



Detection of dicistroviruses RNA in blood of febrile Tanzanian children

Samuel Cordey^{a,b,*}, Florian Laubscher^{a,b,*}, Mary-Anne Hartley^c, Thomas Junier^{d,e}, Francisco J. Pérez-Rodriguez^b, Kristina Keitel^f, Gael Vieille^{a,b}, Josephine Samaka^{g,h}, Tarsis Mlaganile^g, Frank Kagoro^g, Noémie Boillat-Blanco^{f,i}, Zainab Mbarack^j, Mylène Docquier^k, Francisco Brito^{d,l}, Daniel Eibach^{m,n}, Jürgen May^{m,n}, Peter Sothmann^{m,n,o,p}, Cassandra Aldrich^{m,p}, John Lusingu^q, Caroline Tapparel^b, Valérie D'Acremont^{c,f,*} and Laurent Kaiser^{a,b,r,*}

^aDivision of Infectious Diseases and Laboratory of Virology, University of Geneva Hospitals Geneva, Switzerland; ^bUniversity of Geneva Medical School Geneva, Switzerland; ^cDepartment of Ambulatory Care and Community Medicine, Lausanne University Hospital Lausanne, Switzerland; ^dSwiss Institute of Bioinformatics Geneva, Switzerland; ^eGlobal Health Institute, School of Life Sciences, École Polytechnique Fédérale de Lausanne Lausanne, Switzerland; ^fSwiss Tropical and Public Health Institute, University of Basel Basel, Switzerland; ^gIfakara Health Institute, Dar es Salaam, Tanzania; ^hAmana Hospital, Dar es Salaam, Tanzania; ⁱInfectious Diseases Service, Lausanne University Hospital Lausanne, Switzerland; ^jMwananyamala Hospital, Dar es Salaam, Tanzania; ^kiGE3 Genomics Platform, University of Geneva Geneva, Switzerland; ^lDepartment of Genetic Medicine and Development, Faculty of Medicine of Geneva Geneva, Switzerland; ^mDepartment of Infectious Disease Epidemiology, Bernhard Nocht Institute for Tropical Medicine Hamburg, Germany; ⁿGerman Centre for Infection Research (DZIF), Hamburg, Germany; ^oDivision of Tropical Medicine, 1st Department of Medicine, University Medical Center Hamburg-Eppendorf Hamburg, Germany; ^pDivision of Infectious Diseases and Tropical Medicine, Medical Center of the University of Munich (LMU) Munich, Germany; ^qNational Institute for Medical Research, Tanga Research Centre, Tanga, Tanzania; ^rGeneva Centre for Emerging Viral Diseases Geneva, Switzerland

ABSTRACT

Fever is the leading cause of paediatric outpatient consultations in Sub-Saharan Africa. Although most are suspected to be of viral origin, a putative causative pathogen is not identified in over a quarter of these febrile episodes. Using a de novo assembly sequencing approach, we report the detection (15.4%) of dicistroviruses (DicV) RNA in sera collected from 692 febrile Tanzanian children. In contrast, DicV RNA was only detected in 1/77 (1.3%) plasma samples from febrile Tanzanian adults, suggesting that children could represent the primary susceptible population. Estimated viral load by specific quantitative real-time RT-PCR assay ranged from $< 1.32\text{E}3$ to $1.44\text{E}7$ viral RNA copies/mL serum. Three DicV full-length genomes were obtained, and a phylogenetic analysis on the capsid region showed the presence of two clusters representing tentative novel genus. Although DicV-positive cases were detected throughout the year, a significantly higher positivity rate was observed during the rainy season.

This study reveals that novel DicV RNA is frequently detected in the blood of Tanzanian children, paving the way for further investigations to determine if DicV possibly represent a new agent in humans.

ARTICLE HISTORY Received 5 December 2018; Revised 26 February 2019; Accepted 31 March 2019

KEYWORDS Dicistrovirus; de novo analysis; sera; viremia; Tanzanian children

Introduction

Infectious febrile illnesses are leading causes of paediatric outpatient consultations in Sub-Saharan Africa. For a large proportion of these cases, the microbial aetiology remains unknown even after extensive diagnostic panels have been applied; thus, it is suspected that new or divergent viruses that are not detectable by standard targeted diagnostics assay may be involved. High-throughput sequencing (HTS) coupled with de novo assembly offers the possibility of recognizing viral RNA or DNA in an unbiased fashion. HTS-based screenings in both vertebrate and non-vertebrate

biological specimens [1–3] have shown that the diversity of viruses circulating in the animal world is larger than anticipated (including potential human infections such as astroviruses, reoviruses, pseudorabies viruses, and bornaviruses [4–9]), and that many viral families are shared by the two kingdoms, raising the possibility of yet to be discovered viral infections in humans [10].

Using HTS and a de novo assembly approach, we screened sera collected from febrile Tanzanian children for the presence of unrecognized viral agents. These children were enrolled in a large cohort investigating the clinical impact of point-of-care patient management

CONTACT Samuel Cordey samuel.cordey@hcuge.ch Division of Infectious Diseases and Laboratory of Virology, University of Geneva Hospitals, 1211 Geneva, Switzerland University of Geneva Medical School, 1211 Geneva, Switzerland

*Equal contribution

**These senior authors contributed equally to this article

Supplemental data for this article can be accessed at <https://doi.org/10.1080/22221751.2019.1603791>

© 2019 The Author(s). Published by Informa UK Limited, trading as Taylor & Francis Group, on behalf of Shanghai Shangyixun Cultural Communication Co., Ltd
This is an Open Access article distributed under the terms of the Creative Commons Attribution License (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

system featuring an electronic algorithm based on host biomarkers [11]. After identifying an index case harbouring a full-length sequence of dicistrovirus (DicV), we used the complete original sequence obtained to extend screening across the 692 paediatric patients enrolled in this cohort, as well as 77 febrile adults included in a similar study [12]. *Dicistroviridae* are non-enveloped, positive-sense single-stranded RNA viruses belonging to the picorna-like viruses. DicV infect a wide spectrum of arthropods including species that are known to feed on human blood [13].

This report describes the detection of DicV RNA, including full-length genomes, in sera from Tanzanian children with acute febrile illnesses. The presence of DicV RNA was further confirmed by specific quantitative real-time RT-PCR assays as well as conventional sequencing. We also provide a description of epidemiological, demographic and clinical features associated with the presence of DicV sequences and propose strategies for further investigations to explore the role of this virus.

Materials and Methods

Study design

A total of 692 serum samples were selected from a larger paediatric cohort of 3192 patients (2 months to 5 years of age). Participants were consecutively recruited between December 2014 to February 2016 across nine outpatient clinics in Dar es Salaam (Tanzania) according to the inclusion criteria of “acute febrile illness” [11]. The 692 serum samples were randomized from a sub-cohort of patients with the diagnoses of “fever without source”, “severe febrile illness” or “malaria” as detected by PCR (Figure 1). This sub-cohort represents a group of particular clinical interest for expanded diagnostic screening. As sample randomization in children was limited by sample volume availability, we analysed the selection for potential bias. Selected samples were not significantly different for all key demographic variables (such as age and sex) nor for geographic (district) or temporal (season) features. In addition, 77 plasma specimens were randomly selected from an adult cohort of 519 patients. Adult patients were consecutively recruited between July 2013 to May 2014 across outpatient clinics in Dar es Salaam according to the inclusion criteria of “acute febrile illness” [12]. Samples were then randomly selected from a sub-cohort of interest with the same 3 diagnoses as above (Figure 1).

Ethical approval

This study has received local IRB approval in Tanzania by the Ifakara Health Institute and the National Institute for Medical research (paediatric cohort: IHI/IRB/EXT/16-2015 and NIMR/HQ/R.8a/Vol. IX/1789; adult cohort: IHI/IRB/No: 12-2013 and NIMR/HQ/R.8a/

Vol. IX/1561). IRB approval was also granted for analyses undertaken in Switzerland by the Ethikkommission Nordwest- und Zentralschweiz (paediatric cohort: EKNZ UBE-15/03 and adult cohort: EK 1612/13).

De novo analysis and sequence analysis

Unbiased HTS protocols was performed on the 692 sera and 77 plasma specimens from the paediatric and adult cohorts, respectively. Using a de novo assembly approach (see below), an index case harbouring a full-length sequence of DicV was detected. This complete DicV sequence was used to extend screening across all specimens analysed by HTS.

The sample preparation for HTS analysis (i.e. specimen centrifugation, DNase treatment and RNA extraction by TRIzol) was performed as previously published [14]. RNA libraries (TruSeq total RNA preparation protocol (Illumina, San Diego, US)) were multiplexed by 4 and run on a HiSeq 2500 (Illumina, San Diego, CA, USA). Raw data were low-quality and adaptor-trimmed using Trimmomatic (V0.33), human- sequence subtracted (GRCh38, gencode.V23) using snap-aligner (v1.0beta.23) and then analysed using the de novo genome assembly IDBA-UD software (v.1.1.3) [15]. Reads from raw data were mapped to complete-coding genomes obtained in previous steps. Cross-lane contamination was checked using a crosstalk rate of 0.3% [16]. Cross-lane contamination was detected for 22 paediatric samples which were subsequently excluded. The remaining 670 samples from individual patients were retained for further epidemiologic, demographic and clinical investigations (Table 1). The median of total number of read pairs obtained were 45'386'721 and 46'177'840 for DicV- and DicV+ patients, respectively, with a minimum of 15'760'060 read pairs obtained for DicV+ patients.

To check potential reagent contaminants, 38 negative controls (DNase-free, RNase-free H₂O) submitted to the whole HTS process were tested by de novo analysis (Figure 1).

After removal of PCR duplicates, only sequences with a coverage $\geq 5x$ for each individual nucleotide were considered for further analysis (Figure 2a-c). Three complete coding sequences were obtained and submitted to GenBank (accession no. MH536109-11). Phylogenetic trees were inferred for the RNA-dependent RNA polymerase (RdRp) (Figure 2b) and the capsid regions (Figure 2c) using a 500-replicate Maximum Likelihood method based on the LG model [17]. Analyses were performed with a Gamma distributed rate model with invariant sites using MEGA X software [18].

Specific dicistrovirus real-Time RT-PCR

The aligned DicV sequences obtained from de novo analysis were used to design two specific real-time

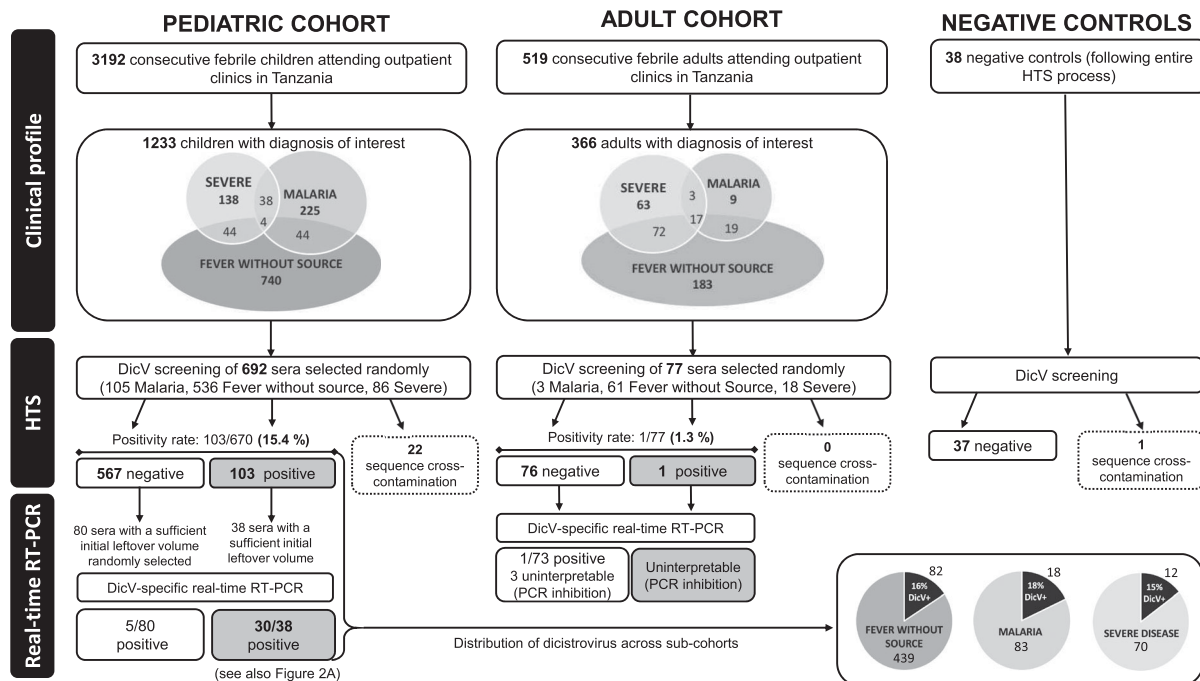


Figure 1. Study flow chart. Severe febrile illness is defined as having severe tachypnea (respiratory rate ≥ 97 th percentile for age and temperature), severe tachycardia (heart rate ≥ 90 th percentile for age and temperature), Hypoxemia ($\text{SaO}_2 < 90\%$), inability to tolerate oral fluids and/or neurological danger signs (such as coma or >2 convulsions in 24 h). The diagnosis of “fever without source” is attributed when the probable cause of a febrile illness cannot be determined by history or physical examination. The diagnosis of “malaria” is attributed in case of positive malaria-specific quantitative PCR, targeting multiple genomic copies of the conserved C-terminus of the var gene family (detection level: 0.1 parasite/ μL). HTS: high-throughput sequencing, DicV: dicistrovirus, RT-PCR: reverse transcription PCR.

RT-PCR assays (Table 2): the quantitative DicV-Capsid and the semi-quantitative DicV-ORF1/RdRp. Samples in which DicV was identified by HTS analysis and for which sufficient volume remained ($n = 38/103$) were checked by these real-time RT-PCR assays. Briefly, the viral genomes were extracted using the NucliSENS easyMAG (bioMérieux, Geneva, Switzerland) nucleic acid kit from 190 μL of sera spiked with 10 μL of standardized canine distemper viruses, as previously described [19]. Nucleic acids were eluted in 25 μL and then screened for the presence of DicV using the DicV-Capsid real-time quantitative assay (upper and lower limits of quantification are $1.32\text{E}8$ and $1.32\text{E}3$ estimated viral RNA copies/mL, respectively) and the DicV-ORF1/RdRp assay. Both assays were performed using the one-step QuantiTect Probe RT-PCR Kit (Qiagen, Hombrechtikon, Switzerland) in a StepOne Plus instrument (Applied Biosystems, Rotkreuz, Switzerland) under the following cycling conditions: 50°C for 30 min, 95°C for 15 min, 45 cycles of 15 s at 94°C , and 1 min at 55°C .

Additional Tanzanian cohort screened for DicV by real-Time RT-PCR

The study was designed to assess the role of rickettsial infection as a source of fever amongst paediatric and adult outpatients in North-Eastern Tanzania. The study was based at the National Institute for Medical

Research’s (NIMR) field station in Korogwe, Tanga Region. The field station is associated with Korogwe District Hospital (KDH).

Between 1 December 2017 and 30 April 2018, patients of five years and above presenting to the Outpatient Department of KDH with fever ($>37.5^\circ\text{C}$ tympanic) or a reported history of fever within the last three days were recruited. Blood was taken on day 0 and day 28 to establish a serological diagnosis of acute rickettsial infection. Amongst 336 patients recruited for the study, 271 (81%) completed the follow up on day 28. Amongst those patients with paired sera available (at least 500 μL per serum), 100 individuals (range from 6 to 80 years of age, median age = 27 years) were randomly selected for DicV screening by the DicV-Capsid and the DicV-ORF1/RdRp real-time RT-PCR assays. Among the 100 screened patients, 60 were > 16 years of age (median age = 45 years of age) and 40 were between 6 and ≤ 16 years of age (median age = 10 years of age). The study was approved by the National Ethics Review Committee of the National Institute for Medical Research, Tanzania.

Respiratory and dicistrovirus detection in respiratory specimens

Of the 21 paediatric patients with respiratory symptoms (Table 1) a nasopharyngeal swab specimen (NPS) was collected for 16 subjects (3 DicV+ and 13

Table 1. Comparative analysis between dicistrovirus-positive and negative febrile children.

	DicV- (n=567)			DicV+ (n=103)				
DEMOGRAPHICS	Mean	CI95%		Mean	CI95%		Difference (mean)	p-value
Age (months)	18.9	(17.9, 20.1)		16.4	(13.8, 18.9)		<div><div></div></div> -2.5	0.02 *
	%	n	total	%	n	total	Difference (%)	p-value
Sex (male)	55.0	312	567	61.2	63	103	6.1 <div><div></div></div>	0.25
EPIDEMIOLOGY	%	n	total	%	n	total	Difference (%)	p-value
Wet season (Minor rains)	17.5	99	567	45.6	47	103	28.2 <div><div></div></div>	<0.001
Wet season (Major rains)	29.8	169		21.4	22		-8.4 <div><div></div></div>	
Dry season	52.7	299		33.0	34		-19.7 <div><div></div></div>	
District (Ilala)†	7.4	40	537	26.8	26	97	19.4 <div><div></div></div>	<0.001 *
District (Temeke)†	24.4	131		20.6	20		-3.8 <div><div></div></div>	
District (Kinondoni)†	68.2	366		52.6	51		-15.6 <div><div></div></div>	
SIGNS AND SYMPTOMS	Mean	CI95%		Mean	CI95%		Difference (mean)	p-value
Axillary temperature (°C)	38.4	(38.4, 38.5)		38.3	(38.2, 38.5)		<div><div></div></div> -0.1	0.26
Days of fever before presentation	1.4	(1.3, 1.5)		1.6	(1.4, 1.8)		0.2 <div><div></div></div>	0.14
	%	n	total	%	n	total	Difference (%)	p-value
Fever as only complaint	37.9	215	567	42.7	44	103	4.8 <div><div></div></div>	0.36
Respiratory distress	2.6	15	567	5.8	6	103	3.2 <div><div></div></div>	0.08
Vomiting	13.6	77	567	16.5	17	103	2.9 <div><div></div></div>	0.43
Ear problem‡	2.1	6	286	4.9	2	41	2.8 <div><div></div></div>	0.28
Diarrhea	1.8	10	567	3.9	4	103	2.1 <div><div></div></div>	0.17
Eye problem	1.1	6	567	1.0	1	103	<div><div></div></div> -0.1	0.94
CNS danger signs	1.6	9	567	1.0	1	103	<div><div></div></div> -0.6	0.72
Danger signs	2.5	14	567	1.9	2	103	<div><div></div></div> -0.5	0.95
Cough	10.4	59	567	7.8	8	103	<div><div></div></div> -2.6	0.41
Any skin problem (excluding abscess)	4.8	27	567	1.0	1	103	<div><div></div></div> -3.8	0.08
INFLAMMATORY MARKERS	Mean	CI95%		Mean	CI95%		Difference (mean)	p-value
PCT level (ug/L) #	4.9	(2.2, 7.7)		9.0	(1.9, 16.2)		4.1 <div><div></div></div>	0.04 *
CRP level (mg/L)	9.0	(6.8, 11.2)		11.1	(5.7, 16.7)		2.1 <div><div></div></div>	0.80
	%	n	total	%	n	total	Difference (%)	p-value
PCT (> 0.5ug/L) # §	29.0	67	231	46.7	28	60	17.7 <div><div></div></div>	0.01 *
CRP > 20 mg/L §	11.8	28	238	13.6	8	59	1.8 <div><div></div></div>	0.71
CO-INFECTIONS	%	n	total	%	n	total	Difference (%)	p-value
Positive malaria rapid diagnostic test	9.9	56	565	15.2	15	99	5.2 <div><div></div></div>	0.12
Positive malaria PCR	14.7	83	565	18.2	18	99	3.5 <div><div></div></div>	0.37
Positive urine dipstick §	13.8	47	341	16.9	13	77	3.1 <div><div></div></div>	0.48
Positive typhoid rapid diagnostic test	0.2	1	567	0.0	0	103	<div><div></div></div> -0.2	0.67
Positive HIV (serology)	2.6	14	534	1.0	1	98	<div><div></div></div> -1.6	0.34
CO-MORBIDITIES	%	n	total	%	n	total	Difference (%)	p-value
Anemia	28.2	160	567	34.0	35	103	5.8 <div><div></div></div>	0.24
Malnutrition	10.4	59	567	11.7	12	103	1.2 <div><div></div></div>	0.71
Sickle cell disease (homozygote) §	1.5	8	531	2.2	2	93	0.6 <div><div></div></div>	0.84
Sickle cell trait (heterozygote) §	14.3	76		12.9	12		-1.4 <div><div></div></div>	
CLINICAL SYNDROMES & OUTCOMES	%	n	total	%	n	total	Difference (%)	p-value
Fever without source (excl. malaria)	84.0	405	482	90.1	73	81	6.1 <div><div></div></div>	0.16
Cured by day 7	96.6	546	565	100.0	103	103	3.4 <div><div></div></div>	0.06
Secondary admission to hospital	1.6	9	567	1.9	2	103	0.4 <div><div></div></div>	0.80
Death by day 30	0.5	3	567	0.0	0	103	<div><div></div></div> -0.5	0.46
Severe febrile illness	12.3	70	567	11.7	12	103	<div><div></div></div> -0.7	0.84

KEY

More prevalent in DicV+

More prevalent in DicV-

* $p < 0.05$, CRP: C-reactive protein, PCT: procalcitonin, HIV: human immunodeficiency virus

† 36 cases with an unknown district excluded

‡ data on this symptom was only collected from children old enough to express the symptom/respond to the question

§ data limited by sample availability (only collected when clinically appropriate or insufficient sample volume)

PCT is an inflammatory marker, values larger than 0.5ug/ml have been associated to a higher risk of bacterial infection

DicV- by HTS). The 16 NPS were extracted using the NucliSENS easyMAG (bioMérieux) and screened for the presence of DicV (using the DicV-Capsid and the DicV-ORF1/RdRp real-time RT-PCR assays) and a large panel of respiratory viruses using the FTD® Respiratory pathogens 21 commercial kit (Fast Track Diagnostics, Sliema, Malta).

Statistical analysis

Statistical analysis compared various demographic (age, sex), epidemiological (seasonality, location) and clinical (signs, symptoms, co-morbidities and outcomes) parameters between the HTS DicV- ($n = 567$) and DicV+ ($n = 103$) paediatric sera samples. Analyses were undertaken using Stata (StataCorp. 2015, College Station, TX, USA) and associations between the variables of interest and the presence of DicV sequences were assessed using chi-squared, Spearman correlation and Kruskal-Wallis tests, as appropriate. Results are presented as means with 95% confidence intervals, frequencies and percentages. A two-sided P -value of <0.05 was considered statistically significant.

Results

Dicistrovirus identification and screening

Of the 670 sera (from 670 individual paediatric subjects), DicV genome sequences were detected in 103 samples (15.4%) (Figure 1; total number of DicV mapping reads ranging from 2 to 121'610, mean = 3'379, median = 240). To confirm the HTS analysis findings, all positive sera with sufficient leftover volume ($n = 38$) were assessed with the quantitative DicV-Capsid and the DicV-ORF1/RdRp real-time RT-PCR assays. Thirty/38 (79%) sera were confirmed DicV+ by real-time RT-PCR (21/30 by the two assays, 9/30 by one assay). The corresponding numbers of DicV-specific reads ranged from 6 to 121'610 for real-time RT-PCR positive sera. Estimated viral load ranged from $< 1.32\text{E}3$ up to $1.44\text{E}7$ viral RNA copies/mL of sera (median viral load = $5.67\text{E}3$ viral RNA copies/mL). The 8 sera not confirmed by real-time RT-PCRs had ≤ 280 DicV-specific reads (median = 16 reads) obtained by the HTS analysis. Of these 8 sera, 7 had a sufficient initial volume for further testing and prepared in two dilutions (1:2 and 1:10) to remove potential PCR inhibitors. None was found positive by real-time RT-PCR. In parallel, 80 sera found DicV-negative by HTS analysis were randomly selected throughout the entire recruitment period of 2015 and assessed by the DicV-Capsid and DicV-ORF1/RdRp real-time RT-PCR assays; 5 were found positive (4 sera collected in February and 1 in July; viral load ranged from $< 1.32\text{E}3$ to $2.11\text{E}4$ viral RNA copies/mL of

sera) suggesting that the DicV prevalence of 15.4% observed by HTS represents a minimal estimate.

In contrast to the HTS analysis results obtained with the paediatric cohort, the presence of DicV RNA was detected in only one of 77 (1.3%) plasma specimens from the adult cohort (2 reads, real-time RT-PCRs were performed but the results were not conclusive due to the presence of a very strong PCR inhibition) (Figure 1). In parallel, the 76 plasma found DicV-negative by HTS analysis were assessed by the DicV-Capsid and DicV-ORF1/RdRp real-time RT-PCR assays; 1 was found positive (collected in November 2013; viral load was $< 1.32\text{E}3$ viral RNA copies/mL of plasma). Of note, among the 38 negative controls submitted to the whole HTS process, none was found positive for DicV.

Dicistrovirus sequence analysis

The phylogenetic trees based on the RdRp and the capsid regions (Figure 2b,c) show that the Tanzanian DicV sequences share the closest amino acid identities with the Hubei picorna-like virus 22 (NC_033227.1) for RdRp (53% identity) and capsid (32% identity) proteins. Based on the divergence of sequences in the capsid region, the three Tanzania sequences (MH536109-11) indicated the presence of two clusters representing tentative novel genus (Figure 2c). Indeed, in the capsid region, an amino acid identity of 95.2% is observed between the two clusters, while an amino acid identity of 99.46% is present in the RdRp region.

Demographic and epidemiological analysis

A comparative analysis of demographic and clinical features between the DicV+ and DicV- paediatric populations are listed in Table 1. While there was a similar sex ratio between DicV+ and DicV- groups (61% vs 55%, $p = 0.25$), DicV+ patients had a lower mean age compared to the DicV- group (16.4 vs 18.9 months, $p = 0.02$).

A monthly distribution analysis shows significantly higher prevalence of positive cases during the wet season (minor rain period) than during the dry season ($p < 0.001$, Table 1). Indeed, 45.6% of DicV infections occurred in the minor rain period (47/103, Figure 3). Significant geographic clustering occurred across three districts in Dar es Salaam ($p > 0.001$). Whilst the Kinondoni and Temeke districts had a 12 and 13% DicV prevalence, respectively (51/417 and 20/151, respectively), infections were 3-fold more frequent in the Ilala district (39%, 26/66).

Clinical analysis

Overall, DicV presence did not vary significantly among the 3 major sub-cohorts of fever without source

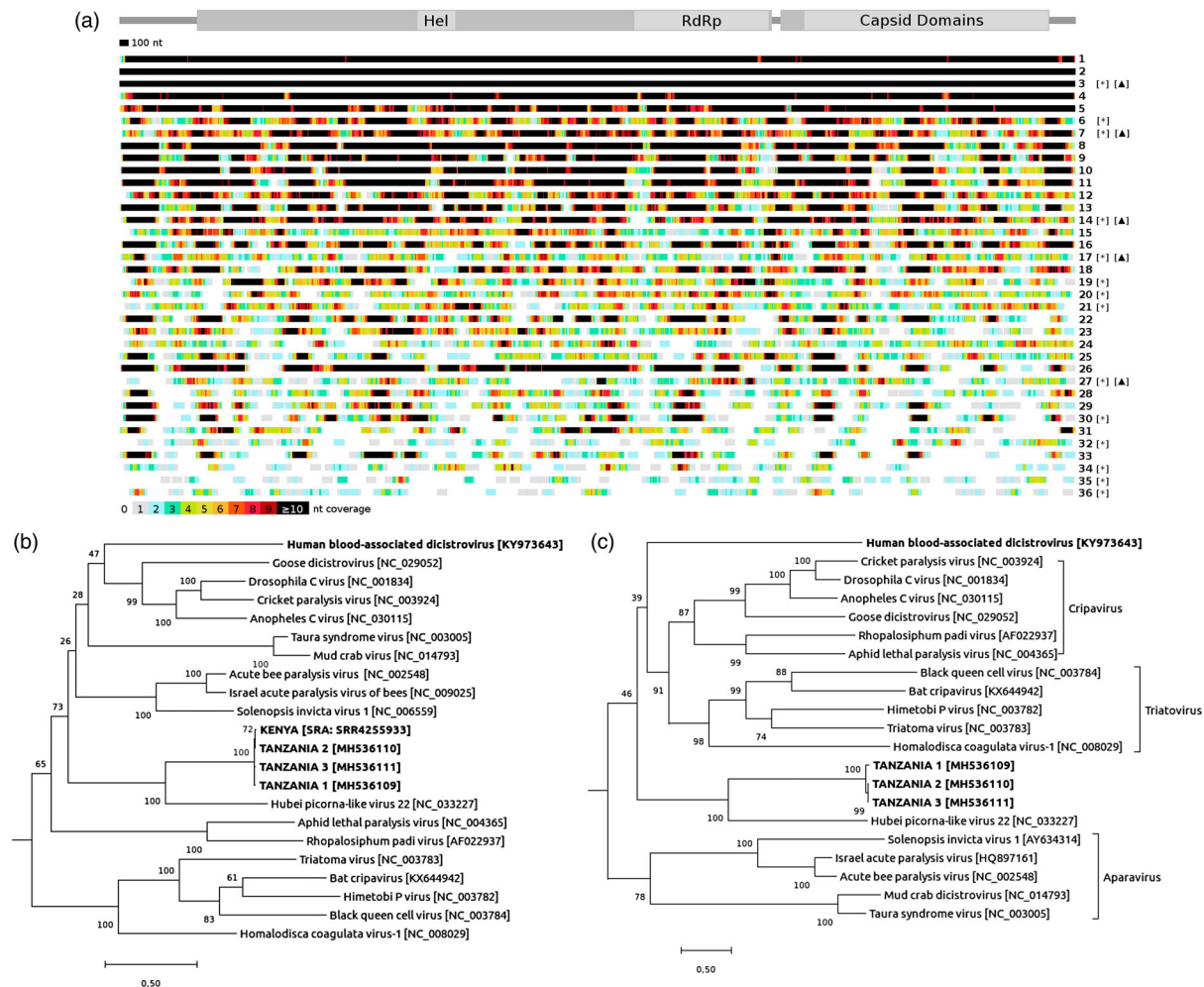


Figure 2. Analysis of Dicistrovirus sequences. (a) Location of DicV sequences with more than fifty percent DicV genome coverage obtained by de novo analysis for positive subjects. A schematic representation of the major ORFs of the DicV genome is featured at the top of the figure (Hel: helicase domain [pfam00910]; RdRp: RNA-dependent RNA polymerase domain [pfam00680]; Capsid domains include two Picornavirus capsid-like domains [cd00205], a DicV VP4 domain and CRPV capsid protein like domain [pfam08762]). On the right side is indicated when the presence of DicV could be verified by real-time RT-PCR (*) in sera with sufficient leftover volume and/or by RT-PCR followed by classical sequencing (▲). (b) Phylogenetic tree of the RdRp region of DicV. The 370-residue amino acid sequence considered for RdRp alignment is between position 1435 and 1804 of ORF1. Dicistrovirus sequences found in human blood are in bold font; the three complete coding sequences obtained in this study (TANZANIA 1–3), the Human blood-associated dicistrovirus sequence reported from a Peruvian patients with fever of unknown aetiology [21], and the KENYA sequence detected by our retrospective analysis in one pool (SRR4255933) from a study that investigated the plasma virome of febrile adult Kenyans [22]. (c) Phylogenetic tree of the capsid region of DicV. The 791-residue amino acid sequence considered for capsid alignment is between position 73 and the terminus of ORF2. Dicistrovirus sequences found in human blood are in bold font; the three complete coding sequences obtained in this study (TANZANIA 1–3), and the Human blood-associated dicistrovirus sequence reported from a Peruvian patients with fever of unknown aetiology [21].

Table 2. Primer and probe sequences. Fwd: primer forward, Rev: primer reverse, AT: PCR annealing temperature, NT: nucleotide.

				Final [uM] Fwd/Rev/Probe	
Assay			Sequence (5' to 3')		
Real-time RT-PCR	DicV-Capsid	Fwd	GGGATGCAYGCTCTTGTA AATT	0.9/0.9/0.25	
		Rev	CRCACCAACGAAACCA		
		Probe	FAM - ATGTCAGGCATTGAAC – MGBNFQ		
	DicV-ORF1/RdRp	Fwd	AGCTGGCGATTCTCAAATTATG	0.9/0.9/0.25	
		Rev	TCGTCGTA AAAATTCGTTGATTATGTC		
		Probe	FAM - TGGTACATTGCATCCCGAYATCTTATTGG - TAMRA		
Target region			Sequence (5' to 3')	Product size (nt)	AT (°C)
Classic RT-PCR	ORF1	Fwd1	TGAAGATTTCGAAGGAAAAATGG	427	50
		Rev1	ACGCGTTCTATTTATTCCTCGA		
		Fwd2	ATGATTCCAATACCGGATGTTGAGA	380	50
		Rev2	ATTCTCGAATAAATTGTGAACACG		

(16%), malaria (18%) and severe febrile illness (15%) (Figure 1).

Clinical features such as axillary temperature, cough, respiratory distress, vomiting, diarrhoea, dysuria, any ear or eye problems, dermatological symptoms, and clinical danger signs were not significantly different between DicV+ and DicV- paediatric patients (Table 1). The presence of DicV was not associated with other concomitant infections such as malaria, HIV or urinary tract infections, nor was it significantly different in patients with common co-morbidities such as malnutrition, anaemia or sickle cell disease. Outcomes and severity measures were also not affected. While the inflammatory marker C-reactive protein (CRP) was not significantly different among the groups, procalcitonin (PCT) was significantly higher in the presence of DicV (9ug/ml vs 4.9ug/ml, $p = 0.04$), where a level of $>0.5\text{ug/L}$ was 18% more frequent in DicV+ patients ($p = 0.01$).

The temporal distribution of diseases representing the major modes of infectious transmission was compared to DicV prevalence (Figure 4). The analysis revealed a temporal symmetry between the predicted probabilities of DicV infection and respiratory disease, which both peaked during the wet season. This is in contrast to both mosquito-vectored disease (malaria) and diseases usually transmitted via the fecal-oral route (diarrheal diseases) which peak in the post-flood and dry season. A total of 21 paediatric patients presented with respiratory distress (Table 1). For 16 subjects (13 DicV- and 3 DicV+ by HTS) a nasopharyngeal swab (NPS) sample was collected and screened by real-time (RT)-PCR for DicV and a large panel of respiratory viruses. Of the 13 DicV- patients, 11 were found positive for a respiratory virus in their respective NPS samples (three human parainfluenza virus type 3, three human rhinovirus, two human parainfluenza virus type 2, two human coronavirus HKU1 and one human parechovirus; no viral co-infections were observed). Of the three DicV+ patients, one respiratory virus was detected (human rhinovirus). Regarding DicV screening, none of the 13 DicV- patients had DicV sequences detectable in their NPS samples. In contrast, one NPS of the three available from the DicV+ patients was also found positive for DicV sequences (low viral load, no viral co-infection observed). This patient had DicV sequences detectable in serum by HTS which was confirmed by specific real-time RT-PCR assay (with an estimated viral load of $1.2\text{E}5$ viral RNA copies/mL of serum).

Discussion

Using an unbiased HTS approach and de novo assembly analysis, we identified distinct novel DicV sequences in 15.4% of paediatric patients attending an outpatient clinics for a febrile illness. Most DicV+

sera contained a large number of reads and a significant proportion of the DicV genomes was thus frequently obtained (Figure 2a), including full-length genomes in 3 cases. These HTS-based findings were subsequently verified by two specific in-house real-time RT-PCR assays in the 38 samples which had sufficient leftover sera (Figure 1) and by conventional sequencing in 5 samples selected for their high viral load (100% nucleotide identity, data not shown). Positivity in 30 of 38 (79%) cases could be confirmed by real-time RT-PCR with a median of $5.7\text{E}3$ viral RNA copies/mL of sera and, in some cases, reached up to $1\text{xE}7$ viral RNA copies/mL, supporting a significant viral load. In contrast to children, only one of 77 adults was found to be DicV-positive by HTS from a parallel cohort, suggesting that children could represent the predominant susceptible population. However, this latter observation should be interpreted with caution since a potential temporal bias cannot be ruled out (sampling times for the paediatric and adult cohorts are not overlapping). In order to expand our screening we tested by real-time RT-PCR 100 paired sera from adult and paediatric patients enrolled in a North-Eastern Tanzania cohort (Tanga region) between December 2017 and end of April 2018; we did not identify any DicV positive cases. However, not only a restricted number of adult ($n = 60$) and paediatric ($n = 40$) cases could be tested in this later cohort but also the paediatric populations studied in the Tanga and Dar es Salaam cohorts are different (e.g. age of enrolment) and limit our conclusions.

We also observed a clear seasonal distribution with clustering in the minor rain period of the wet season while few cases observed during the dry season (Figure 3). Taken together, these observations point toward the potential identification of a previously uncharacterized viral agent in human blood. This represents only a first step since we could unfortunately not have access to follow-up samples or any other validation cohort. According to Lipkin proposal concerning the levels of certainty in pathogen discovery, our finding fulfil the Level 1 criteria [20].

Whatever the molecular technology used, the discovery of new viral genomic sequences in body fluids must first exclude the possibility of contamination and/or spurious bioinformatics results particularly when HTS is used. In the present study, contamination has been carefully considered and is considered unlikely for the following reasons: Firstly, none of the 38 HTS negative controls that followed the entire HTS procedure revealed any DicV sequences, excluding the possibility of reagent contamination. Secondly, DicV positivity by HTS analysis was successfully cross-validated by separate extraction and real-time RT-PCR assays in 30/38 DicV+ patient sera with sufficient leftover volume, excluding the probability of spurious bioinformatics results and further

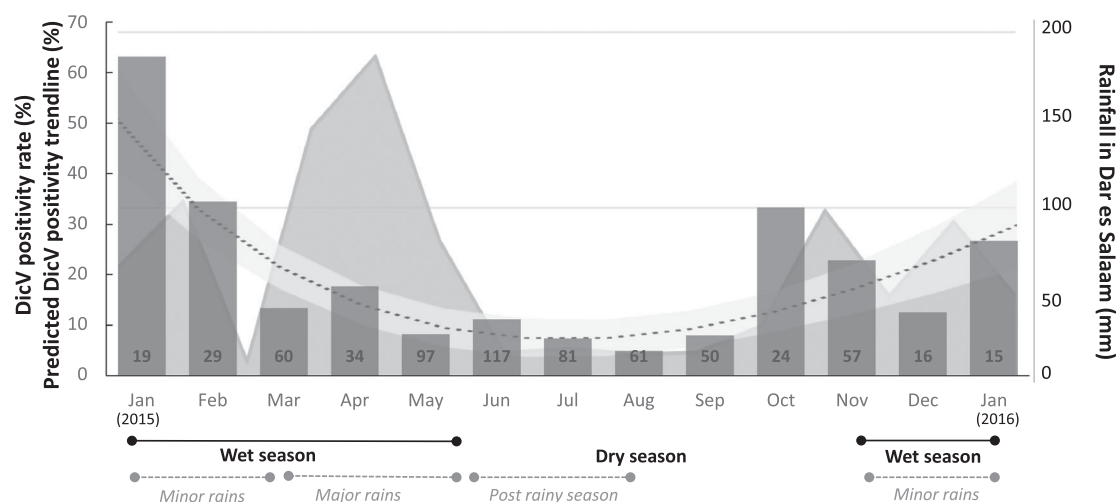


Figure 3. Monthly Dicistrovirus-positivity rate in febrile children. Bar graphs represent the positivity rates (%) (Left Y axis) of DicV RNA across each month of the study. Too few specimens ($n = 10$) were collected in December 2014 and February 2016 to provide reliable positivity rates, and, thus, are not represented on the graph. The total number of sera collected appear in the base of each bar. A dotted line with shaded CI95% shows a fractional polynomial trendline for predicted DicV positivity (also featured on the left Y axis). The rainfall recorded in Dar es Salaam (mm; adapted from <https://www.worldweatheronline.com/dar-es-salaam-weatheraverages/dar-es-salaam/tz.aspx>) over the patient recruitment period appears as a shaded graphic in the background (Right Y axis). Corresponding seasons are indicated below.

eliminating contamination of HTS reagents (especially considering the absence of any common reagent between the HTS and real-time RT-PCR protocols). Five sera with the highest number of DicV reads were confirmed by conventional RT-PCR sequencing (Figure 2a; Table 2). The viral loads detected by real-time RT-PCR corresponded in most cases to the HTS read coverage and ranged from $< 1.32E3$ up to $1x10^7$ viral RNA copies/mL; this viral load heterogeneity speaks against a common and systematic source of contamination and would rather be compatible with the expected kinetics of an acute viral infection.

The sequence heterogeneity (amino acids identity is 97.8% between the 3 DicV full-length sequences MH536109-11) found across DicV+ samples indicate different origins and speaks against a unique viral RNA source or spurious bioinformatics findings. Additionally, a phylogenetic analyse on the capsid region show the presence of two DicV clusters (amino acid identity was 95.2%) without any evidence of co-detections. Although only observed in a single case in blood samples, the presence of the DicV genome was previously reported in a Peruvian cohort of patients with fever of unknown aetiology [21]. In

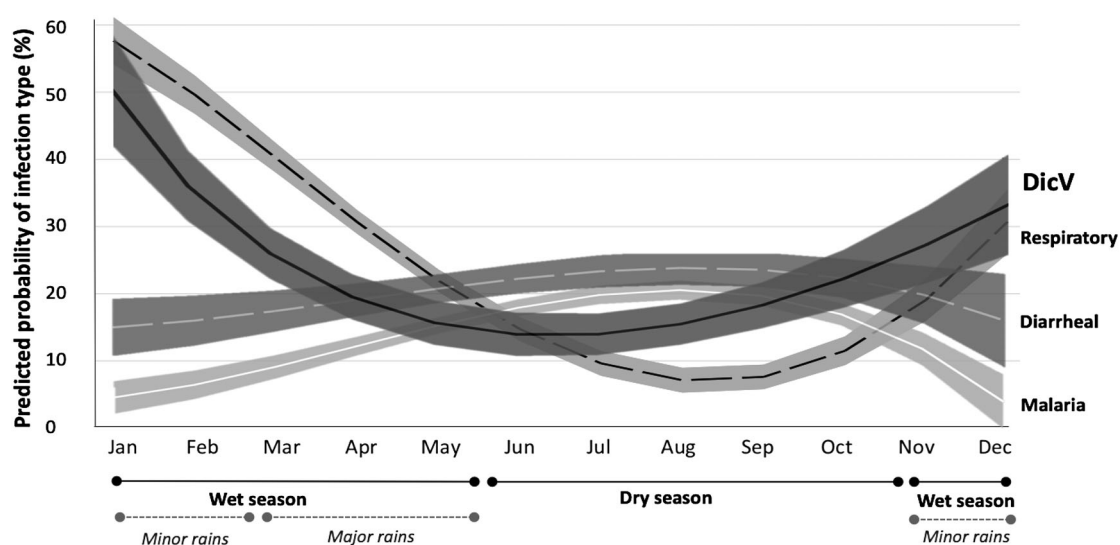


Figure 4: Temporal distribution of diseases representing the major modes of infectious transmission (airborne, fecal-oral and mosquito-borne) compared to DicV prevalence. The mean predicted prevalences (fractional polynomial trendlines) and their 95% confidence intervals (shaded areas around trendlines) are depicted for the probability of having the diagnoses of 1: "Respiratory infection" (dashed black line), 2: "Diarrheal disease" (white dashed line) and 3: Malaria (white solid line). The predicted prevalence of DicV is shown as a solid black line. The months of the entire 2015 recruitment period are depicted. Seasons are indicated below. DicV: Dicistrovirus.

comparison to the Tanzanian sequences, the Peruvian DicV viral sequence showed amino acid identities as low as 35.5% and 22% in the RdRp and the capsid regions, respectively, confirming that they belong to different genera (Figure 2b,c). Finally, when analysing publicly available sequence databases from a recent study that investigated the plasma virome of febrile adult Kenyans [22], and thus has absolutely no link with our investigations, we identified 2'290 DicV reads covering 96.1% (5'350 nt) of ORF1 and 97.5% (2'525 nt) of ORF2 in one pool (SRR4255933). For ORF1, the covered sequence showed 95.4–96.4% nucleotide identity and 98.9–99.2% amino acid identity compared to the Tanzanian DicV sequences. ORF2 had similar results, with 91.5–97.5% nucleotide identity and 94.5–99.2% amino acid identities. The phylogenetic tree of the RdRp region shows that the Kenyan sequence clusters with the Tanzanian ones (Figure 2b), which represents an additional argument for the presence of this DicV novel genus in this geographic region.

Of note, specimens with as few as 6 DicV-specific reads could be confirmed by real-time RT-PCR, whereas 8 with a limited number of reads (≤ 240 , median = 16) could not be confirmed. This latter result could be explained either by a higher sensitivity of the HTS in specific situations (particularly in relation to the additional freeze–thaw cycle for the real-time RT-PCR) and/or by the presence of primer-probe mismatches (this hypothesis could not be checked since only a very limited number of reads were obtained). The number of reads obtained through HTS provides a rough estimate of the viral load and depends on various factors, such as cellularity, the presence of other microorganisms and the efficiency of the library preparation [23].

Since dicistroviruses have been observed in a large spectrum of arthropods worldwide, one hypothesis raised by our findings is whether DicV could represent a viral agent or reflects a viral RNA transmission through contact with vector that still has to be discovered. Indeed, the seasonality clustering observed in this study, may suggest vector-borne transmission although others transmission routes cannot be ruled out. A zoonotic reservoir is also possible since DicVs have been detected in stools from various non-human primates (such as gorillas [24]) as well as giant pandas [25], and bats [26], which also raises the possibility of a faecal-oral or sylvatic transmission cycle. Furthermore, the presence of DicV sequences were recently reported in the blood of a fruit bat captured in 2015 from the Republic of Congo [27]. In humans, unbiased sequencing analysis of stool samples collected from 35 South Asian children presenting with non-polio acute flaccid paralysis revealed the presence of DicV in 3 subjects coinfecting with other viruses [28]. Another study reported DicV sequences in stool samples from

Australian children with acute diarrhoea [29]. In both studies, whether the presence of DicV in these stools resulted from a viral replication in the human gastrointestinal tract or from environmental contamination was not determined. At the *in vitro* level, the cellular tropism of DicV has been investigated using the Taura syndrome virus (TSV): a DicV that infects shrimp. Exposing various primate cells to this virus revealed a range of susceptibilities, where RD, Hep-2C and BGM cells were considered vulnerable to infection [30], while FRhK-4, MA-104 and BGMK remained non-permissive [31]. Both animal models as well as reverse genetics approaches will be needed to further investigate the tropism and likely replication sites of DicV in humans.

The present cohort offered the unique opportunity to assess demographics, symptoms and signs associated with the presence of DicV RNA. At this time, comparative analysis between DicV-positive and negative patients (HTS analysis) did not reveal any significant clinical differences, although the positive patients were significantly younger. The presence of DicV was not associated with any other co-infections (such as malaria, HIV or skin infections) nor was it linked to any other comorbidities. However, our cohort specifically enrolled patients at the outpatient level who presented unspecific febrile illnesses and thus covered a homogenous spectrum of clinical syndromes from which to draw conclusions. In the future, larger and more diverse cohorts should be screened, including more severe hospitalized children and asymptomatic blood donors.

In contrast to the paediatric population, DicV RNA was infrequently detected in adults. Whether this finding reflects acquired protective immunity in adults, a more frequent transmission from the vector, a greater susceptibility in the paediatric population or as mentioned above a temporal bias needs to be evaluated. A recent study investigated the seroprevalence of antibodies against another DicV from another genus named *Triatoma virus* (TrV) [32]. TrV is known to infect the vector responsible for Chagas disease transmission. Their findings revealed a high prevalence of anti-TrV antibodies against the TrV structural proteins in human sera of subjects living in Argentina, Bolivia and Mexico and suggested a passive transmission of the virus (i.e. absence of viral replication in human hosts). To fully investigate the possible transmission of DicV, future serological investigations should be carried out on a more diverse collection of symptomatic and asymptomatic adult and paediatric cohorts from varied geographical regions inside. At this time, we may raise the hypothesis of a vector-borne transmission and that, among the potential vectors, triatomines should be considered.

While this study does not establish the pathogenicity or even a true human infection, our findings

suggest that DicV could represent a novel agent to consider in humans. We lack an afebrile cohort to assess the association. Other than the small increase in the inflammatory marker, PCT, we could not identify any specific clinical features, diagnoses or co-morbidities associated to DicV infection despite the detailed clinical database gathered at the time of sample collection. Elevated PCT is commonly associated (although not strongly) with bacterial infections; it may be increased either directly by bacterial endotoxins and lipopolysaccharides or indirectly by inflammatory mediators (such as tumour necrosis factor- α , interleukin-6, interleukin-1). Mediators of viral infection (interferon- γ) are thought to attenuate PCT levels. Thus, its increase in DicV+ patients could be an indication that DicV may have a pathological immunomodulatory effect on its human host or an as yet unidentified association to another bacterial infection. However, given the small size effect size, no clinically relevant conclusions could be drawn from this observation. Notably, the *de novo* analysis performed on all sera and plasma specimens (from the paediatric and adult cohorts respectively) revealed the presence RNA sequences from viruses known to infect humans (Supplementary Figure S1). Of the 103 paediatric patients who were DicV+ by HTS, only ten presented a co-detection with sequences from these known human RNA viruses (two human pegivirus-1, two enteroviruses, four rhinoviruses and two rotaviruses). Human pegivirus-1 sequences were also detected in the plasma of the unique DicV+ adult patient by HTS. Studies in larger and more clinically diverse patients (including asymptomatic subjects) are needed to clarify whether DicV is pathogenic or if it should rather be considered as a member of the commensal blood virome, similarly to torque tenovirus or human pegivirus. Screening other types of samples will also help to define the *in vivo* tropism of DicV.

In conclusion, our findings that DicV RNA is frequently detected in the blood of Tanzanian children deserve attention. This study provides the first step toward a possible new agent in human, although without proven viral replication and open the door to further investigation to confirm and contextualize these results by screening other populations, and fulfilling all or part of Koch's postulates.

Acknowledgments

We thank Guy Boivin (Laval University), Gary Kobinger (Laval University), Francois Chappuis (University of Geneva Hospitals), Isabella Eckerle (University of Geneva Hospitals) and Philippe Lemercier (Swiss Institute of Bioinformatics) for their valuable comments and advice. We are also grateful to Lara Turin (University of Geneva Hospitals) for technical assistance. Finally, we thank the patients and their caretakers for their participation and contribution to improved medical care for febrile illness in Tanzania. This study was

supported by the Bill and Melinda Gates Foundation (grant n°. OPP1163434).

Disclosure statement

No potential conflict of interest was reported by the authors.

Funding

This study was supported by the Bill and Melinda Gates Foundation (grant number OPP1163434).

ORCID

Samuel Cordey  <http://orcid.org/0000-0002-2684-5680>

References

- [1] Anthony SJ, et al. A strategy to estimate unknown viral diversity in mammals. *mBio*. 2013;4:e00598–e00513.
- [2] Shi M, et al. Redefining the invertebrate RNA virosphere. *Nature*. 2016;540:539–543.
- [3] van Aerle R, Santos EM. Advances in the application of high-throughput sequencing in invertebrate virology. *J Invertebr Pathol*. 2017;147:145–156.
- [4] Brown JR, et al. Astrovirus VA1/HMO-C: an increasingly recognized neurotropic pathogen in immunocompromised patients. *Clin Infect Dis*. 2015;60:881–888.
- [5] Cordey S, et al. Astrovirus MLB2, a New Gastroenteric virus associated with Meningitis and Disseminated infection. *Emerg Infect Dis*. 2016;22:846–853.
- [6] Hoffmann B, et al. A Variegated Squirrel Bornavirus associated with Fatal human Encephalitis. *N Engl J Med*. 2015;373:154–162.
- [7] Komoto S, Adah MI, Ide T, et al. Whole genomic analysis of human and bovine G8P[1] rotavirus strains isolated in Nigeria provides evidence for direct bovine-to-human interspecies transmission. *Infect Genet Evol*. 2016;43:424–433.
- [8] Naccache SN, et al. Diagnosis of neuroinvasive astrovirus infection in an immunocompromised adult with encephalitis by unbiased next-generation sequencing. *Clin Infect Dis*. 2015;60:919–923.
- [9] Ai JW, et al. Human Endophthalmitis Caused By pseudorabies virus infection, China, 2017. *Emerg Infect Dis*. 2018;24:1087–1090.
- [10] Holmes EC. The Expanding Virosphere. *Cell Host Microbe*. 2016;20:279–280.
- [11] Keitel K, Kagoro F. A novel electronic algorithm using host biomarker point-of-care tests for the management of febrile illnesses in Tanzanian children (e-POCT): A randomized, controlled non-inferiority trial. *PLoS Med*. 2017;14:e1002411.
- [12] D'Acremont V. Etiologies of Acute Febrile Illness Among Adults Attending an Outpatient Department in Dar es Salaam. *Clinical Trials.gov Identifier: NCT01947075*.
- [13] Bonning BC, Miller WA. Dicistroviruses. *Annu Rev Entomol*. 2010;55:129–150.
- [14] Petty T, et al. Comprehensive human virus screening using high-throughput sequencing with a user-friendly representation of bioinformatics analysis: a pilot study. *J Clin Microbiol*. 2014;52:3351–3361.

- [15] Peng Y, Leung HC, Yiu SM, et al. IDBA-UD: a de novo assembler for single-cell and metagenomic sequencing data with highly uneven depth. *Bioinformatics*. [2012;28:1420–1428](#).
- [16] Wright ES, Vetsigian KH. Quality filtering of Illumina index reads mitigates sample cross-talk. *BMC Genomics*. [2016;17:876](#).
- [17] Le SQ, Gascuel O. An improved general amino acid replacement matrix. *Mol Biol Evol*. [2008;25:1307–1320](#).
- [18] Kumar S, Stecher G, Li M, et al. MEGA x: molecular Evolutionary genetics analysis across Computing Platforms. *Mol Biol Evol*. [2018;35:1547–1549](#).
- [19] Schibler M, et al. Critical analysis of rhinovirus RNA load quantification by real-time reverse transcription-PCR. *J Clin Microbiol*. [2012;50:2868–2872](#).
- [20] Lipkin WI. The changing face of pathogen discovery and surveillance. *Nature Rev Microbiol*. [2013;11:133–141](#).
- [21] Phan TG, et al. Sera of Peruvians with fever of unknown origins include viral nucleic acids from non-vertebrate hosts. *Virus Genes*. [2018;54:33–40](#).
- [22] Ngoi CN, et al. The plasma virome of febrile adult Kenyans shows frequent parvovirus B19 infections and a novel arbovirus (Kadipiro virus). *J Gen Virol*. [2016;97:3359–3367](#).
- [23] Schlaberg R, Chiu CY, Miller S, et al. Validation of Metagenomic Next-Generation sequencing tests for Universal pathogen detection. *Arch Pathol Lab Med*. [2017;141:776–786](#).
- [24] Duraisamy R, et al. Detection of novel RNA viruses from free-living gorillas, Republic of the Congo: genetic diversity of picobirnaviruses. *Virus Genes*. [2018;54:256–271](#).
- [25] Zhang W, et al. Virome comparisons in wild-diseased and healthy captive giant pandas. *Microbiome*. [2017;5:90](#).
- [26] Reuter G, Pankovics P, Gyongyi Z, et al. Novel dicistrovirus from bat guano. *Arch Virol*. [2014;159:3453–3456](#).
- [27] Bennett AJ, et al. Diverse RNA viruses of arthropod origin in the blood of fruit bats suggest a link between bat and arthropod viromes. *Virology*. [2018;528:64–72](#).
- [28] Victoria JG, et al. Metagenomic analyses of viruses in stool samples from children with acute flaccid paralysis. *J Virol*. [2009;83:4642–4651](#).
- [29] Holtz LR, et al. Geographic variation in the eukaryotic virome of human diarrhea. *Virology*. [2014;468-470:556–564](#).
- [30] Audelo-del-Valle J, et al. Infection of cultured human and monkey cell lines with extract of penaeid shrimp infected with Taura syndrome virus. *Emerg Infect Dis*. [2003;9:265–266](#).
- [31] Pantoja CR, Navarro SA, Naranjo J, et al. Nonsusceptibility of primate cells to Taura syndrome virus. *Emerg Infect Dis*. [2004;10:2106–2112](#).
- [32] Querido JF, et al. Seroprevalence of *Triatoma* virus (Dicistroviridae: Cripaviridae) antibodies in Chagas disease patients. *Parasites Vectors*. [2015;8:29](#).