Title
Effect of Geographic Isolation on the Nasal Virome of Indigenous Children.

Journal
Journal of Virology, 93(17)

ISSN
0022-538X

Authors
Altan, Eda
Dib, Juan Carlos
Gulloso, Andres Rojas
et al.

Publication Date
2019-09-01

DOI
10.1128/jvi.00681-19

Peer reviewed

AMERICAN SOCIETY FOR MICROBIOLOGY | Journal of Virology®

GENETIC DIVERSITY AND EVOLUTION

Effect of Geographic Isolation on the Nasal Virome of Indigenous Children

Eda Altan,[a,b] Juan Carlos Dib,[c,d] Andres Rojas Gulloso,[d] Duamaco Escribano Juandigua,[d] Xutao Deng,[a,b] Roberta Bruhn,[a,b] Kristen Hildebrand,[e] Pamela Freiden,[e] Janie Yamamoto,[a,b] Stacey Schultz-Cherry,[e] Eric Delwart[a,b]

[a]Vitalant Research Institute, San Francisco, California, USA
[b]Department of Laboratory Medicine, University of California, San Francisco, San Francisco, California, USA
[c]Universidad del Norte, Barranquilla, Atlantico, Colombia
[d]Fundación Salud Para el Trópico-Tropical Health Foundation, Santa Marta, Magdalena, Colombia
[e]Department of Infectious Diseases, St. Jude Children's Research Hospital, Memphis, Tennessee, USA

ABSTRACT The influence of living in small remote villages on the diversity of viruses in the nasal mucosa was investigated in three Colombian villages with very different levels of geographic isolation. Metagenomic analysis was used to characterize viral nucleic acids in nasal swabs from 63 apparently healthy young children. Sequences from human virus members of the families *Anelloviridae*, *Papillomaviridae*, *Picornaviridae*, *Herpesviridae*, *Polyomaviridae*, *Adenoviridae*, and *Paramyxoviridae* were detected in decreasing proportions of children. The number of papillomavirus infections detected was greater among Hispanic children most exposed to outside contacts, while anellovirus infections were more common in the isolated indigenous villages. The diversity of the other human viruses detected did not differ among the villages. Closely related variants of rhinovirus A or B were identified in 2 to 4 children from each village, reflecting ongoing transmission clusters. Genomes of viruses not currently known to infect humans, including members of the families *Parvoviridae*, *Partitiviridae*, *Dicistroviridae*, and *Iflaviridae* and circular Rep-encoding single-stranded DNA (CRESS-DNA) virus, were also detected in nasal swabs, possibly reflecting environmental contamination from insect, fungal, or unknown sources. Despite the high levels of geographic and cultural isolation, the overall diversity of human viruses in the nasal passages of children was not reduced in highly isolated indigenous villages, indicating ongoing exposure to globally circulating viruses.

IMPORTANCE Extreme geographic and cultural isolation can still be found in some indigenous South American villages. Such isolation may be expected to limit the introduction of otherwise common and widely distributed viruses. Very small population sizes may also result in rapid local viral extinction due to a lack of seronegative subjects to maintain transmission chains for rapidly cleared viruses. We compared the viruses in the nasal passages of young children in three villages with increasing levels of geographic isolation. We found that isolation did not reduce the overall diversity of viral infections. Multiple infections with nearly identical rhinoviruses could be detected within each village, likely reflecting recent viral introductions and transmission clusters among epidemiologically linked members of these very small communities. We conclude that, despite their geographic isolation, remote indigenous villages show evidence of ongoing exposure to globally circulating viruses.

KEYWORDS Colombia, nasal virome, next-generation sequencing, children, metagenomics, papillomavirus, rhinovirus

The impact of extreme geographic isolation in shaping the respiratory virome remains largely unknown. In the preagricultural era, people typically lived widely dispersed in small nomadic groups, a lifestyle that might have minimized the spread

Citation Altan E, Dib JC, Gulloso AR, Escribano Juandigua D, Deng X, Bruhn R, Hildebrand K, Freiden P, Yamamoto J, Schultz-Cherry S, Delwart E. 2019. Effect of geographic isolation on the nasal virome of indigenous children. J Virol 93:e00681-19. https://doi.org/10.1128/JVI.00681-19.

Editor Julie K. Pfeiffer, University of Texas Southwestern Medical Center

Address correspondence to Eric Delwart, eric.delwart@ucsf.edu.

Received 25 April 2019
Accepted 29 May 2019
Accepted manuscript posted online 12 June 2019
Published 13 August 2019

TABLE 1 Age and gender of children analyzed

Village	No. of children			
	2–5 yr of age	6–9 yr of age	Girl	Boy
Calabazo	11	10	8	13
Seywiaka	12	9	9	12
Umandita	13	8	17	4

and maintenance of infectious diseases that did not establish long-lasting or chronic infections. Small populations now settled in hard-to-explore regions may still be relatively isolated from repeated exposures to highly prevalent viruses circulating in larger, more connected, communities. Therefore, inhabitants of such highly isolated villages might have lost viruses that are dependent on the large numbers of young, seronegative, susceptible hosts found in larger populations (1).

Coincident with the arrival of Europeans, native Amerindian populations underwent strong population bottlenecks, possibly due to imported airborne epidemics (such as smallpox, measles, and, more recently, influenza viruses) to which they had no prior exposure (2, 3). A recent study of feces collected from children in isolated villages in the Venezuelan Amazon region showed that extreme geographic isolation did not result in a reduction in enteric human virus diversity (4). To determine whether reduced rates of outside contact coincide with a reduction in the diversity of human viruses in the respiratory tract, we compared the nasal mucosal viromes of children in two highly isolated Amerindian Kogi villages in a tropical forest of northern Colombia and in one largely Hispanic village alongside a coastal highway. In order to detect all human viruses, viral metagenomic analysis was applied to nasal swabs collected from children 2 to 9 years of age.

(This article was submitted to an online preprint archive [5].)

RESULTS

Sample collection and village locations. Nasal swabs were collected from 63 children (53.9% female), with a mean age of 5 years (Table 1). The children lived in three villages in northern Columbia that differed in the degree of outside contact. Samples used for comparison were from age- and gender-matched children (Table 1). Satellite pictures of these villages can be seen on Google Maps, using their Global Positioning System (GPS) coordinates. The first village, Calabazo (GPS coordinates, 11.28448, −74.00195), is located along a major road (highway 90) running alongside the National Natural Park of Tayrona and is frequently visited by tourists. Calabazo has a 2005 census population of 499, and the main language is Spanish. Seywiaka (GPS coordinates, 11.2174, −73.5794) is an isolated village with a population size of 250 to 300 that is accessible only by foot (1.5-h walk from the nearest road) and is inhabited by Kogi people speaking their indigenous language (Fig. 1, top). Nonindigenous people visit every week to work in Seywiaka's school and health care center. Indigenous adults from Seywiaka travel to the town of Palomino (located along highway 90, with a 2000 census population of 3,900) approximately twice a month. An even more isolated Kogi village, Umandita (GPS coordinates, 11.09698, −73.64781), with an estimated population of 350 to 400 inhabitants, is accessible after a 9- to 10-h walk from Seywiaka (Fig. 1, bottom). Only a few adults travel from Umandita to Seywiaka or further to Palomino, approximately twice a year. The health team personnel collecting the biological samples analyzed here visit Seywiaka and Umandita twice a year to collect samples.

Nasal mucosa virome. Following viral metagenomic enrichment of viral particle-associated nucleic acids in nasal swabs, random RNA and DNA amplification, and deep sequencing of 63 individual nasal swab supernatants, a total of approximately 63 million reads were generated. We found 92% of samples (58/63 samples) to be positive for at least one human virus. Human viruses belonging to seven viral families (in decreasing prevalence of detection (*Anelloviridae*, *Papillomaviridae*, *Picornaviridae*, *Herpesviridae*, *Polyomaviridae*, *Adenoviridae*, and *Paramyxoviridae*) were detected.

FIG 1 View of isolated indigenous village of Seywiaka (top) and the most isolated indigenous village of Umandita (bottom).

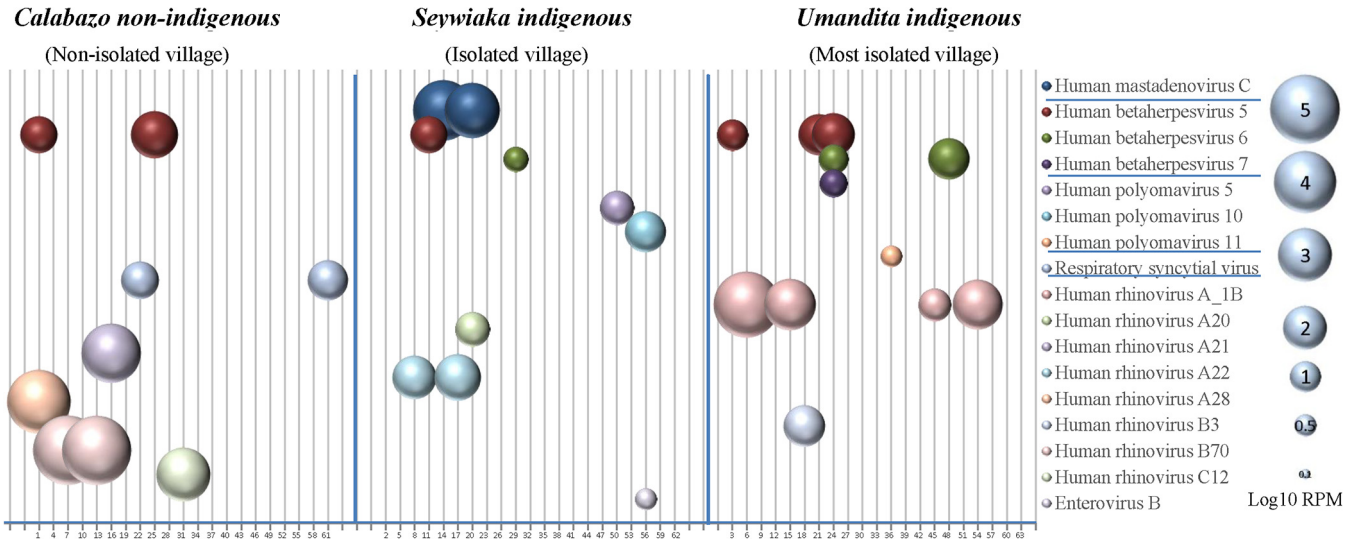

FIG 2 Distribution and levels of reads for human viruses within villages. To display the amounts of viral reads in each sample, a bubble chart was generated based on the number of reads for named viruses (BLASTx E scores of $<10^{-10}$) per 1 million total reads (reads per million [RPM]), with $\log_{10}$ transformation. For example, 10 and 1,000 viral reads in 10^6 total reads yield values of 1 and 3, respectively.

Anelloviridae family reads were the most commonly detected viral sequences and were found in 49/63 children (77.7%); 0.16% ($n = 100,957$ sequence reads) of 63 million total reads could be mapped to the *Anelloviridae* family, with BLASTx E scores of $<10^{-10}$. The second most commonly detected human virus reads belonged to the *Papillomaviridae* family and were detected in 44.4% of children (28/63 children), with 0.087% of total reads ($n = 55,248$ sequence reads). Next, with a prevalence of 23.8% (15/63 children), were reads from the *Picornaviridae* family. Of these, 0.094% of reads ($n = 59,819$ sequence reads) were picornavirus reads from the species rhinovirus A (10/63 children [15.8%]), rhinovirus B (3/63 children [4.7%]), rhinovirus C (1/63 children [1.58%]), and enterovirus B (1/63 children [1.58%]). *Herpesviridae* family members were next in prevalence, being detected in 7/63 children (11.1%) and including human betaherpesvirus 5 (HHV5 or cytomegalovirus) (6/63 children [9.52%]), human betaherpesvirus 6 (HHV6 or roseolovirus) (3/63 children [4.76%]), and human betaherpesvirus 7 (HHV7 or Kaposi sarcoma virus) (1/63 children [1.58%]). In the *Polyomaviridae* family, we detected human polyomavirus 5 (HPyV5) (Merkel cell carcinoma virus) (6) (1/63 children [1.58%]), HPyV10 (Malawi polyomavirus) (7, 8) (1/63 children [1.58%]), and HPyV11 (Saint Louis polyomavirus) (9) (1/63 children [1.58%]). Adenovirus C reads were detected in 2/63 children (3.17%). Respiratory syncytial virus (RSV), belonging to the *Paramyxoviridae* family, was found in 2/63 children (3.17%). This was the only viral family detected exclusively in the most exposed Calabazo village. The fractions of total reads from each sample encoding proteins with high-level similarity (E scores of $<10^{-10}$) to human viruses are shown in Fig. 2, with the exception of the papillomaviruses and anelloviruses, which are further analyzed below.

Family *Picornaviridae*. Fourteen children showed the presence of picornavirus sequences, 13 from rhinovirus A, B, or C species and 1 from enterovirus B species. Rhinovirus reads generated contigs ranging in size from 481 to 7,089 nucleotides. In total, 8 contigs included the complete 5′-untranslated region (UTR) -VP4-VP2, 2 contigs the complete UTR-VP4, and 1 contig the complete VP4-VP2. The complete VP4-VP2 regions were used for phylogenetic analysis (Fig. 3). Four rhinovirus contigs from the Umandita region showed the closest nucleotide identity (90% to 92%) to genotype 1B of rhinovirus A. Contigs from 3 of these children overlapped almost the entire genome (6.6 kb without gaps) and showed nucleotide identity of 99.8 to 99.9%. Two rhinovirus A genotype 1B contigs included the VP4-VP2 region and clustered tightly together, reflecting a recent common origin and an ongoing transmission cluster in the most

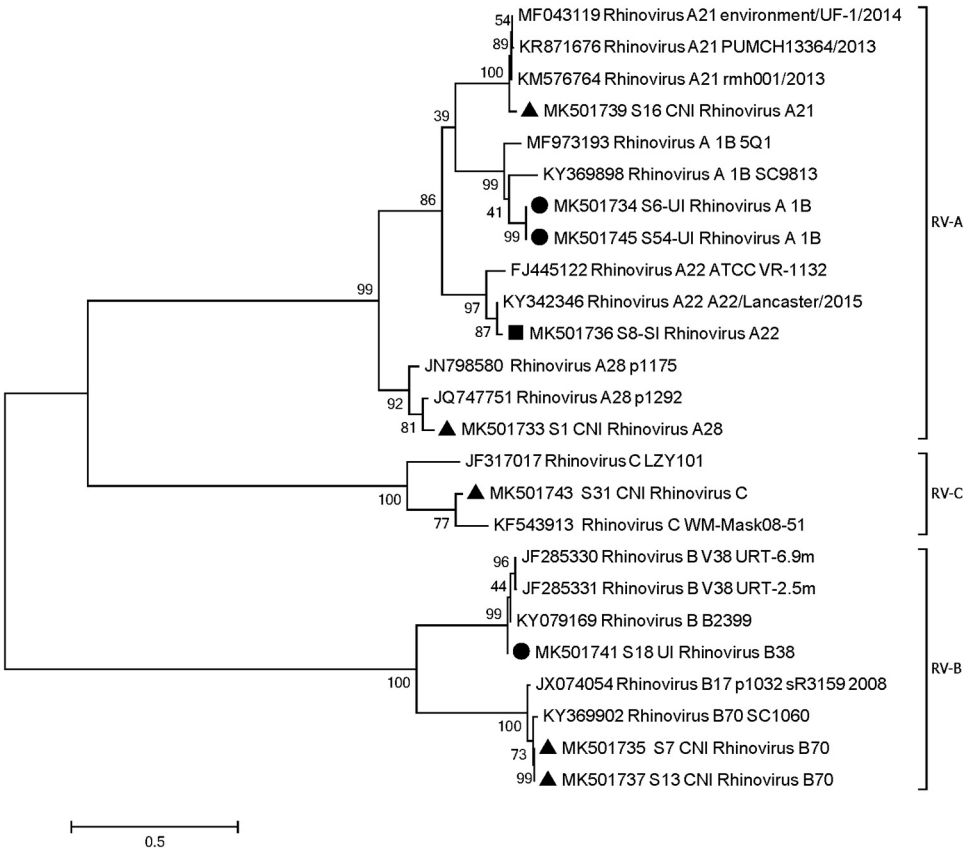

FIG 3 Phylogenetic analysis of VP4-VP2 region of rhinoviruses inferred using the maximum likelihood method, based on the general time reversible model. A discrete gamma distribution was used to model evolutionary rate differences among sites. The tree is drawn to scale, with branch lengths measured in the number of substitutions per site. There were a total of 518 positions in the final data set. Evolutionary analyses were conducted in MEGA6. ●, Calabazo nonindigenous (CNI); ▲, Seywiaka indigenous (SI); ■, Umandita indigenous (UI).

isolated village, Umandita (Fig. 3). Two children from Calabazo were infected with rhinovirus B70 with 99.9% nucleotide identity, indicating another transmission cluster occurring at the time of sampling. Two children from Seywiaka were shedding rhinovirus A22, but only 1 generated enough sequence reads to be included in VP4-VP2 phylogenetic analysis (Fig. 3). Reads from these two children did overlap by 154 bases, showing a single mismatch, which indicates another possible transmission cluster. Therefore, rhinovirus transmission clusters (genotype A1B in Umandita, genotype B70 in Calabazo, and genotype A22 in Seywiaka) were detected in each village. The enterovirus B reads from a Seywiaka child showed closest amino acid identity (93%) to echovirus E15 (GenBank accession number AY302541).

Polyomaviruses. Polyomavirus sequences were also found in the isolated indigenous villages of Seywiaka ($n = 2$) and Umandita ($n = 1$). Two Seywiaka villagers were shedding HPyV5 or HPyV10, and 1 child from Umandita was shedding HPyV11 (Fig. 2). HPyV5 reads showed nucleotide identity of 98 to 99%, HPyV10 identity of 92 to 98%, and HPyV11 identity of 91 to 99% with genomes available in GenBank, falling within the range of diversity seen in these viral species.

Herpesviruses. Sequences of human CMV, roseolovirus, and Kaposi sarcoma virus were identified. CMV sequences were found in 6 children (2, 1 and 3 children from Calabazo, Seywiaka, and Umandita respectively). Three children shed roseolovirus (1 and 2 from Seywiaka and Umandita respectively). A single Kaposi sarcoma virus infection was detected, in a child from Umandita (Fig. 2). All contigs showed 98 to 100% nucleotide identities to genomes in GenBank.

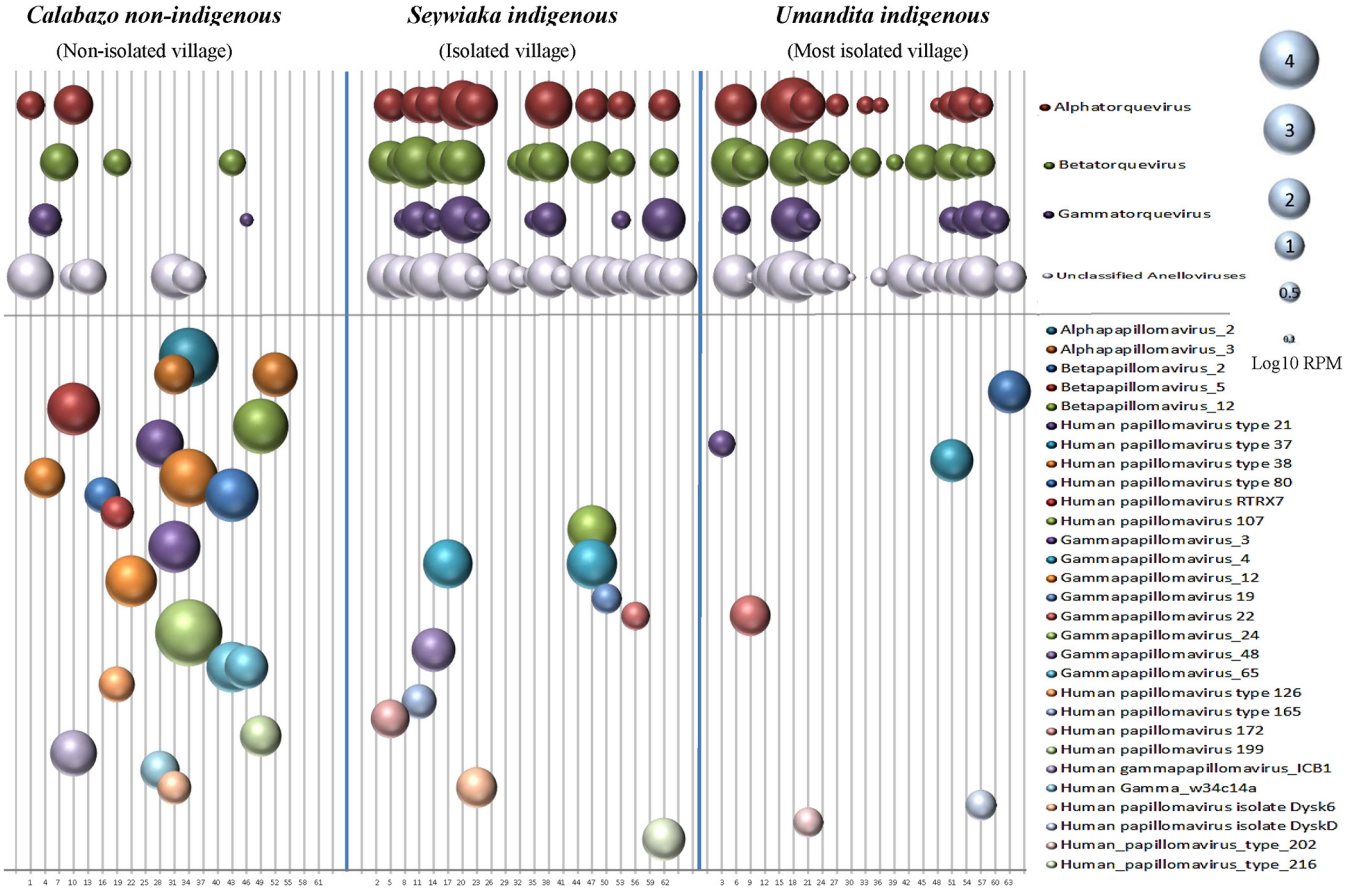

FIG 4 Distribution and levels of anellovirus (top) and papillomavirus (bottom) reads within villages. The bubble charts were generated based on the number of reads for named viruses (BLASTx E scores of $<10^{-10}$) per 1 million total reads (reads per million [RPM]), with $\log_{10}$ transformation.

Adenovirus, pneumovirus, and parvovirus. Sequences from human mastadenovirus C species in the *Adenoviridae* family, ranging in size from 250 to 1831 nucleotides, were identified in 2 children from Seywiaka village (Fig. 2). Different regions (within E3, E4, E1a, and L3) showed overlap between reads from the 2 children, with nucleotide identity of 83% to 97%, likely reflecting 2 different strains of adenovirus C. Similarity of adenovirus reads to genomes in GenBank ranged from 92% to 99%.

Two RSV contigs, of 363 and 888 nucleotides, were identified from 2 children in Calabazo, both showing 99% nucleotide identity with RSV strain A (GenBank accession number MG793382). Contigs from these 2 children overlapped in the G gene (350 bp), showing nucleotide identity of 99.1%. The close sequence identity of these two RSV strains may also reflect an ongoing transmission cluster within that village.

Unexpectedly, 2 reads of canine bocavirus were also detected in 1 swab sample from Calabazo village, showing 92 and 97% amino acid identity to the canine bocavirus NS1 gene region (GenBank accession number MG025952).

Anelloviruses. The prevalence of anellovirus detection was 42% (10/21 children), 90% (19/21 children), and 95.2% (20/21 children) in children from Calabazo, Seywiaka, and Umandita respectively. Therefore, anellovirus DNA was detected in 77.7% of all children (49/63 children). The overall fractions of children infected with different anellovirus genera were 34% with alphatorquevirus, 44.4% with betatorquevirus, 28.5% with gammatorquevirus, and 65% with unclassified anelloviruses (Fig. 4, top).

Papillomaviruses. Altogether, we detected 29 papillomaviruses (17 genotypes) in 13 Calabazo children, 10 papillomaviruses (9 genotypes) in 9 Seywiaka children, and 6 papillomaviruses (6 genotypes) in 6 Umandita children (Fig. 4, bottom). Thirty-seven partial papillomavirus genome contigs, ranging in size from 261 to 7,392 nucleotides,

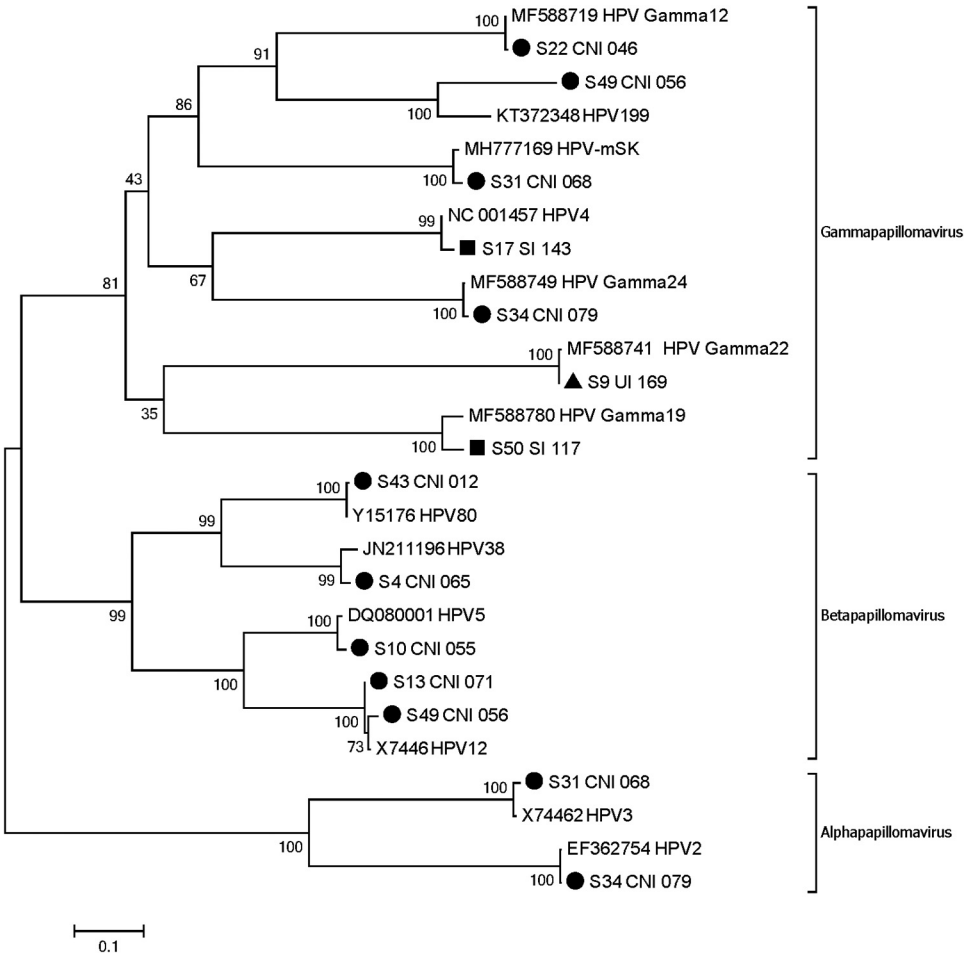

FIG 5 Phylogenetic analysis of the major capsid (L1) protein of papillomaviruses inferred using the maximum likelihood method, based on the general time reversible model. A discrete gamma distribution was used to model evolutionary rate differences among sites. There were a total of 1,693 positions in the final data set. Evolutionary analyses were conducted in MEGA6. ●, Calabazo nonindigenous (CNI); ▲, Seywiaka indigenous (SI); ■, Umandita indigenous (UI).

were generated, 14 of which included a partial L1 gene region. Phylogenetic analysis of these L1 sequences was performed (Fig. 5). All papillomavirus contigs showed 97 to 100% amino acid identity to papillomavirus proteins in GenBank. Papillomavirus (human papillomavirus 12) contigs in 2 children from Calabazo village (S13-CNI and S49-CNI) were closely linked (Fig. 5), showing 99% nucleotide identity.

Virome comparison between villages. We next compared the distribution of the two viral families that yielded the most reads, anelloviruses and papillomaviruses, among the three villages (Fig. 4). The numbers of anellovirus infections were significantly different among the villages ($P = 0.0001$) (see Materials and Methods). Inspection of the anellovirus distribution indicated that more anellovirus infections were detected in the most isolated villages of Seywiaka and Umandita (Fig. 4, top).

An analysis of papillomavirus read distribution similarly showed that the numbers of papillomavirus infections were significantly different among the villages ($P = 0.043$) (see Materials and Methods). Contrary to anelloviruses, a significantly greater number of papillomavirus infections was detected in the most exposed, Hispanic village of Calabazo (Fig. 4, bottom).

Viral families of nonvertebrate or unknown host tropism. Sequences from viral families not known to infect humans (or vertebrates), likely representing airborne mucosal surface contamination, were detected in 44/63 children (69.8% of children). Members of the following viral clades, ranked from highest to lowest prevalence, were detected: *Parvoviri-*

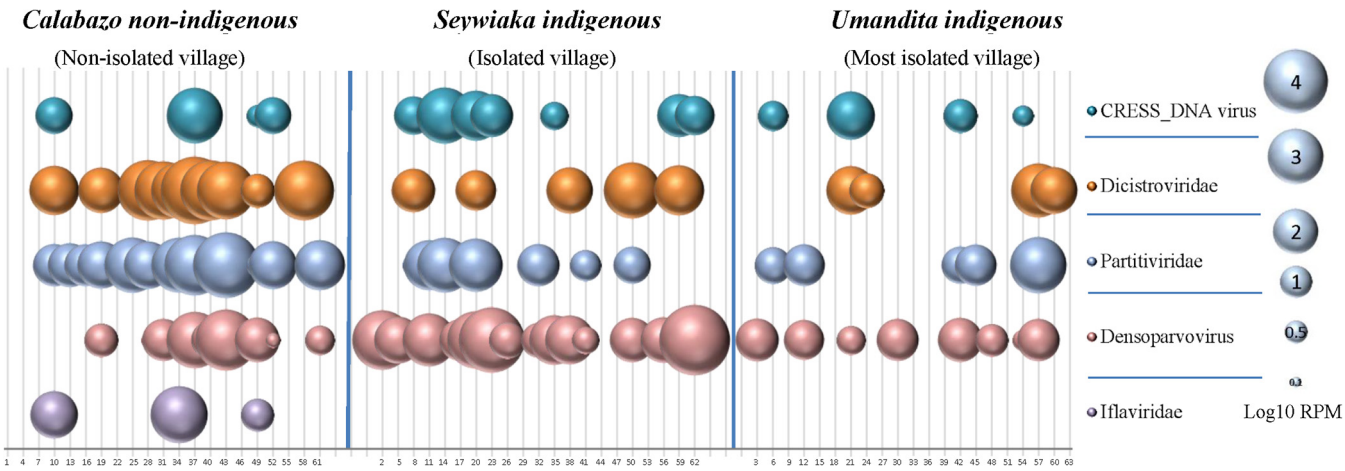

FIG 6 Distribution and levels of reads for viral groups not known to infect vertebrates, or with unknown tropism, within villages. The bubble chart was generated based on the number of reads for named viruses (BLASTx E scores of $<10^{-10}$) per 1 million total reads (reads per million [RPM]), with $\log_{10}$ transformation.

dae (densoviruses), *Partitiviridae*, *Dicistroviridae*, circular Rep-encoding single-stranded DNA (CRESS-DNA) viruses, and *Iflaviridae* (found in 49.2%, 38.09%, 30.1%, 23.8%, and 4.7% of the nasal swab samples respectively) (Fig. 6).

Densoparvoviruses, dicistroviruses, and iflaviruses are known to infect invertebrates, and partitiviruses are known to infect fungi and plants. Some CRESS-DNA viruses can infect fungi, plants, or mammals, but the tropism of most CRESS-DNA virus genomes, largely identified through metagenomic analyses of environmental samples (including those detected here), remains unknown.

DISCUSSION

Viral metagenomics of human respiratory secretions have analyzed different sample types, mainly from clinical cases. Early studies of nasopharyngeal aspirates from lower respiratory tract infections in China (10) and Sweden (11) revealed numerous viruses, with members of the families *Paramyxoviridae* (RSV, metapneumovirus, and parainfluenza virus), *Picornaviridae* (rhinovirus), and *Orthomyxoviridae* (influenza virus) predominating. Numerous other metagenomic studies of clinical respiratory samples have confirmed the ability of this method to detect diverse human viruses (12–17). Metagenomic analyses thus show great promise as a supplement to or even replacement for more specific viral genome detection assays, although sensitivity issues remain (18–21). A more limited number of metagenomic studies have analyzed the respiratory tract viromes of healthy children. Double-stranded DNA from the anterior nares of healthy individuals in the Human Microbiome Project showed that betapapillomaviruses and gammapapillomaviruses were the most common viruses detected, followed by roseolovirus/HHV6 (22). A PCR study of sinonasal mucosa from sinus surgery patients, testing for 16 common respiratory viruses, indicated that HHV6 was the most frequently detected virus (23). A study of nasopharyngeal swabs from healthy 18-month-old children, using quantitative PCR, showed the presence of human rhinoviruses, adenoviruses, bocaviruses, and parainfluenza virus (24). Nasopharyngeal swabs from healthy children that were analyzed by metagenomics and pan-viral group PCR showed anelloviruses, HHV6, and HHV7 to be the most common infections (14).

Here we characterize the nasal mucosal viromes of 63 age- and gender-matched children from three villages with different degrees of geographic isolation. The fraction of all reads encoding human viral sequences was 0.6%, with the most reads being derived from anelloviruses, followed by papillomaviruses and picornaviruses. We show that the geographic and cultural isolation of the two most isolated (indigenous) villages did not result in the elimination or even in a reduction of the diversity of their human viruses.

Three different herpesviruses (HHV5 to HHV7) and three different polyomaviruses (HPyV5, HPyV10, and HPyV11) were detected in the two isolated villages, while only one herpesvirus (HHV5) and no polyomaviruses were found in the village with frequent outside contact (Fig. 2). RSV was the only virus found exclusively in the most exposed village. Four different rhinovirus genotypes and 1 enterovirus B genotype were found in the two isolated villages, while 4 rhinovirus genotypes were detected in the more exposed village. There was no overlap in the picornavirus genotypes in the different villages. Among 0- to 5-year-old children, rates of asymptomatic rhinovirus detection of 12.5% to 33% have been reported (25–28). A rhinovirus detection rate of 33% among <3-year-old asymptomatic children did not significantly differ from that found in matched hospitalized children (27). Here, we found an average rate of rhinovirus detection of 20.6% in healthy 2- to 9-year-old children, ranging from 23% in Calabazo and Umandita villages to 19% in Seywiaka.

Outbreaks, as reflected by the detection of closely related variants of the same rhinovirus genotypes, were detected in both the most isolated (4 cases of rhinovirus A1B in Umandita) and the least isolated (2 cases of rhinovirus B70 in Calabazo) villages. Two rhinovirus A22 infections in isolated Seywiaka were also very closely related and likely were epidemiologically linked. Because rhinovirus infections are of short duration, it seems likely that each of these clusters resulted from recent introductions within these communities.

The origin of sequence reads from viral clades not known to infect humans, namely, *Parvoviridae* genus (densovirus), *Partitiviridae*, *Dicistroviridae*, *Iflaviridae*, and CRESS-DNA viruses, remains unknown, but deposition onto nasal mucosal surfaces from environmental sources such as the ambient air remains a likely possibility. Possible sources for such viruses include plants and fungi for the partitiviruses and insects for the dicistroviruses, iflaviviruses, and densoparvoviruses. The origins of CRESS-DNA viral genomes are unknown. The detection of only 2 reads of canine bocavirus, a virus reported in dogs as well as cats (29–33), might similarly reflect environmental contamination from local pets.

More frequent infections with anelloviruses were detected in the more isolated villages of Seywiaka and Umandita. Anellovirus concentrations in blood are highly dependent on the host's immunocompetence, and viral titers have been shown to increase in febrile patients (34), immunosuppressed transplant recipients (35–37), and AIDS patients (38–43). The higher rates of detectable anellovirus infections in the most isolated villages may thus reflect generally weaker immunocompetence, leading to more readily detectable anelloviruses. The converse relationship was found for papillomaviruses, which were more commonly detected in the most exposed village of Calabazo (21 distinct infections), compared with the more isolated villages (10 and 6 infections in Seywiaka and Umandita, respectively). Carcinogenic papillomaviruses were not detected. The number of papillomavirus infections thus appears to correlate with the amount of exposure to people from outside the villages, and *Papillomaviridae* was the only virus family for which geographic isolation was associated with reduced viral diversity. Whether papillomavirus infections are consistently less prevalent in other small isolated villages, relative to more connected or larger populations, remains to be tested and confirmed.

We show here that children from both connected and highly isolated villages in northern Colombia carry diverse human viruses in their nasal mucosa and that extreme geographic and cultural isolation did not result in a general reduction in viral diversity. A recent study of fecal viromes in children also showed that extreme isolation in Venezuelan Amazon villages did not reduce the diversity of their enteric human viruses (4). The genetically very close similarities between the human viruses reported here or in the prior enteric virome study (4) and genomes in GenBank (derived from much larger, interconnected, populations) points to frequent virus introduction into such villages despite their high degree of geographic isolation.

Rhinovirus transmission clusters could also be readily detected within these small villages (Fig. 3). Closely related enteric picornaviruses (and caliciviruses) were also

identified in fecal samples from children in highly isolated Amazonian villages (4). These observations likely reflect ongoing outbreaks of enteric and respiratory tract viruses among epidemiologically close children within very small villages.

Overall, these results support our conclusion that the reach of common human viruses, both enteric (4) and here respiratory, commonly extends to some of the most geographically isolated populations still in existence.

MATERIALS AND METHODS

Study population and study design. Nasal swab samples were collected between September 2016 and February 2017 from children with no apparent clinical signs who were enrolled in an influenza surveillance study located in three different villages in the Magdalena Department of Colombia, by the Caribbean Sea (Fig. 1). Nasal swabs from 21 children from each village were analyzed, totaling 63 samples from 34 girls and 29 boys (Table 1). Dry sterile swabs (nylon flocked swabs; Fisher) were used in both nostrils and stored in 1 ml universal transport medium (Quidel). Samples were kept on ice for 4 to 7 days and then stored at −80°C.

Viral metagenomics. To reduce possible batch effects, samples from the three locations were processed in an interdigitated manner (first sample from village 1, 2, 3, then second sample from village 1, 2, 3, etc.) using two Illumina MiSeq runs. Individual swab supernatants (150 μl) were filtered using a 0.45-μm filter (Millipore). The filtrates were treated with a mixture of DNases (Turbo DNase [Ambion], Baseline-ZERO DNase [Epicentre], and Benzonase [Novagen]) and RNase (Fermentas) at 37°C for 90 min to enrich for viral capsid-protected nucleic acids, which were then extracted using a Maxwell 16 automated extractor (Promega) (44). Random reverse transcription-PCR and the Nextera XT sample preparation kit (Illumina) were used to generate a library for Illumina MiSeq sequencing (2 × 250 bases) with dual barcoding, as described previously (45).

Bioinformatic analysis. An in-house analysis pipeline was used to analyze sequence data. Before analysis, raw data were preprocessed by subtracting human and bacterial sequences, duplicate sequences, and low-quality reads. Following *de novo* assembly using the Ensemble program (46), both contigs and singlet viral sequences were then analyzed by using a translated protein sequence similarity search (BLASTx v.2.2.7) for all annotated viral proteins available in GenBank. Candidate viral hits were then compared to a nonvirus nonredundant protein database to remove false-positive viral hits. To align reads and contigs to reference viral genomes from GenBank and to generate complete or partial genome sequences, the Geneious R10 program was used. To plot read numbers for different viruses, the numbers of reads with BLASTx E scores of $<10^{-10}$ for named viruses per 1 million reads were $\log_{10}$ transformed to determine the size of the colored circles in Fig. 2, 4, and 6 using Excel.

Phylogenetic analyses. Phylogenetic trees were constructed from VP4-VP2 nucleotide sequences for rhinoviruses and amino acid sequences for papillomaviruses. Evolutionary analyses were conducted in MEGA6 using the maximum likelihood method, based on the general time reversible model (47, 48).

Statistical methods. To evaluate the proportional distribution of viral types among villages, a nonparametric one-way analysis of variance was performed using the Kruskal-Wallis test, with ties and an *a priori* statistical significance level set at a P value of <0.05. Stata/MP 15.1 (StataCorp, College Station, TX) was used for statistical analysis. The Kruskal-Wallis equality of population rank test was performed using 2 degrees of freedom.

Ethics statement. Studies were approved by the Indigenous Health Council, the Tropical Health Foundation ethics committee, and St. Jude Children's Research Hospital institutional review board. The investigators ensured that the study was conducted in full conformity with the principles set forth in The Belmont Report (49) and codified (50–53). The investigators' institutions hold current federalwide assurances issued by the Office of Human Research Protection for federally funded research.

Data availability. Sequence data are available under BioProject accession number PRJNA530270 and GenBank accession numbers MK501733 to MK501745.

ACKNOWLEDGMENTS

Funding sources consisted of support from Vitalant Research Institute to E.D. and from ALSAC and NIAID contract HHSN272201400006C to S.S.-C.

REFERENCES

1. Black FL. 1975. Infectious diseases in primitive societies. Science 187:515–518. https://doi.org/10.1126/science.163483.
2. O'Fallon BD, Fehren-Schmitz L. 2011. Native Americans experienced a strong population bottleneck coincident with European contact. Proc Natl Acad Sci U S A 108:20444–20448. https://doi.org/10.1073/pnas.1112563108.
3. Walker RS, Sattenspiel L, Hill KR. 2015. Mortality from contact-related epidemics among indigenous populations in Greater Amazonia. Sci Rep 5:14032. https://doi.org/10.1038/srep14032.
4. Siqueira JD, Dominguez-Bello MG, Contreras M, Lander O, Caballero-Arias H, Xutao D, Noya-Alarcon O, Delwart E. 2018. Complex virome in feces from Amerindian children in isolated Amazonian villages. Nat Commun 9:4270. https://doi.org/10.1038/s41467-018-06502-9.
5. Altan E, Dib JC, Gulloso AR, Escribano Juandigua D, Deng X, Bruhn R, Hildebrand K, Freiden P, Yamamoto J, Schultz-Cherry S, Delwart E. 2019. Effect of geographic isolation on the nasal virome of indigenous children. bioRxiv https://doi.org/10.1101/616342.
6. Feng H, Shuda M, Chang Y, Moore PS. 2008. Clonal integration of a polyomavirus in human Merkel cell carcinoma. Science 319:1096–1100. https://doi.org/10.1126/science.1152586.
7. Siebrasse EA, Reyes A, Lim ES, Zhao G, Mkakosya RS, Manary MJ, Gordon JI, Wang D. 2012. Identification of MW polyomavirus, a novel polyoma-

virus in human stool. J Virol 86:10321–10326. https://doi.org/10.1128/JVI.01210-12.
8. Yu G, Greninger AL, Isa P, Phan TG, Martinez MA, de la Luz Sanchez M, Contreras JF, Santos-Preciado JI, Parsonnet J, Miller S, DeRisi JL, Delwart E, Arias CF, Chiu CY. 2012. Discovery of a novel polyomavirus in acute diarrheal samples from children. PLoS One 7:e49449. https://doi.org/10.1371/journal.pone.0049449.
9. Lim ES, Reyes A, Antonio M, Saha D, Ikumapayi UN, Adeyemi M, Stine OC, Skelton R, Brennan DC, Mkakosya RS, Manary MJ, Gordon JI, Wang D. 2013. Discovery of STL polyomavirus, a polyomavirus of ancestral recombinant origin that encodes a unique T antigen by alternative splicing. Virology 436:295–303. https://doi.org/10.1016/j.virol.2012.12.005.
10. Yang J, Yang F, Ren L, Xiong Z, Wu Z, Dong J, Sun L, Zhang T, Hu Y, Du J, Wang J, Jin Q. 2011. Unbiased parallel detection of viral pathogens in clinical samples by use of a metagenomic approach. J Clin Microbiol 49:3463–3469. https://doi.org/10.1128/JCM.00273-11.
11. Lysholm F, Wetterbom A, Lindau C, Darban H, Bjerkner A, Fahlander K, Lindberg AM, Persson B, Allander T, Andersson B. 2012. Characterization of the viral microbiome in patients with severe lower respiratory tract infections, using metagenomic sequencing. PLoS One 7:e30875. https://doi.org/10.1371/journal.pone.0030875.
12. Bal A, Pichon M, Picard C, Casalegno JS, Valette M, Schuffenecker I, Billard L, Vallet S, Vilchez G, Cheynet V, Oriol G, Trouillet-Assant S, Gillet Y, Lina B, Brengel-Pesce K, Morfin F, Josset L. 2018. Quality control implementation for universal characterization of DNA and RNA viruses in clinical respiratory samples using single metagenomic next-generation sequencing workflow. BMC Infect Dis 18:537. https://doi.org/10.1186/s12879-018-3446-5.
13. Graf EH, Simmon KE, Tardif KD, 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. https://doi.org/10.1128/JCM.03060-15.
14. Schlaberg R, Queen K, Simmon K, Tardif K, Stockmann C, Flygare S, Kennedy B, Voelkerding K, Bramley A, Zhang J, Eilbeck K, Yandell M, Jain S, Pavia AT, Tong S, Ampofo K. 2017. Viral pathogen detection by metagenomics and pan-viral group polymerase chain reaction in children with pneumonia lacking identifiable etiology. J Infect Dis 215:1407–1415. https://doi.org/10.1093/infdis/jix148.
15. Xu L, Zhu Y, Ren L, Xu B, Liu C, Xie Z, Shen K. 2017. Characterization of the nasopharyngeal viral microbiome from children with community-acquired pneumonia but negative for Luminex xTAG respiratory viral panel assay detection. J Med Virol 89:2098–2107. https://doi.org/10.1002/jmv.24895.
16. Lewandowska DW, Schreiber PW, Schuurmans MM, Ruehe B, Zagordi O, Bayard C, Greiner M, Geissberger FD, Capaul R, Zbinden A, Boni J, Benden C, Mueller NJ, Trkola A, Huber M. 2017. Metagenomic sequencing complements routine diagnostics in identifying viral pathogens in lung transplant recipients with unknown etiology of respiratory infection. PLoS One 12:e0177340. https://doi.org/10.1371/journal.pone.0177340.
17. Zinter MS, Dvorak CC, Mayday MY, Iwanaga K, Ly NP, McGarry ME, Church GD, Faricy LE, Rowan CM, Hume JR, Steiner ME, Crawford ED, Langelier C, Kalantar K, Chow ED, Miller S, Shimano K, Melton A, Yanik GA, Sapru A, DeRisi JL. 2019. Pulmonary metagenomic sequencing suggests missed infections in immunocompromised children. Clin Infect Dis 68:1847–1855. https://doi.org/10.1093/cid/ciy802.
18. Blauwkamp TA, Thair S, Rosen MJ, Blair L, Lindner MS, Vilfan ID, Kawli T, Christians FC, Venkatasubrahmanyam S, Wall GD, Cheung A, Rogers ZN, Meshulam-Simon G, Huijse L, Balakrishnan S, Quinn JV, Hollemon D, Hong DK, Vaughn ML, Kertesz M, Bercovici S, Wilber JC, Yang S. 2019. Analytical and clinical validation of a microbial cell-free DNA sequencing test for infectious disease. Nat Microbiol 4:663–674. https://doi.org/10.1038/s41564-018-0349-6.
19. Gu W, Miller S, Chiu CY. 2019. Clinical metagenomic next-generation sequencing for pathogen detection. Annu Rev Pathol 14:319–338. https://doi.org/10.1146/annurev-pathmechdis-012418-012751.
20. Parize P, Muth E, Richaud C, Gratigny M, Pilmis B, Lamamy A, Mainardi JL, Cheval J, de Visser L, Jagorel F, Ben Yahia L, Bamba G, Dubois M, Join-Lambert O, Leruez-Ville M, Nassif X, Lefort A, Lanternier F, Suarez F, Lortholary O, Lecuit M, Eloit M. 2017. Untargeted next-generation sequencing-based first-line diagnosis of infection in immunocompromised adults: a multicentre, blinded, prospective study. Clin Microbiol Infect 23:574.e1–574.e6. https://doi.org/10.1016/j.cmi.2017.02.006.
21. Schlaberg R, Chiu CY, Miller S, Procop GW, Weinstock G. 2017. Validation of metagenomic next-generation sequencing tests for universal pathogen detection. Arch Pathol Lab Med 141:776–786. https://doi.org/10.5858/arpa.2016-0539-RA.
22. Wylie KM, Mihindukulasuriya KA, Zhou Y, Sodergren E, Storch GA, Weinstock GM. 2014. Metagenomic analysis of double-stranded DNA viruses in healthy adults. BMC Biol 12:71. https://doi.org/10.1186/s12915-014-0071-7.
23. Goggin RK, Bennett CA, Bassiouni A, Bialasiewicz S, Vreugde S, Wormald PJ, Psaltis AJ. 2018. Comparative viral sampling in the sinonasal passages: different viruses at different sites. Front Cell Infect Microbiol 8:334. https://doi.org/10.3389/fcimb.2018.00334.
24. Bogaert D, Keijser B, Huse S, Rossen J, Veenhoven R, van Gils E, Bruin J, Montijn R, Bonten M, Sanders E. 2011. Variability and diversity of nasopharyngeal microbiota in children: a metagenomic analysis. PLoS One 6:e17035. https://doi.org/10.1371/journal.pone.0017035.
25. Fry AM, Lu X, Olsen SJ, Chittaganpitch M, Sawatwong P, Chantra S, Baggett HC, Erdman D. 2011. Human rhinovirus infections in rural Thailand: epidemiological evidence for rhinovirus as both pathogen and bystander. PLoS One 6:e17780. https://doi.org/10.1371/journal.pone.0017780.
26. Iwane MK, Prill MM, Lu X, Miller EK, Edwards KM, Hall CB, Griffin MR, Staat MA, Anderson LJ, Williams JV, Weinberg GA, Ali A, Szilagyi PG, Zhu Y, Erdman DD. 2011. Human rhinovirus species associated with hospitalizations for acute respiratory illness in young US children. J Infect Dis 204:1702–1710. https://doi.org/10.1093/infdis/jir634.
27. Singleton RJ, Bulkow LR, Miernyk K, DeByle C, Pruitt L, Hummel KB, Bruden D, Englund JA, Anderson LJ, Lucher L, Holman RC, Hennessy TW. 2010. Viral respiratory infections in hospitalized and community control children in Alaska. J Med Virol 82:1282–1290. https://doi.org/10.1002/jmv.21790.
28. van Benten I, Koopman L, Niesters B, Hop W, van Middelkoop B, de Waal L, van Drunen K, Osterhaus A, Neijens H, Fokkens W. 2003. Predominance of rhinovirus in the nose of symptomatic and asymptomatic infants. Pediatr Allergy Immunol 14:363–370. https://doi.org/10.1034/j.1399-3038.2003.00064.x.
29. Kapoor A, Mehta N, Dubovi EJ, Simmonds P, Govindasamy L, Medina JL, Street C, Shields S, Lipkin WI. 2012. Characterization of novel canine bocaviruses and their association with respiratory disease. J Gen Virol 93:341–346. https://doi.org/10.1099/vir.0.036624-0.
30. Piewbang C, Jo WK, Puff C, Ludlow M, van der Vries E, Banlunara W, Rungsipipat A, Kruppa J, Jung K, Techangamsuwan S, Baumgärtner W, Osterhaus ADME. 2018. Canine bocavirus type 2 infection associated with intestinal lesions. Vet Pathol 55:434–441. https://doi.org/10.1177/0300985818755253.
31. Bodewes R, Lapp S, Hahn K, Habierski A, Forster C, Konig M, Wohlsein P, Osterhaus AD, Baumgartner W. 2014. Novel canine bocavirus strain associated with severe enteritis in a dog litter. Vet Microbiol 174:1–8. https://doi.org/10.1016/j.vetmic.2014.08.025.
32. Niu J, Yi S, Wang H, Dong G, Zhao Y, Guo Y, Dong H, Wang K, Hu G. 2019. Complete genome sequence analysis of canine bocavirus 1 identified for the first time in domestic cats. Arch Virol 164:601–605. https://doi.org/10.1007/s00705-018-4096-z.
33. Lau SK, Woo PC, Yeung HC, Teng JL, Wu Y, Bai R, Fan RY, Chan KH, Yuen KY. 2012. Identification and characterization of bocaviruses in cats and dogs reveals a novel feline bocavirus and a novel genetic group of canine bocavirus. J Gen Virol 93:1573–1582. https://doi.org/10.1099/vir.0.042531-0.
34. McElvania TeKippe E, Wylie KM, Deych E, Sodergren E, Weinstock G, Storch GA. 2012. Increased prevalence of anellovirus in pediatric patients with fever. PLoS One 7:e50937. https://doi.org/10.1371/journal.pone.0050937.
35. Young JC, Chehoud C, Bittinger K, Bailey A, Diamond JM, Cantu E, Haas AR, Abbas A, Frye L, Christie JD, Bushman FD, Collman RG. 2015. Viral metagenomics reveal blooms of anelloviruses in the respiratory tract of lung transplant recipients. Am J Transplant 15:200–209. https://doi.org/10.1111/ajt.13031.
36. De Vlaminck I, Khush KK, Strehl C, Kohli B, Luikart H, Neff NF, Okamoto J, Snyder TM, Cornfield DN, Nicolls MR, Weill D, Bernstein D, Valantine HA, Quake SR. 2013. Temporal response of the human virome to immunosuppression and antiviral therapy. Cell 155:1178–1187. https://doi.org/10.1016/j.cell.2013.10.034.
37. Blatter JA, Sweet SC, Conrad C, Danziger-Isakov LA, Faro A, Goldfarb SB, Hayes D, Jr, Melicoff E, Schecter M, Storch G, Visner GA, Williams NM, Wang D. 2018. Anellovirus loads are associated with outcomes in pediatric lung transplantation. Pediatr Transplant 22:e13069. https://doi.org/10.1111/petr.13069.
38. Li L, Deng X, Linsuwanon P, Bangsberg D, Bwana MB, Hunt P, Martin JN,

Deeks SG, Delwart E. 2013. AIDS alters the commensal plasma virome. J Virol 87:10912–10915. https://doi.org/10.1128/JVI.01839-13.
39. Sherman KE, Rouster SD, Feinberg J. 2001. Prevalence and genotypic variability of TTV in HIV-infected patients. Dig Dis Sci 46:2401–2407. https://doi.org/10.1023/A:1012307416122.
40. Touinssi M, Gallian P, Biagini P, Attoui H, Vialettes B, Berland Y, Tamalet C, Dhiver C, Ravaux I, De Micco P, De Lamballerie X. 2001. TT virus infection: prevalence of elevated viraemia and arguments for the immune control of viral load. J Clin Virol 21:135–141. https://doi.org/10.1016/S1386-6532(01)00157-3.
41. Shibayama T, Masuda G, Ajisawa A, Takahashi M, Nishizawa T, Tsuda F, Okamoto H. 2001. Inverse relationship between the titre of TT virus DNA and the CD4 cell count in patients infected with HIV. AIDS 15:563–570. https://doi.org/10.1097/00002030-200103300-00004.
42. 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. https://doi.org/10.1002/jmv.20756.
43. Christensen JK, Eugen-Olsen J, Sørensen M, Ullum H, Gjedde SB, Pedersen BK, Nielsen JO, Krogsgaard K. 2000. Prevalence and prognostic significance of infection with TT virus in patients infected with human immunodeficiency virus. J Infect Dis 181:1796–1799. https://doi.org/10.1086/315440.
44. Victoria JG, Kapoor A, Li L, Blinkova O, Slikas B, Wang C, Naeem A, Zaidi S, Delwart E. 2009. Metagenomic analyses of viruses in stool samples from children with acute flaccid paralysis. J Virol 83:4642–4651. https://doi.org/10.1128/JVI.02301-08.
45. Li L, Deng X, Mee ET, Collot-Teixeira S, Anderson R, Schepelmann S, Minor PD, Delwart E. 2015. Comparing viral metagenomics methods using a highly multiplexed human viral pathogens reagent. J Virol Methods 213:139–146. https://doi.org/10.1016/j.jviromet.2014.12.002.
46. Deng X, Naccache SN, Ng T, Federman S, Li L, Chiu CY, Delwart EL. 2015. An ensemble strategy that significantly improves de novo assembly of microbial genomes from metagenomic next-generation sequencing data. Nucleic Acids Res 43:e46. https://doi.org/10.1093/nar/gkv002.
47. Nei M, Kumar S. 2000. Molecular evolution and phylogenetics. Oxford University Press, New York, NY.
48. Tamura K, Stecher G, Peterson D, Filipski A, Kumar S. 2013. MEGA6: Molecular Evolutionary Genetics Analysis version 6.0. Mol Biol Evol 30:2725–2729. https://doi.org/10.1093/molbev/mst197.
49. National Commission for the Protection of Human Subjects of Biomedical and Behavioral Research. 1979. The Belmont Report: ethical principles and guidelines for the protection of human subjects of research. National Commission for the Protection of Human Subjects of Biomedical and Behavioral Research, Washington, DC. https://www.hhs.gov/ohrp/sites/default/files/the-belmont-report-508c_FINAL.pdf.
50. Code of Federal Regulations. 2018. Title 45. Public welfare. Part 46. Protection of human subjects. 45 CFR §46. https://www.ecfr.gov/cgi-bin/retrieveECFR?gp=&SID=83cd09e1c0f5c6937cd9d7513160fc3f&pitd=20180719&n=pt45.1.46&r=PART&ty=HTML.
51. Code of Federal Regulations. 2018. Title 21. Food and drugs. Chapter I. Food and Drug Administration. Subchapter A. General. Part 50. Protection of human subjects. 21 CFR §50. https://www.accessdata.fda.gov/scripts/cdrh/cfdocs/cfcfr/CFRSearch.cfm?CFRPart=50.
52. Code of Federal Regulations. 2018. Title 21. Food and drugs. Chapter I. Food and Drug Administration. Subchapter A. General. Part 56.Institutional review boards. 21 CFR §56. https://www.accessdata.fda.gov/scripts/cdrh/cfdocs/cfcfr/CFRSearch.cfm?CFRPart=56.
53. Federal Register. 1997. International Conference on Harmonisation: good clinical practice: consolidated guideline. Fed Reg 62:25692–25709. https://www.govinfo.gov/content/pkg/FR-1997-05-09/pdf/97-12138.pdf.