

The plasma virome of febrile adult Kenyans shows frequent parvovirus B19 infections and a novel arbovirus (Kadipiro virus)

Carolyn N. Ngoi,^{1,2} Juliana Siqueira,^{2,3} Linlin Li,² Xutao Deng,² Peter Mugo,¹ Susan M. Graham,^{1,4} Matt A. Price,^{5,6} Eduard J. Sanders^{1,7} and Eric Delwart^{2,8}

Correspondence

Eric Delwart

delwarte@medicine.ucsf.edu

¹Centre for Geographic Medicine Research – Coast, Kenya Medical Research Institute, Kilifi, Kenya

²Blood Systems Research Institute, San Francisco, CA, USA

³Programa de Oncovirologia, Instituto Nacional de Câncer, Rio de Janeiro, Brazil

⁴University of Washington, Seattle, WA, USA

⁵International AIDS Vaccine Initiative, New York, NY, USA

⁶Department of Epidemiology and Biostatistics, University of California at San Francisco, CA, USA

⁷Nuffield Department of Medicine, University of Oxford, Headington, UK

⁸Department of Laboratory Medicine, University of California at San Francisco, CA, USA

Viral nucleic acids present in the plasma of 498 Kenyan adults with unexplained fever were characterized by metagenomics analysis of 51 sample pools. The highest to lowest fraction of plasma pools was positive for parvovirus B19 (75 %), pegivirus C (GBV-C) (67 %), alpha anellovirus (59 %), gamma anellovirus (55 %), beta anellovirus (41 %), dengue virus genotype 2 (DENV-2) (16 %), human immunodeficiency virus type 1 (6 %), human herpesvirus 6 (6 %), HBV (4 %), rotavirus (4 %), hepatitis B virus (4 %), rhinovirus C (2 %), Merkel cell polyomavirus (MCPyV; 2 %) and Kadipiro virus (2 %). Ranking by overall percentage of viral reads yielded similar results. Characterization of viral nucleic acids in the plasma of a febrile East African population showed a high frequency of parvovirus B19 and DENV infections and detected a reovirus (Kadipiro virus) previously reported only in Asian *Culex* mosquitoes, providing a baseline to compare with future virome studies to detect emerging viruses in this region.

Received 16 June 2016

Accepted 23 October 2016

INTRODUCTION

Cryopreserved samples from patients with unexplained fever can be used for emerging virus surveillance. Enrichment of viral nucleic acids followed by random nucleic acid amplification, deep DNA sequencing, and similarity searches to all previously sequenced viral genomes or proteomes have each been used to identify known and previously uncharacterized ‘new’ viral genomes. Such viral metagenomics approaches were pioneered in 2001 using mammalian samples (Allander *et al.*, 2001) and in 2002 with environmental samples (Breitbart *et al.*, 2002). The advent of next-generation sequencing has led to the

characterization of numerous viral genomes from clinical samples, cell culture supernatants, or diverse environmental sources.

To describe the plasma virome and identify emerging viruses in a population of adults with unexplained fever from East Africa, we enriched, sequenced, and identified viral nucleic acids from 498 febrile adults enrolled in a study of acute human immunodeficiency virus type 1 (HIV-1) infection (AHI) at two sites in coastal Kenya.

RESULTS

Febrile population analysed

We enrolled 994 eligible participants for AHI evaluation, of whom 672 were febrile (Fig. 1). We excluded 98 (14.6 %) participants diagnosed with malaria, two (0.3 %) participants

The raw sequence data for each pool is available at the National Center for Biotechnology Information (NCBI) Short Read Archive under GenBank accession number SRP090133.

diagnosed with AHI, and one (0.15 %) participant diagnosed with AHI and malaria co-infection. Of the 571 remaining participants, 20 had insufficient samples and two had not consented to testing their samples for causes of fever other than AHI and malaria. We further excluded 41 (8.4 %) of 489 samples tested for dengue virus (DENV) that were positive for DENV RNA. In total, 498 samples were available for sequencing, including 60 samples that had not been tested for DENV and two DENV-RNA-positive samples that were incorrectly included among sequenced samples (Fig. 1). Febrile patients without known causes of fever were preferentially selected to increase the possibility of detecting novel viral pathogens.

The demographic characteristics of study participants are described in Table 1. Of the 498 study participants whose samples were analysed, 61.5 % were female. Of all participants, 68.9 % were from Mtwapa, and 59.6 % were enrolled during the rainy season (March through June or October through December). The vast majority (86.5 %) of patients presented within 7 days of symptom onset. Prevalent HIV-1 infection was newly diagnosed at study enrolment in 4.8 % of participants. The most common symptoms of these febrile patients were generalized myalgia (i.e. muscle pain), loss of appetite, sore throat, vomiting and diarrhoea (Fig. 2).

Plasma virome

Following enrichment of viral particles-associated nucleic acid, random nucleic acid amplification, deep sequencing and bioinformatics analysis, the percentage of sequence reads belonging to different viruses was calculated for each of the 51 pools (Methods). A total number of 519 million reads were generated. The average number of reads was 10.18 million per plasma pool. The raw data for each pool is available at NCBI's Short Reads Archive under GenBank accession number SRP090133. A threshold number of sequence reads for virus detection was established using known HIV-RNA-positive samples (Methods).

In Fig. 3, the percentage of viral reads for each of the 14 different viruses detected is shown for each plasma pool. The percentage of positive plasma pools was also plotted (Fig. 4). Parvovirus B19 sequences were found in 75 % of the pools followed by, in ranking order, pegivirus C (GBV-C) (found in 67 % of pools), alpha anellovirus (59 %), gamma anellovirus (55 %), beta anellovirus (41 %), DENV-2 (16 %), HIV-1 (6 %), human herpesvirus 6 (HHV-6) (6 %), hepatitis B virus (HBV) (4 %), rotavirus (4 %), norovirus (4 %), rhinovirus C (2 %), Merkel cell polyomavirus (MCPyV; 2 %) and Kadipiro virus (2 %) (Fig. 4). The ranking was largely similar when the total overall percentage of viral reads, rather than percentage of positive pools, was calculated with parvovirus B19>pegivirus C (GBV-C)>alpha anellovirus>gamma anellovirus>DENV-2>beta anellovirus>HIV-1>HBV>rhinovirus C>Kadipiro virus>MCPyV>HHV-6>rotavirus>norovirus.

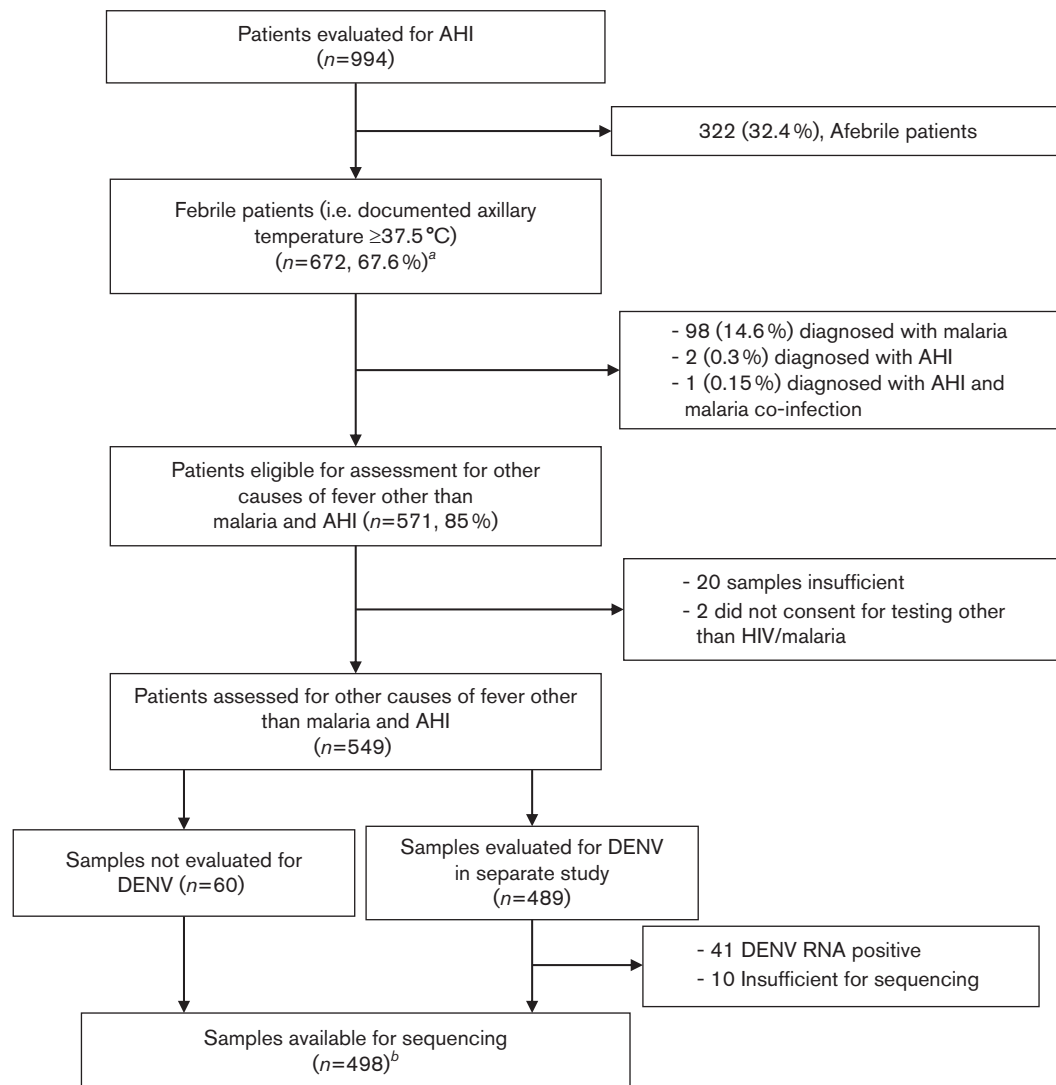
Sixty plasma samples out of 498 samples in this study had not been previously tested for DENV RNA (included in pools 4, 8, 14, 22, 23, 24, 25, 29, 33, 34, 37, 38, 39, 40, 41, 42 and 44). Given the 8.8 % detection rate in our prior study (data not shown), approximately five of these samples would be expected to be DENV-RNA-positive. Pool 19 which was positive with 0.018 % DENV reads included two samples that had been reverse transcription(RT)-PCR tested and were DENV RNA-positive. Five other pools (22, 23, 24, 37 and 41), each containing plasma samples not tested for DENV RNA, were positive with 0.17 to 0.0019 % DENV reads. Unexpectedly, pools 13 and 45 were also positive with 0.0012 and 0.0021 % DENV reads, respectively. This result may reflect cross-contamination, as the percentage of DENV reads was only slightly above the threshold set using HIV (Methods). Sequence analysis also showed that the binding site of the F1 primer used for DENV-2 RT-PCR had a mutation at the third base from the 3' end, which might have reduced its sensitivity for detection of DENV-2 RNA.

Pool 3 also showed translated protein matches to Kadipiro virus, a member of the genus *Seadornavirus* in the family *Reoviridae* previously isolated from mosquitoes in Indonesia and China (GenBank BioProject PRJNA14858). Reads and contigs with strong matches to five viral Kadipiro virus proteins (VP1–VP5) were detected using all available sequence data. These Kadipiro virus matches were 356–785 nt long, with 69–75 % nucleotide identity and 76–83 % translated peptide identity (GenBank accession numbers KU697364–KU697367, KX247778 and KX247779). A phylogenetic analysis of the available protein sequences of the VP1 and VP2 showed that the Kenyan viral genome clustered tightly with the reported Kadipiro genome (Fig. 5).

Based on best BLASTX matches, we determined that the overall number of reads to the three B19 genotypes was 97.1 % G1, 2.8 % G3, and 0.03 % G2. As the viral load of B19 can be highly variable these values do not necessarily reflect the proportion of Kenyan patients infected with these genotypes. Both strains of HBV belonged to genotype A. The three HIV strains belonged to subtypes A, D, and an A/D recombinant, and all three HHV-6 isolates belonged to subtype B.

Contamination due to the use of silica-containing columns

Other viral hits were also recorded but assigned to contamination from the use of silica columns derived from marine diatoms for nucleic acid extraction, as previously reported for a small circular DNA virus (NIH-CQV) (Naccache *et al.*, 2013, 2014; Smuts *et al.*, 2014). Because pools 1–27 were extracted with QIAamp viral RNA mini kits (Qiagen) and pools 28–51 were extracted using the magnetic beads method of the MagMax viral RNA extraction kit, we were able to compare the viral hits derived using both methods. All 1083 reads to NIH-CQV were detected only in pool 1–27. Of 3984 hits to *Iridoviridae* (a family known to infect invertebrates, fish and amphibia), all but 23 reads were found in the Qiagen



^aA total of 672 febrile patients including 30 HIV-1 seropositive patients were enrolled, of whom 24 had samples available for sequencing. The remaining 642 patients were further evaluated for AHI using p24 antigen detection.

^bTwo DENV RNA positive samples were incorrectly included among samples for sequencing.

Fig. 1. Flow chart of outpatients aged 18–35 years evaluated for viral causes of fever at the time they sought care in Kilifi or Mtwapa, February 2014–January 2015.

kit extracted pools. All but 13 of 1683 hits to *Circoviridae*/circovirus, all but 71 of 2316 hits to *Baculoviridae*, and all 305 hits to gemycircularviruses were seen exclusively in Qiagen kit extracted pools 1–27.

DISCUSSION

A viral metagenomics analysis of plasma pools from 498 febrile patients from two sites in Kenya showed that the ssDNA parvovirus B19 was found in the largest fraction of plasma pools (75%). A prior study analysing a febrile

population in southeast Nigeria did not report parvovirus B19 detection, likely due to its focus on RNA viruses (Strem-lau *et al.*, 2015). Typically transmitted by the respiratory route, parvovirus B19 can cause fever and an evanescent rash (fifth disease) in otherwise healthy children (Rogo *et al.*, 2014). Parvovirus B19 infection in adults may cause arthritis, myalgia and anaemia, especially in AIDS patients, and can decrease erythropoiesis, leading to aplastic anaemia, often in patients with sickle cell anaemia (Slavov *et al.*, 2011). When early foetal infection occurs with parvovirus B19, developmental anomalies, hydrops fetalis, and miscarriage can result (Stegmann & Carey, 2002). The frequent detection of B19

Table 1. Demographic and clinical characteristics of febrile patients aged 18–35 years seeking outpatient care between February 2014 and January 2015 in Kilifi and Mtwapa ($n=498$)

	Number of participants (%)		
	Male ($n=192$)	Female ($n=306$)	Total ($n=498$)
Age category			
18–24 years	102 (53.1)	169 (55.2)	271 (54.4)
25–30 years	50 (26.1)	89 (29.1)	139 (27.9)
30–35 years	40 (20.8)	48 (15.7)	88 (17.7)
Residence			
Kilifi	36 (18.8)	119 (38.9)	155 (31.1)
Mtwapa	156 (81.2)	187 (61.1)	343 (68.9)
Marital status			
Single	124 (64.6)	96 (31.4)	220 (44.2)
Ever married	68 (35.4)	209 (68.3)	277 (55.6)
Data missing	0 (0.0)	1 (0.3)	1 (0.2)
Education level			
No education/primary level education	74 (38.5)	202 (66.0)	276 (55.4)
Secondary/higher level	118 (61.5)	103 (33.7)	221 (44.4)
Data missing	0	1 (0.3)	1 (0.2)
Employment			
Employed	137 (71.4)	115 (37.6)	252 (50.6)
Unemployed	55 (28.6)	191 (62.4)	246 (49.4)
Seasonal presentation			
Dry season	84 (43.8)	117 (38.2)	201 (40.4)
Rainy season	108 (56.2)	189 (61.8)	297 (59.6)
Duration of illness			
Less than 7 days	166 (84.7)	270 (87.7)	436 (86.5)
More than 7 days	30 (15.3)	38 (12.3)	68 (13.5)
Prevalent HIV diagnosed at enrolment			
HIV-negative	186 (96.9)	288 (94.2)	474 (95.2)
HIV-positive	6 (3.1)	18 (5.8)	24 (4.8)

DNA in the adults tested indicates that parvovirus B19 may be an underappreciated cause of fever and other symptoms reported by the febrile Kenyan adults analysed here. Myalgia, reported by ~70 % of the Kenyan patients (Fig. 2), has been frequently reported in parvovirus B19-infected adults (Haya-kawa *et al.*, 2002; Oiwa *et al.*, 2011). Testing of an epidemiologically matched healthy control group will be required to more firmly support an association of parvovirus B19 and the symptoms reported here in febrile Kenyan adults.

The next four most commonly detected viruses in this study are usually described as commensal viruses, and have been reported in a large fraction of diverse populations. Pegivirus C (GBV-C or HPegV-1), an RNA virus in the genus *Flavivirus*, was the second most common viral nucleic acid detected in this study and was also commonly detected in a study of both healthy and febrile Nigerians (Stremlau *et al.*, 2015). In contrast, this virus was rarely detected in a study of febrile Nicaraguan children (Yozwiak *et al.*, 2012). The next most commonly detected infections in this study were the three known species of human anelloviruses, another group of ssDNA viruses. Anelloviruses were not reported in the Nigerian study, which focused on RNA viruses

(Stremlau *et al.*, 2015), but were the most commonly detected viruses in the study of Nicaraguan children (Yozwiak *et al.*, 2012). Anellovirus infection is nearly universal, acquired early in life, often results in chronic viraemia, and frequently involves multiple species and genotypes (Bernardin *et al.*, 2010). Although anellovirus infections are generally considered asymptomatic (Okamoto, 2009; Spandole *et al.*, 2015), increased levels of anelloviruses have been associated with AIDS (Li *et al.*, 2013; Thom & Petrik, 2007) and with immunosuppression in transplant patients (De Vlamincx *et al.*, 2013; Young *et al.*, 2015).

Dengue infection can induce high fever and has been repeatedly detected in East Africa starting in 1982 (Baba *et al.*, 2016; Johnson *et al.*, 1982; Mease *et al.*, 2011). Its detection in febrile Kenyan adults sampled for this study is also consistent with its detection in an outbreak reported from Mombasa, Kenya in 2013 (Ellis *et al.*, 2015).

We also detected a small number of pools (2–6 %) that were positive for other viruses that may have played a role in patients' symptoms. Of note, however, the read numbers for these viruses indicated very low plasma viral loads.

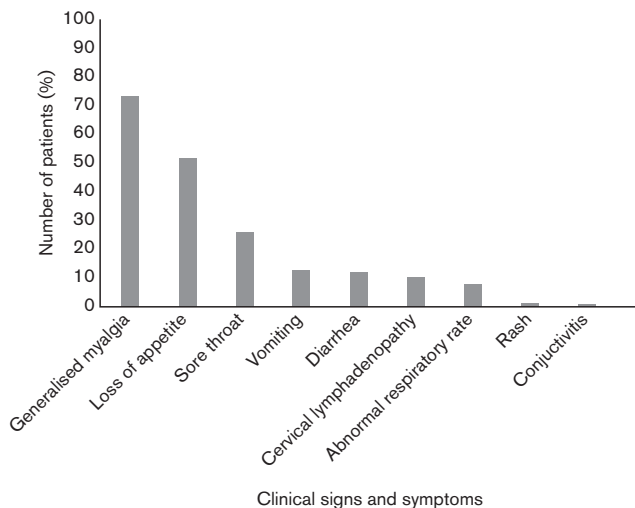


Fig. 2. Clinical signs and symptoms of febrile patients aged 18–35 years seeking outpatient care between February 2014 and January 2015 in Kilifi and Mtwapa ($n=498$).

HHV-6 subtype B and HBV subtype A viraemia were found in 6 and 4 % of pools, respectively. Enteric viruses (rotavirus in 4 % and norovirus in 4 % of pools), a respiratory virus (rhinovirus C in 2 % of pools), and a skin infection virus (MCPyV in 2 % of pools) were also detected. Detection in blood of rotavirus (Franco & Greenberg, 2013), norovirus (Frangé *et al.*, 2012; Fumian *et al.*, 2013; Newman *et al.*, 2015; Takanashi *et al.*, 2009), rhinovirus C (Esposito *et al.*, 2014; Fuji *et al.*, 2011; Lupo *et al.*, 2015) and MCPyV (Fukumoto *et al.*, 2013) have been previously reported and may reflect spill-over from ongoing high-level replication in enteric, respiratory, or epithelial target tissues.

Kadipiro virus was initially isolated from *Culex fuscocephalus* mosquitoes from Java, Indonesia in 1981 by replication in the *Aedes albopictus*-derived C6/36 cell line (Brown *et al.*, 1993) and was sequenced in 2000 (Attoui *et al.*, 2000). Using C6/36 cells, the virus was also isolated from mosquitoes from Northwest Yunnan province in China (Sun *et al.*, 2009). To date, Kadipiro virus has not been reported to be amplified in mammalian cell lines or recovered following mice inoculations. The most closely related seadornavirus currently known

Pool no.	Parvovirus B19	HPeV-1 (GBV-C)	Anellovirus (alpha)	Anellovirus (gamma)	Dengue virus 2	Anellovirus (beta)	HHV-1	HBV	Rhinovirus C	Kadipiro virus	MCPyV	HHV6	Rotavirus A	Norovirus
1	0.00	1.31E-04	5.17E-05	2.15E-05	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
2	0.00	1.40E-05	6.90E-04	2.82E-04	0.00	1.07E-04	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
3	2.16E-05	0.00	3.29E-04	2.13E-04	0.00	8.50E-05	0.00	0.00	0.00	1.43E-05	0.00	0.00	0.00	0.00
4	1.82E-04	0.00	1.41E-04	2.14E-04	0.00	1.63E-05	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
5	0.00	0.00	2.75E-04	6.43E-04	0.00	3.49E-05	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
6	0.00	0.00	5.47E-03	2.16E-03	0.00	9.80E-04	2.71E-04	0.00	0.00	0.00	0.00	0.00	0.00	0.00
7	1.73E-04	4.22E-03	6.56E-05	9.83E-05	0.00	6.82E-05	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
8	0.00	0.00	9.27E-05	3.10E-05	0.00	0.00	0.00	0.00	0.00	0.00	0.00	1.49E-06	0.00	0.00
9	8.84E-05	4.42E-03	5.95E-05	1.67E-04	0.00	3.83E-05	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
10	7.29E-05	6.89E-03	8.21E-05	1.20E-04	0.00	2.97E-05	1.72E-05	0.00	0.00	0.00	0.00	0.00	0.00	0.00
11	1.36E-04	2.46E-03	4.81E-05	5.99E-05	0.00	1.55E-05	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
12	2.78E-02	3.64E-04	2.48E-05	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
13	5.41E-05	3.73E-05	1.51E-03	1.39E-05	1.21E-05	1.33E-05	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
14	5.50E-05	4.28E-05	6.05E-05	1.57E-05	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
15	5.52E-02	7.38E-05	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
16	1.02E-03	2.77E-03	2.63E-05	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
17	2.25E-05	0.00	3.70E-04	1.52E-04	0.00	2.79E-05	3.02E-05	0.00	0.00	0.00	0.00	0.00	0.00	0.00
18	3.34E-05	1.62E-04	2.91E-04	1.52E-04	0.00	2.46E-05	0.00	0.00	3.07E-05	0.00	0.00	0.00	0.00	0.00
19	2.34E-04	1.64E-03	1.71E-04	8.06E-05	1.81E-04	1.45E-05	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
20	3.47E-05	3.76E-04	1.01E-04	2.15E-05	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
21	0.00	3.41E-04	6.02E-04	1.82E-04	0.00	3.44E-05	0.00	0.00	0.00	0.00	3.37E-06	0.00	0.00	0.00
22	0.00	6.09E-05	1.40E-05	0.00E+00	2.89E-05	0.00	0.00	1.76E-05	0.00	0.00	0.00	0.00	0.00	0.00
23	3.39E-05	0.00	3.04E-04	2.69E-04	1.86E-04	2.48E-05	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
24	3.42E-04	0.00	0.00	0.00	3.30E-04	0.00	0.00	0.00	0.00	0.00	0.00	1.31E-06	0.00	0.00
25	0.00	4.45E-04	1.72E-05	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
26	0.00	0.00	3.94E-05	9.22E-05	0.00	4.68E-05	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
27	2.69E-04	5.75E-03	0.00	5.69E-05	0.00	2.35E-05	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
28	0.00	2.83E-05	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
29	1.71E-05	4.09E-05	0.00	0.00	0.00	0.00	0.00	9.03E-05	0.00	0.00	0.00	0.00	0.00	0.00
30	0.00	3.34E-04	1.47E-04	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
31	3.36E-05	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
32	1.43E-04	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
33	4.06E-04	2.65E-05	0.00	1.76E-05	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
34	0.00	0.00	7.58E-05	2.36E-05	0.00	1.60E-05	0.00	0.00	0.00	0.00	0.00	0.00	0.00	8.42E-07
35	3.46E-03	8.38E-04	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
36	0.00	1.20E-03	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
37	1.27E-04	2.86E-04	0.00	0.00	1.74E-03	0.00	0.00	0.00	0.00	0.00	0.00	0.00	1.84E-07	0.00
38	1.89E-04	5.92E-05	4.21E-05	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
39	1.22E-04	2.37E-05	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
40	2.21E-02	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
41	1.50E-03	2.21E-05	3.47E-05	2.84E-05	1.89E-05	0.00	0.00	0.00	0.00	0.00	0.00	0.00	3.15E-06	0.00
42	8.16E-04	0.00	1.69E-04	5.67E-05	0.00	2.40E-05	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
43	1.90E-03	1.71E-04	0.00	1.93E-05	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
44	1.04E-04	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	2.37E-07	0.00	0.00
45	8.34E-04	4.39E-05	0.00	0.00	2.14E-05	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
46	2.87E-04	3.04E-05	0.00	0.00	0.00	1.61E-05	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
47	6.82E-04	1.79E-05	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
48	5.06E-04	4.70E-04	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
49	1.04E-04	2.26E-04	0.00	2.45E-05	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
50	9.37E-04	0.00	7.18E-04	4.15E-05	0.00	2.33E-05	0.00	0.00	0.00	0.00	0.00	0.00	0.00	7.91E-07
51	2.56E-02	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00

Fig. 3. Viral reads (%) for 14 viruses in 51 plasma pools. Pool 19 included two samples that were DENV RNA-positive.

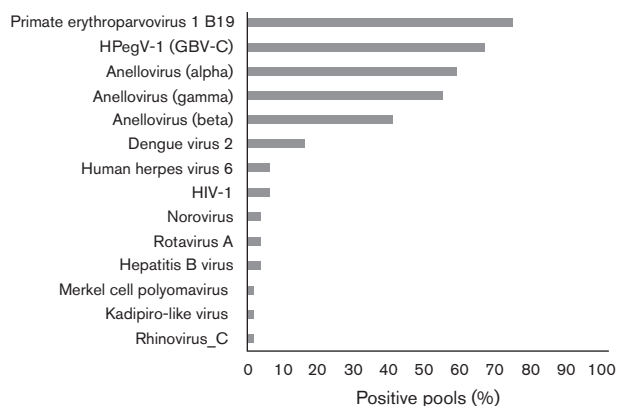


Fig. 4. Virus identification in pooled plasma sample. Percentage of positive plasma pools for each identified virus.

include Banna virus, which has been isolated from human cases of encephalitis (Liu *et al.*, 2010), as well as from pigs and cattle, and may be considered an emerging human pathogen (Attoui *et al.*, 2005; Chen & Tao, 1996). Other known seadornaviruses are Liao Ning virus (Attoui *et al.*, 2006), Balaton virus (Reuter *et al.*, 2013), and Mangshi virus (Wang *et al.*, 2015), which together with Kadipiro viruses, were all isolated from mosquitoes (Attoui *et al.*, 2005). Vectors for members of this viral genus include *Anopheles*, *Culex* and *Aedes* mosquitoes (Lv *et al.*, 2012). Based on the 2012 International Committee on Taxonomy of Viruses (ICTV) guidelines, the close similarity of the available VP1 segments of the Kadipiro viral segments from Kenya to the sequence of the Indonesian isolate indicates that it is a strain of the Kadipiro virus rather than a novel *Seadornavirus* species. Identifying Kadipiro RNA in a plasma pool indicates that its tropism includes humans.

The role of the Kadipiro virus in inducing fever and other symptoms, and its seroprevalence, will require further studies.

The presence/absence of viral sequence reads (after subtracting estimated background) was used here to estimate the number of viraemic pools. That the viral read numbers reflects the viral loads is supported by studies reporting a correlation between the percentage of viral reads and RT-PCR-estimated viral loads (Graf *et al.*, 2016; Li *et al.*, 2015; Monaco *et al.*, 2016). Viral read numbers also seem to correlate with the concentration of different types of RNA and DNA viral genomes (Li *et al.*, 2015). We used here a high number of PCR cycles to generate randomly amplified DNA for deep sequencing (see Methods). We acknowledge the possibility that our amplification method may have distorted the viral/non-viral nucleic acids ratio initially present and therefore that the percentage viral reads in Fig. 3 may not reflect relative viral loads. Our conclusions were based on the fraction of positive plasma pools, irrespective of their read numbers, rather than the percentage viral reads in the positive pools.

The major plasma viruses associated with fever have differed in the three viral metagenomics studies published to date. These have been Lassa virus [in southeast Nigerian patients of all ages (Stremlau *et al.*, 2015)], HHV-6 [in DENV-negative children from Nicaragua (Yozwiak *et al.*, 2012)], and parvovirus B19 in this study. Such differences may reflect regional or temporal variation in the major viral causes of fever other than DENV, the inclusion of patients from different age groups, or methodological differences in the generation of viral sequences. In conclusion, our study characterizing viral nucleic acids in the plasma of a febrile East African population has demonstrated a relatively high frequency of parvovirus B19 and dengue infections and revealed a novel human arbovirus, providing a baseline to

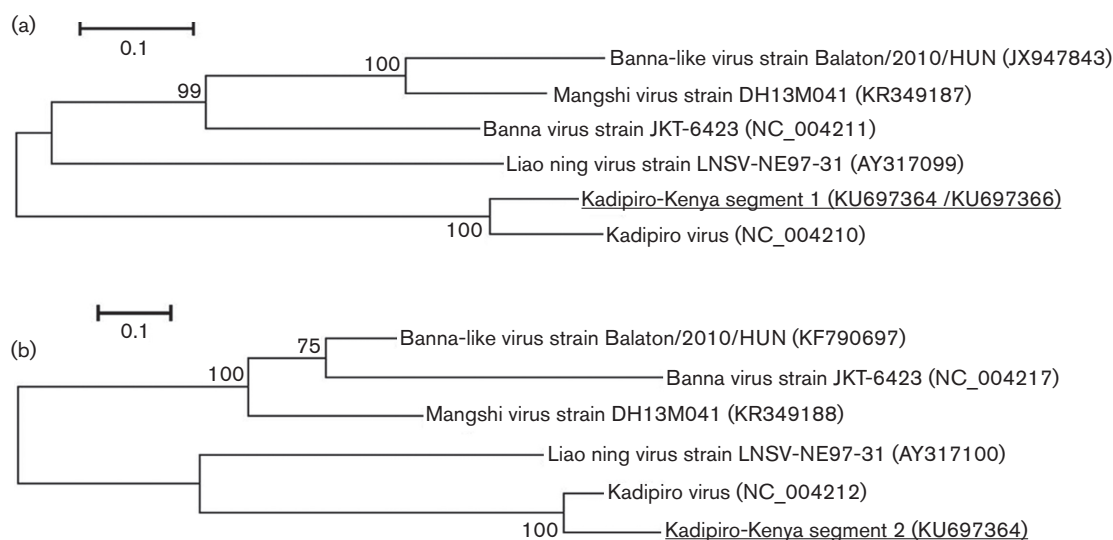


Fig. 5. Phylogenetic analysis of Kadipiro virus Kenya strain segments 1 (a) and 2 (b). Bars, 0.1 nucleotide substitutions per site.

compare with future virome studies to detect emerging viruses in this population.

METHODS

Study population. The study was conducted in two towns on the Kenyan coast: Mtwapa, a peri-urban area located 13 km north of Mombasa, and Kilifi, a small town located 43 km north of Mtwapa. Blood samples were collected between January 2014 and February 2015 from patients aged 18–35 years with HIV-1-seronegative status or unknown HIV status who enrolled in a study of AHI as previously described (Sanders *et al.*, 2014). Patients with known HIV-1-seropositive status were excluded. All samples were tested for prevalent HIV at enrolment using two rapid antibody tests in parallel, Alere Determine HIV-1/2 (Alere Medical) and Uni-Gold (Trinity Biotech). Samples from febrile patients (i.e. documented axillary temperature of $\geq 37.5^{\circ}\text{C}$) were tested for malaria using a rapid diagnostic test (Optimal; Flow Inc.). HIV-antibody-negative samples were evaluated for AHI using a p24 antigen assay (miniVidas; bioMérieux) (Sanders *et al.*, 2014). A subset of samples that were negative for both malaria and AHI were tested for DENV infection using RT-PCR (Santiago *et al.*, 2013). All samples from febrile patients that tested positive for malaria or AHI, as well as all samples that tested positive for DENV infection were excluded from the current study except for two samples that were incorrectly included (Fig. 1). Samples of patients newly diagnosed with prevalent HIV infection at enrolment were included in this study. Plasma from epidemiologically matched healthy subjects were not available as a control group since febrile patient sample collection was performed to measure HIV incidence and to initiate early anti-retroviral therapy.

Viral particle enrichment and random nucleic acid amplification. Particle-associated viral genomes were enriched, randomly amplified, and sequenced using the Illumina platform and viral sequences identified through sequence homology searches. A total of 498 plasma samples were randomly grouped into 51 pools of 8 to 11 samples. Pools were filtered using a 450 nm-pore-size filter, followed by nuclease digestion (Li *et al.*, 2015). Pools 1–27 were then extracted using QIAamp Viral RNA Mini Kit (Qiagen), while pools 28–51 were extracted using the magnetic beads method of the MagMax viral RNA extraction kit (Ambion). Kits purify both RNA and DNA. Viral nucleic acids were then amplified using a primer with a randomized 3' end. Complementary strands were generated using Klenow DNA polymerase and then further amplified by 39 cycles of PCR using primers complementary to the conserved 5' end of the random primers. This DNA was then prepared for Illumina deep sequencing using Illumina Nextera kits with dual barcoding.

Bioinformatic analysis. A total of 519.4 million reads were generated using two lanes of HiSeq 150PE reads. Sequences were binned based on dual barcode labelling, then RT-PCR primer sequences were removed and reads within each pool were *de novo* assembled using Ensemble software (Deng *et al.*, 2015). Viral read numbers were identified using BLASTX against all viral protein sequences from reference viral genome sequences in RefSeq (Deng *et al.*, 2015).

Calculation of viral read numbers and threshold determination for virus detection. The number of viral reads from each virus was determined using a cut-off sequence similarity E score of $<10^{-5}$. The number of viral reads over the total number of reads was used to calculate the percentage of reads for each virus in each pool. The number of reads from each pools ranged from 4.5×10^5 to 2.9×10^7 with an average of 1.1×10^7 reads. Rather than report as little as a single sequence read within a pool as evidence of viral presence, we established a percent of reads threshold to account for 'leakage' observed when multiple samples (here pools tagged with different barcodes) are sequenced on the same flow cell using Illumina technology. This phenomenon has been

observed in our laboratory (data not shown) and by others (Greninger *et al.*, 2015; Quail *et al.*, 2014), and may occur due to cross-contamination during sample preparation and/or misclassification due to acquiring a barcode from another sample during bridge PCR to generate DNA clusters on the same Illumina flow cell. In order to establish a stringent threshold criterion for viral detection, we took advantage of the presence of known HIV-seropositive samples included in nine plasma pools.

The percentage of HIV reads from the nine pools known to contain a total of 24 HIV prevalent samples (seropositive but p24 antigen negative) ranged from 0.027% (271 reads per million) to 0.0000035% (0.035 reads per million). The percentage of HIV reads from the 42 pools known to contain only samples that were both HIV-seronegative and p24-antigen-negative ranged from 0.00072 to 0%. The mid-point between the highest percentage HIV reads from a HIV-negative pool (0.00072%) and the next highest percentage reads from a pool containing a HIV-seropositive sample (0.0017%) was used as the threshold for viral nucleic acid detection and was calculated as 0.0012% (12 in a million reads). By this criterion, three of the nine pools known to contain HIV-seropositive p24-negative plasma were positive for HIV RNA detection. The negative results for the other six pools may be due to low HIV-1 RNA levels below levels of detection by deep sequencing or to dilution from sample pooling.

As potentially misleading barcode leakage occurs most efficiently from samples or pools with a high number of reads for a specific virus, this threshold was not applied to the four viruses for which no pools had greater than 0.0012% reads. A threshold was therefore not applied to the very low number of reads (range 0.00015 to 0.000012%) for HHV-6, MCPyV, Norwalk virus and rotavirus (found in three, one, two and two pools, respectively).

ACKNOWLEDGEMENTS

This work was supported by the Blood Systems Research Institute and National Institutes of Health R01 HL105770 to E. D. This work was partially funded by IAVI with the generous support of USAID and other donors; a full list of IAVI donors is available at www.iavi.org. The KWTRP at the Centre for Geographical Medicine Research-Kilifi is supported by core funding from the Wellcome Trust (#077092). The contents of this manuscript are the responsibility of the authors and do not necessarily reflect the views of USAID or the US Government. This manuscript was submitted for publication with permission from the Director of the Kenya Medical Research Institute (KEMRI). The study was approved by the KEMRI Scientific and Ethics Review Unit, and by the University of Oxford. All participants provided written informed consent for study participation.

REFERENCES

- Allander, T., Emerson, S. U., Engle, R. E., Purcell, R. H. & Bukh, J. (2001). A virus discovery method incorporating DNase treatment and its application to the identification of two bovine parvovirus species. *Proc Natl Acad Sci USA* **98**, 11609–11614.
- Attoui, H., Billoir, F., Biagini, P., de Micco, P. & de Lamballerie, X. (2000). Complete sequence determination and genetic analysis of Banna virus and Kadipiro virus: proposal for assignment to a new genus (*Seadornavirus*) within the family *Reoviridae*. *J Gen Virol* **81**, 1507–1515.
- Attoui, H., Mohd Jaafar, F., Belhouchet, M., Tao, S., Chen, B., Liang, G., Tesh, R. B., de Micco, P. & de Lamballerie, X. (2006). *Liao ning virus*, a new Chinese seadornavirus that replicates in transformed and embryonic mammalian cells. *J Gen Virol* **87**, 199–208.
- Attoui, H., Mohd Jaafar, F., de Micco, P. & de Lamballerie, X. (2005). Coltiviruses and seadornaviruses in North America, Europe, and Asia. *Emerg Infect Dis* **11**, 1673–1679.

- Baba, M., Villinger, J. & Masiga, D. K. (2016). Repetitive dengue outbreaks in East Africa: a proposed phased mitigation approach may reduce its impact. *Rev Med Virol* 26, 183–196.
- Bernardin, F., Operskalski, E., Busch, M. & Delwart, E. (2010). Transfusion transmission of highly prevalent commensal human viruses. *Transfusion* 50, 2474–2483.
- Breitbart, M., Salamon, P., Andresen, B., Mahaffy, J. M., Segall, A. M., Mead, D., Azam, F. & Rohwer, F. (2002). Genomic analysis of uncultured marine viral communities. *Proc Natl Acad Sci U S A* 99, 14250–14255.
- Brown, S. E., Gorman, B. M., Tesh, R. B. & Knudson, D. L. (1993). Coltiviruses isolated from mosquitoes collected in Indonesia. *Virology* 196, 363–367.
- Chen, B. & Tao, S. (1996). Arbovirus survey in China in recent ten years. *Chin Med J* 109, 13–15.
- De Vlaminc, I., Khush, K. K., Strehl, C., Kohli, B., Luikart, H., Neff, N. F., Okamoto, J., Snyder, T. M., Cornfield, D. N. & other authors (2013). Temporal response of the human virome to immunosuppression and antiviral therapy. *Cell* 155, 1178–1187.
- Deng, X., Naccache, S. N., Ng, T., Federman, S., Li, L., Chiu, C. Y. & Delwart, E. L. (2015). An ensemble strategy that significantly improves *de novo* assembly of microbial genomes from metagenomic next-generation sequencing data. *Nucleic Acids Res* 43, e46.
- Ellis, E. M., Neatherlin, J. C., Delorey, M., Ochieng, M., Mohamed, A. H., Mogeni, D. O., Hunsperger, E., Patta, S., Gikunju, S. & other authors (2015). A household serosurvey to estimate the magnitude of a dengue outbreak in Mombasa, Kenya, 2013. *PLoS Negl Trop Dis* 9, e0003733.
- Esposito, S., Daleno, C., Scala, A., Castellazzi, L., Terranova, L., Sferazza Papa, S., Longo, M. R., Pelucchi, C. & Principi, N. (2014). Impact of rhinovirus nasopharyngeal viral load and viremia on severity of respiratory infections in children. *Eur J Clin Microbiol Infect Dis* 33, 41–48.
- Franco, M. A. & Greenberg, H. B. (2013). Rotavirus. *Microbiol Spectr* 1.
- Frange, P., Touzot, F., Debré, M., Héritier, S., Leruez-Ville, M., Cros, G., Rouzioux, C., Blanche, S., Fischer, A. & Avettand-Fenoël, V. (2012). Prevalence and clinical impact of norovirus fecal shedding in children with inherited immune deficiencies. *J Infect Dis* 206, 1269–1274.
- Fuji, N., Suzuki, A., Lupisan, S., Sombrero, L., Galang, H., Kamigaki, T., Tamaki, R., Saito, M., Aniceto, R. & other authors (2011). Detection of human rhinovirus C viral genome in blood among children with severe respiratory infections in the Philippines. *PLoS One* 6, e27247.
- Fukamoto, H., Sato, Y., Hasegawa, H. & Katano, H. (2013). Frequent detection of Merkel cell polyomavirus DNA in sera of HIV-1-positive patients. *Virol J* 10, 84.
- Fumian, T. M., Justino, M. C., D'Arc Pereira Mascarenhas, J., Reyman, T. K., Abreu, E., Soares, L., Linhares, A. C. & Gabbay, Y. B. (2013). Quantitative and molecular analysis of noroviruses RNA in blood from children hospitalized for acute gastroenteritis in Belém, Brazil. *J Clin Virol* 58, 31–35.
- Graf, E. H., Simmon, K. E., Tardif, K. D., Hymas, W., Flygare, S., Eilbeck, K., Yandell, M. & Schlaberg, R. (2016). Unbiased detection of respiratory viruses by use of RNA sequencing-based metagenomics: a systematic comparison to a commercial PCR panel. *J Clin Microbiol* 54, 1000–1007.
- Greninger, A. L., Naccache, S. N., Messacar, K., Clayton, A., Yu, G., Somasekar, S., Federman, S., Stryke, D., Anderson, C. & other authors (2015). A novel outbreak enterovirus D68 strain associated with acute flaccid myelitis cases in the USA (2012–14): a retrospective cohort study. *Lancet Infect Dis* 15, 671–682.
- Hayakawa, H., Tara, M., Niina, K. & Osame, M. (2002). A clinical study of adult human parvovirus B19 infection. *Intern Med* 41, 295–299.
- Johnson, B. K., Ocheng, D., Gichogo, A., Okiro, M., Libondo, D., Kinyanjui, P. & Tukei, P. M. (1982). Epidemic dengue fever caused by dengue type 2 virus in Kenya: preliminary results of human virological and serological studies. *East Afr Med J* 59, 781–784.
- Li, L., Deng, X., Linsuwanon, P., Bangsberg, D., Bwana, M. B., Hunt, P., Martin, J. N., Deeks, S. G. & Delwart, E. (2013). AIDS alters the commensal plasma virome. *J Virol* 87, 10912–10915.
- Li, L., Deng, X., Mee, E. T., Collot-Teixeira, S., Anderson, R., Schepelmann, S., Minor, P. D. & Delwart, E. (2015). Comparing viral metagenomics methods using a highly multiplexed human viral pathogens reagent. *J Virol Methods* 213, 139–146.
- Liu, H., Li, M. H., Zhai, Y. G., Meng, W. S., Sun, X. H., Cao, Y. X., Fu, S. H., Wang, H. Y., Xu, L. H. & other authors (2010). Banna virus, China, 1987–2007. *Emerg Infect Dis* 16, 514–517.
- Lupo, J., Schuffenecker, I., Morel-Baccard, C., Bardet, J., Payen, V., Kaiser, L., Constant, S., Lobrinus, J. A., Lin-Marq, N. & other authors (2015). Disseminated rhinovirus C8 infection with infectious virus in blood and fatal outcome in a child with repeated episodes of bronchiolitis. *J Clin Microbiol* 53, 1775–1777.
- Lv, X., Mohd Jaafar, F., Sun, X., Belhouchet, M., Fu, S., Zhang, S., Tong, S. X., Lv, Z., Mertens, P. P. & other authors (2012). Isolates of Liao ning virus from wild-caught mosquitoes in the Xinjiang province of China in 2005. *PLoS One* 7, e37732.
- Mease, L. E., Coldren, R. L., Musila, L. A., Prosser, T., Ogolla, F., Ofula, V. O., Schoepp, R. J., Rossi, C. A. & Adungo, N. (2011). Seroprevalence and distribution of arboviral infections among rural Kenyan adults: a cross-sectional study. *Virol J* 8, 371.
- Monaco, C. L., Gootenberg, D. B., Zhao, G., Handley, S. A., Ghebremichael, M. S., Lim, E. S., Lankowski, A., Baldrige, M. T., Wilen, C. B. & other authors (2016). Altered virome and bacterial microbiome in human immunodeficiency virus-associated acquired immunodeficiency syndrome. *Cell Host Microbe* 19, 311–322.
- Naccache, S. N., Greninger, A. L., Lee, D., Coffey, L. L., Phan, T., Rein-Weston, A., Aronsohn, A., Hackett, J., Delwart, E. L. & Chiu, C. Y. (2013). The perils of pathogen discovery: origin of a novel parvovirus-like hybrid genome traced to nucleic acid extraction spin columns. *J Virol* 87, 11966–11977.
- Naccache, S. N., Hackett, J., Delwart, E. L. & Chiu, C. Y. (2014). Concerns over the origin of NIH-CQV, a novel virus discovered in Chinese patients with seronegative hepatitis. *Proc Natl Acad Sci U S A* 111, E976.
- Newman, K. L., Marsh, Z., Kirby, A. E., Moe, C. L. & Leon, J. S. (2015). Immunocompetent adults from human norovirus challenge studies do not exhibit norovirus viremia. *J Virol* 89, 6968–6969.
- Oiwa, H., Shimada, T., Hashimoto, M., Kawaguchi, A., Ueda, T., Sugiyama, E. & Kamiya, T. (2011). Clinical findings in parvovirus B19 infection in 30 adult patients in Kyoto. *Mod Rheumatol* 21, 24–31.
- Okamoto, H. (2009). History of discoveries and pathogenicity of TT viruses. *Curr Top Microbiol Immunol* 331, 1–20.
- Quail, M. A., Smith, M., Jackson, D., Leonard, S., Skelly, T., Swerdlow, H. P., Gu, Y. & Ellis, P. (2014). SASI-Seq: sample assurance Spike-Ins, and highly differentiating 384 barcoding for Illumina sequencing. *BMC Genomics* 15, 110.
- Reuter, G., Boros, A., Delwart, E. & Pankovics, P. (2013). Novel sea-dornavirus (family *Reoviridae*) related to Banna virus in Europe. *Arch Virol* 158, 2163–2167.
- Rogo, L. D., Mokhtari-Azad, T., Kabir, M. H. & Rezaei, F. (2014). Human parvovirus B19: a review. *Acta Virol* 58, 199–213.
- Sanders, E. J., Mugo, P., Prins, H. A., Wahome, E., Thiong'o, A. N., Mwashigadi, G., van der Elst, E. M., Omar, A., Smith, A. D. & Graham, S. M. (2014). Acute HIV-1 infection is as common as malaria in young febrile adults seeking care in coastal Kenya. *AIDS* 28, 1357–1363.

- Santiago, G. A., Vergne, E., Quiles, Y., Cosme, J., Vazquez, J., Medina, J. F., Medina, F., Colón, C., Margolis, H. & Muñoz-Jordán, J. L. (2013). Analytical and clinical performance of the CDC real time RT-PCR assay for detection and typing of dengue virus. *PLoS Negl Trop Dis* 7, e2311.
- Slavov, S. N., Kashima, S., Pinto, A. C. & Covas, D. T. (2011). Human parvovirus B19: general considerations and impact on patients with sickle-cell disease and thalassemia and on blood transfusions. *FEMS Immunol Med Microbiol* 62, 247–262.
- Smuts, H., Kew, M., Khan, A. & Korsman, S. (2014). Novel hybrid parvovirus-like virus, NIH-CQV/PHV, contaminants in silica column-based nucleic acid extraction kits. *J Virol* 88, 1398.
- Spandole, S., Cimponeriu, D., Berca, L. M. & Mihăescu, G. (2015). Human anelloviruses: an update of molecular, epidemiological and clinical aspects. *Arch Virol* 160, 893–908.
- Stegmann, B. J. & Carey, J. C. (2002). TORCH infections. Toxoplasmosis, other (syphilis, varicella-zoster, parvovirus B19), Rubella, Cytomegalovirus (CMV), and Herpes infections. *Curr Womens Health Rep* 2, 253–258.
- Stremlau, M. H., Andersen, K. G., Folarin, O. A., Grove, J. N., Odia, I., Ehiane, P. E., Omoniwa, O., Omoregie, O., Jiang, P. P. & other authors (2015). Discovery of novel rhabdoviruses in the blood of healthy individuals from West Africa. *PLoS Negl Trop Dis* 9, e0003631.
- Sun, X. H., Meng, W. S., Fu, S. H., Feng, Y., Zhai, Y. G., Wang, J. L., Wang, H. Q., Lv, X. J. & Liang, G. D. (2009). [The first report of Kadipiro virus isolation in China]. *Bing Du Xue Bao = Chinese Journal of Virology/ [Bian Ji, Bing Du Xue Bao Bian Ji Wei Yuan Hui]* 25, 173–177.
- Takanashi, S., Hashira, S., Matsunaga, T., Yoshida, A., Shiota, T., Tung, P. G., Khamrin, P., Okitsu, S., Mizuguchi, M. & other authors (2009). Detection, genetic characterization, and quantification of norovirus RNA from sera of children with gastroenteritis. *J Clin Virol* 44, 161–163.
- Thom, K. & Petrik, J. (2007). Progression towards AIDS leads to increased Torque teno virus and Torque teno minivirus titers in tissues of HIV infected individuals. *J Med Virol* 79, 1–7.
- Wang, J., Li, H., He, Y., Zhou, Y., Meng, J., Zhu, W., Chen, H., Liao, D. & Man, Y. (2015). Isolation and genetic characterization of Mangshi virus: a newly discovered *Seadornavirus* of the *Reoviridae* family found in Yunnan Province, China. *PLoS One* 10, e0143601.
- Young, J. C., Chehoud, C., Bittinger, K., Bailey, A., Diamond, J. M., Cantu, E., Haas, A. R., Abbas, A., Frye, L. & other authors (2015). Viral metagenomics reveal blooms of anelloviruses in the respiratory tract of lung transplant recipients. *Am J Transplant* 15, 200–209.
- Yozwiak, N. L., Skewes-Cox, P., Stenglein, M. D., Balmaseda, A., Harris, E. & DeRisi, J. L. (2012). Virus identification in unknown tropical febrile illness cases using deep sequencing. *PLoS Negl Trop Dis* 6, e1485.