Virology 468-470 (2014) 556–564

Contents lists available at ScienceDirect

Virology

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

Geographic variation in the eukaryotic virome of human diarrhea

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

[a] Washington University School of Medicine, Department of Pediatrics, St. Louis, MO, USA
[b] Washington University School of Medicine, Department of Pathology and Immunology, St. Louis, MO, USA
[c] Washington University School of Medicine, Department of Molecular Microbiology, St. Louis, MO, USA
[d] University of Washington, Department of Pediatrics, Seattle, WA, USA
[e] University of Washington, Department of Global Health, Seattle, WA, USA
[f] Medical Research Council Unit, Gambia
[g] Department of Epidemiology and Public Health, University of Maryland School of Medicine, Baltimore, MD, USA
[h] Menzies School of Health Research, Charles Darwin University, Northern Territory, Australia
[i] Telethon Institute for Child Health Research, Centre for Child Health Research, University of Western Australia, Australia
[j] Enteric Virus Research Group, Murdoch Childrens Research Institute, Royal Children's Hospital, Victoria, Australia

ARTICLE INFO

Article history:
Received 12 May 2014
Returned to author for revisions
10 June 2014
Accepted 10 September 2014

Keywords:
Diarrhea
Virome
Metagenomics

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 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.

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

* Corresponding author at: Campus Box 8230, Department of Molecular Microbiology, 660 S. Euclid Ave., St. Louis, MO 63110, USA. Fax: +1 314 362 1232.
E-mail addresses: Holtz_l@kids.wustl.edu (L.R. Holtz), scao@path.wustl.edu (S. Cao), GZhao@pathology.wustl.edu (G. Zhao), Bauer_I@kids.wustl.edu (I.K. Bauer), ddenno@u.washington.edu (D.M. Denno), eileen.klein@seattlechildrens.org (E.J. Klein), mantonio@mrc.gm (M. Antonio), ostin001@umaryland.edu (O.C. Stine), tsnelling@ichr.uwa.edu.au (T.L. Snelling), carl.kirkwood@mcri.edu.au (C.D. Kirkwood), davewang@wustl.edu (D. Wang).

http://dx.doi.org/10.1016/j.virol.2014.09.012
0042-6822/

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% 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.

Table 1
Sequencing statistics by cohort.

	Northern Territory	Melbourne
Average read length	329 nt	339 nt
Total reads	406,252	646,630
Unique reads	118,420	76,347
Unique viral reads	2754 (2.3%)	1266 (1.7%)
Unique bacterial reads	73,164 (61.8%)	50,196 (65.7%)
Unique phage reads	3417 (2.9%)	3680 (4.8%)
Unique fungi reads	12,643 (10.7%)	1897 (2.5%)
Unique human reads	8183 (6.9%)	3212 (4.2%)
Unique other reads	1055 (0.9%)	1317 (1.7%)
No similarity to sequences in GenBank	17,204 (14.5%)	14,779 (19.4%)

Viruses detected by metagenomic analysis

Since 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 (Supplementary 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 (Fig. 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 and Roos, 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 9898 nt (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