND, Muller MN, Hahn BH, Wolfe ND, Triki H, Bartkus J, Zaidi SZ, Delwart E. Multiple diverse circoviruses infect farm animals and are commonly found in human and chimpanzee feces. *J Virol*. 2010; 84:1674–1682. [PubMed: 20007276]

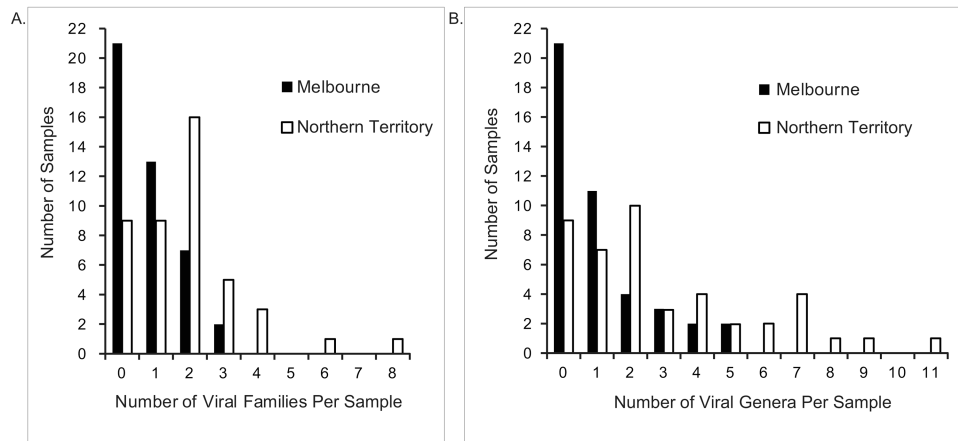


- Li W, Godzik A. Cd-hit: a fast program for clustering and comparing large sets of protein or nucleotide sequences. *Bioinformatics*. 2006; 22:1658–1659. [PubMed: 16731699]
- Lindenbaum J. Small intestine dysfunction in Pakistanis and Americans resident in Pakistan. *Am J Clin Nutr*. 1968; 21:1023–1029. [PubMed: 5675841]
- Lindenbaum J, Gerson CD, Kent TH. Recovery of small-intestinal structure and function after residence in the tropics. I. Studies in Peace Corps volunteers. *Ann Intern Med*. 1971; 74:218–222. [PubMed: 5545229]
- Magwalivha M, Wolfaardt M, Kiulia NM, van Zyl WB, Mwenda JM, Taylor MB. High prevalence of species D human adenoviruses in fecal specimens from Urban Kenyan children with diarrhea. *J Med Virol*. 2010; 82:77–84. [PubMed: 19950234]
- Mai V, Young CM, Ukhanova M, Wang X, Sun Y, Casella G, Theriaque D, Li N, Sharma R, Hudak M, Neu J. Fecal microbiota in premature infants prior to necrotizing enterocolitis. *PLoS One*. 2011; 6:e20647. [PubMed: 21674011]
- Menzies IS, Zuckerman MJ, Nukajam WS, Somasundaram SG, Murphy B, Jenkins AP, Crane RS, Gregory GG. Geography of intestinal permeability and absorption. *Gut*. 1999; 44:483–489. [PubMed: 10075954]
- Minot S, Bryson A, Chehoud C, Wu GD, Lewis JD, Bushman FD. Rapid evolution of the human gut virome. *Proc Natl Acad Sci U S A*. 2013; 110:12450–12455. [PubMed: 23836644]
- Minot S, Sinha R, Chen J, Li H, Keilbaugh SA, Wu GD, Lewis JD, Bushman FD. The human gut virome: inter-individual variation and dynamic response to diet. *Genome Res*. 2011; 21:1616–1625. [PubMed: 21880779]
- Nakamura S, Yang CS, Sakon N, Ueda M, Tougan T, Yamashita A, Goto N, Takahashi K, Yasunaga T, Ikuta K, Mizutani T, Okamoto Y, Tagami M, Morita R, Maeda N, Kawai J, Hayashizaki Y, Nagai Y, Horii T, Iida T, Nakaya T. Direct metagenomic detection of viral pathogens in nasal and fecal specimens using an unbiased high-throughput sequencing approach. *PLoS One*. 2009; 4:e4219. [PubMed: 19156205]
- Ng TF, Marine R, Wang C, Simmonds P, Kapusinszky B, Bodhidatta L, Oderinde BS, Wommack KE, Delwart E. High variety of known and new RNA and DNA viruses of diverse origins in untreated sewage. *J Virol*. 2012; 86:12161–12175. [PubMed: 22933275]
- Okamoto H, Akahane Y, Ukita M, Fukuda M, Tsuda F, Miyakawa Y, Mayumi M. Fecal excretion of a nonenveloped DNA virus (TTV) associated with posttransfusion non-A-G hepatitis. *Journal of medical virology*. 1998; 56:128–132. [PubMed: 9746068]
- Otatume S, Addy P. Ecology of enteroviruses in tropics. I. Circulation of enteroviruses in healthy infants in tropical urban area. *Japanese journal of microbiology*. 1975; 19:201–209. [PubMed: 1165613]
- Pallansch, M.; Roos, R. Enteroviruses: Polioviruses, Coxsackieviruses, Echoviruses, and Newer Enteroviruses. In: Knipe, D.; Howley, P., editors. *Fields Virology*. 5th. Lippincott Williams and Wilkins; Philadelphia: 2007. p. 839-894.
- Phan TG, Vo NP, Bonkougou IJ, Kapoor A, Barro N, O’Ryan M, Kapusinszky B, Wang C, Delwart E. Acute diarrhea in West African children: diverse enteric viruses and a novel parvovirus genus. *J Virol*. 2012; 86:11024–11030. [PubMed: 22855485]
- Rao CD, Das M, Ilango P, Lalwani R, Rao BS, Gowda K. Comparative nucleotide and amino acid sequence analysis of the sequence-specific RNA-binding rotavirus nonstructural protein NSP3. *Virology*. 1995; 207:327–333. [PubMed: 7871749]
- Reyes A, Haynes M, Hanson N, Angly FE, Heath AC, Rohwer F, Gordon JI. Viruses in the faecal microbiota of monozygotic twins and their mothers. *Nature*. 2010; 466:334–338. [PubMed: 20631792]
- Risku M, Lappalainen S, Rasanen S, Vesikari T. Detection of human coronaviruses in children with acute gastroenteritis. *J Clin Virol*. 2010; 48:27–30. [PubMed: 20233673]
- Rolfe KJ, Parmar S, Mururi D, Wreghitt TG, Jalal H, Zhang H, Curran MD. An internally controlled, one-step, real-time RT-PCR assay for norovirus detection and genogrouping. *J Clin Virol*. 2007; 39:318–321. [PubMed: 17604686]
- Rosario K, Nilsson C, Lim YW, Ruan Y, Breitbart M. Metagenomic analysis of viruses in reclaimed water. *Environ Microbiol*. 2009; 11:2806–2820. [PubMed: 19555373]

- Ruben AR, Walker AC. Malnutrition among rural aboriginal children in the Top End of the Northern Territory. *The Medical journal of Australia*. 1995; 162:400–403. [PubMed: 7746170]
- Sartor RB. Microbial influences in inflammatory bowel diseases. *Gastroenterology*. 2008; 134:577–594. [PubMed: 18242222]
- Smits SL, Schapendonk CM, van Beek J, Vennema H, Schurch AC, Schipper D, Bodewes R, Haagmans BL, Osterhaus AD, Koopmans MP. New viruses in idiopathic human diarrhea cases, the Netherlands. *Emerg Infect Dis*. 2014; 20:1218–1222. [PubMed: 24964003]
- Strachan DP. Hay fever, hygiene, and household size. *BMJ*. 1989; 299:1259–1260. [PubMed: 2513902]
- Turnbaugh PJ, Hamady M, Yatsunenko T, Cantarel BL, Duncan A, Ley RE, Sogin ML, Jones WJ, Roe BA, Affourtit JP, Egholm M, Henrissat B, Heath AC, Knight R, Gordon JI. A core gut microbiome in obese and lean twins. *Nature*. 2009; 457:480–484. [PubMed: 19043404]
- van Leeuwen M, Williams MM, Koraka P, Simon JH, Smits SL, Osterhaus AD. Human picobirnaviruses identified by molecular screening of diarrhea samples. *Journal of clinical microbiology*. 2010; 48:1787–1794. [PubMed: 20335418]
- Verstrepen WA, Kuhn S, Kockx MM, Van De Vyvere ME, Mertens AH. Rapid detection of enterovirus RNA in cerebrospinal fluid specimens with a novel single-tube real-time reverse transcription-PCR assay. *Journal of clinical microbiology*. 2001; 39:4093–4096. [PubMed: 11682535]
- Victoria JG, Kapoor A, Li L, Blinkova O, Slikas B, Wang C, Naeem A, Zaidi S, Delwart E. Metagenomic analyses of viruses in stool samples from children with acute flaccid paralysis. *J Virol*. 2009; 83:4642–4651. [PubMed: 19211756]
- Vijgen L, Moes E, Keyaerts E, Li S, Van Ranst M. A pancoronavirus RT-PCR assay for detection of all known coronaviruses. *Methods Mol Biol*. 2008; 454:3–12. [PubMed: 19057882]
- Wang D, Urisman A, Liu YT, Springer M, Ksiazek TG, Erdman DD, Mardis ER, Hickenbotham M, Magrini V, Eldred J, Latreille JP, Wilson RK, Ganem D, DeRisi JL. Viral discovery and sequence recovery using DNA microarrays. *PLoS biology*. 2003; 1:E2. [PubMed: 14624234]
- Wootton SH, Scheifele DW, Mak A, Petric M, Skowronski DM. Detection of human influenza virus in the stool of children. *The Pediatric infectious disease journal*. 2006; 25:1194–1195. [PubMed: 17133173]
- Wylie KM, Mihindukulasuriya KA, Sodergren E, Weinstock GM, Storch GA. Sequence analysis of the human virome in febrile and afebrile children. *PLoS One*. 2012; 7:e27735. [PubMed: 22719819]
- Yatsunenko T, Rey FE, Manary MJ, Trehan I, Dominguez-Bello MG, Contreras M, Magris M, Hidalgo G, Baldassano RN, Anokhin AP, Heath AC, Warner B, Reeder J, Kuczynski J, Caporaso JG, Lozupone CA, Lauber C, Clemente JC, Knights D, Knight R, Gordon JI. Human gut microbiome viewed across age and geography. *Nature*. 2012; 486:222–227. [PubMed: 22699611]
- Zhang T, Breitbart M, Lee WH, Run JQ, Wei CL, Soh SW, Hibberd ML, Liu ET, Rohwer F, Ruan Y. RNA viral community in human feces: prevalence of plant pathogenic viruses. *PLoS biology*. 2006; 4:e3. [PubMed: 16336043]
- 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]

### Highlights

- 22 different viral families detected in pediatric diarrhea
- More viral families in diarrhea from Northern Territory than diarrhea from Melbourne
- *Adenoviridae* and *Picornaviridae* more common in Northern Territory than in Melbourne
- qRT-PCR/PCR confirmed the increased prevalence of adenoviruses and enteroviruses
- Viromes in patients with diarrhea differ between two different geographic sites



**Figure 1.** Number of viral families (A) and genera (B) in Northern Territory and Melbourne. Diarrhea samples from the Northern Territory contained more viral families and genera per sample than diarrhea samples from Melbourne.

**Table 1**  
**Sequencing statistics by cohort**

	<b>Northern Territory</b>	<b>Melbourne</b>
Average read length	329 nt	339 nt
Total reads	406,252	646,630
Unique reads	118,420	76,347
Unique viral reads	2,754 (2.3%)	1,266 (1.7%)
Unique bacterial reads	73,164 (61.8%)	50,196 (65.7%)
Unique phage reads	3,417 (2.9%)	3,680 (4.8%)
Unique fungi reads	12,643 (10.7%)	1,897 (2.5%)
Unique human reads	8,183 (6.9%)	3,212 (4.2%)
Unique other reads	1055 (0.9%)	1,317 (1.7%)
No similarity to sequences in GenBank	17,204 (14.5%)	14,779 (19.4%)

**Table 2**  
**Number of samples with picornavirus genera detected by sequencing**

	<b>Northern Territory</b>	<b>Melbourne</b>
Cardiovirus	1	1
Enterovirus	17	2
Klassevirus	1	0
Kobuvirus	1	0
Parechovirus	6	2



**Table 3****a. qPCR results**

	<b>Adenovirus qPCR positive</b>	<b>Enterovirus qRT-PCR positive</b>
Melbourne (n=159)	10 (6.3%)	1 (0.6%)
Northern Territory (n=165)	31 (18.8%)	29 (18.0%)
Seattle (n=80)	17 (21.3%)	0 (0%)
The Gambia (n=160)	67 (41.9%)	46 (28.8%)

**Table 3b. Adenovirus qPCR site comparison p values**

	<b>Northern Territory</b>	<b>Seattle</b>	<b>The Gambia</b>
Melbourne	0.0042	0.0036	0.0006
Northern Territory		0.9981	0.0006
Seattle			0.0096
The Gambia			

**Table 3c. Enterovirus qRT-PCR site comparison p values**

	<b>Northern Territory</b>	<b>Seattle</b>	<b>The Gambia</b>
Melbourne	0.0006	0.9796	0.0006
Northern Territory		0.0006	0.0967
Seattle			0.0006
The Gambia			