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

Carolyne 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. Sanders1,7 and Eric Delwart2,8

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.

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...
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 belonging to different viruses was calculated for each and bioinformatics analysis, the percentage of sequence reads to NIH-CQV were detected only in pool 1

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.
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 (Strelau *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

---

**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.
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). 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 Vlaminck 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.

### 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)

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

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

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

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 (Franque 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 fusocephalus 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 saerdornavirus currently known...
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 reflect 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. 4. Virus identification in pooled plasma sample. Percentage of positive plasma pools for each identified virus.](image_url)

![Fig. 5. Phylogenetic analysis of Kadipiro virus Kenya strain segments 1 (a) and 2 (b). Bars, 0.1 nucleotide substitutions per site.](image_url)
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 (Sanderson 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 ≥37.5 °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) (Sanderson 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 samples from random pools 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. Supplementary 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 150 PE 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 pool ranged from 4.5×104 to 2.9×105 with an average of 1.1×105 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 criterion for viral detection. The percentage of reads for each virus in each pool above the 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 0.000035 % (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 (#070792). 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


