

Identifying viral infections through metagenomic Next Generation Sequencing of undiagnosed respiratory fevers in Madagascar (2014-2019)

Sophie Lockwood

slockwood@berkeley.edu

University of California Berkeley

Hafaliana Christian Ranaivoson

University of California Berkeley

Tsiry Hasina Randriambolamanantsoa

Institut Pasteur de Madagascar

Noroso Razanajatovo

Institut Pasteur de Madagascar

Vololoniaina Raharinosy

Institut Pasteur de Madagascar

Vida Ahyong

Chan Zuckerberg Biohub

Jean-Michel Heraud

Institut Pasteur de Madagascar

Philippe Dussart

Institut Pasteur de Madagascar

Vincent Lacoste

Institut Pasteur de Madagascar

Cara E. Brook

University of California Berkeley

Research Article

Keywords: respiratory viruses, mNGS, public health surveillance

Posted Date: March 11th, 2026

DOI: <https://doi.org/10.21203/rs.3.rs-9086867/v1>

License: © ⓘ This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

Additional Declarations: The authors declare no competing interests.

Abstract

Background

Respiratory illness contributes to substantial global morbidity and mortality. In Madagascar, an island nation off the southeastern coast of the African continent, hospital-based public health surveillance for respiratory pathogens screens for common respiratory viruses. However, many cases remain undiagnosed.

Methods

We conducted metagenomic Next Generation Sequencing (mNGS) to identify the pathogen profile of 102 undiagnosed febrile patients who presented to public hospitals with respiratory symptoms and screened negative on a 14-virus multiplex RT-qPCR. We analyzed the diversity of the respiratory microbiome of each patient from mNGS data and identified viral infections potentially linked to undiagnosed fever. We assembled whole genome consensus sequences of viruses with sufficient read depth and coverage, characterized each phylogenetically, and identified any discrepancies with the primers used in the multiplex RT-qPCR panel. Finally, we compared all whole genome sequences against publicly available global databases in a phylogenetic analysis.

Results

We identified evidence of infection by a wide range of known human viruses in approximately two thirds (64.7%) of study participants from nine different families of viruses and generated 30 complete or nearly complete consensus sequences of known respiratory viruses including orthopneumoviruses, metapneumoviruses, rhinoviruses, coronaviruses, parainfluenza virus, and bocaparvovirus. mNGS-attributed evidence of infection was predominantly due to orthopneumovirus (also called respiratory syncytial virus [RSV]; $n = 24$) and enterovirus ($n = 20$) detections, despite previous negative RT-qPCR results for these and other common respiratory viruses for most patients investigated. Finally, phylogenetic analysis identified two distinct phylogenetic clusters of RSV subtype A, suggesting local transmission following distinct international introductions for this virus.

Conclusion

mNGS provides a sensitive pan-pathogenic tool for virus detection. We demonstrate the diversity of viruses driving respiratory fevers in Madagascar, emphasize the importance and relevance of the existing respiratory surveillance in the country, and highlight the interconnectedness of regional respiratory infection dynamics with global networks of respiratory pathogen transmission.

BACKGROUND

Respiratory illness is responsible for substantial global morbidity and mortality. A leading cause of death due to infectious disease, respiratory infections result in the highest burden in lower- and middle-income countries and among children under age five¹⁻³. Viruses frequently cause acute respiratory infections, although bacterial and parasitic infections can also result in respiratory symptoms^{2,4}.

Madagascar, an island nation of nearly 30 million people off the southeastern coast of the African continent, experiences a high burden of respiratory disease⁵, resulting from a high diversity of pathogens⁶. In Madagascar, viral infections contribute to approximately one fifth of hospital visits for non-specific febrile disease⁶ and more than half of hospital visits for respiratory symptoms in both urban⁷ and rural⁸ settings. Madagascar is a relatively closed island population with temperate to tropical climates where respiratory virus dynamics likely reflect both local circulation and repeated introductions from outside sources. Respiratory pathogen surveillance in Madagascar has permitted the identification of temporal⁷, climatic⁹ and geographic¹⁰ patterns of acute respiratory viral infections. Interestingly, the seasonal patterns observed in influenza and parainfluenza transmission in Madagascar are asynchronous with global patterns, even those of countries connected by relatively frequent travel^{11,12}. Further, Madagascar hosts unique viruses relative to global trends, with distinct circulating strains of influenza virus in comparison to outside strains used to determine annual vaccine recommendations¹³. Madagascar's unusual geographic position and infection seasonality underscore the importance of investigating patterns in respiratory illness in the country.

The Ministry of Health of Madagascar, in collaboration with the Institut Pasteur de Madagascar (IPM), established a nationwide sentinel syndromic surveillance system for febrile illness in 2007, expanding coverage from surveillance that had previously only been conducted in the capital, Antananarivo¹⁴. The system included screening for influenza-like illnesses via nasopharyngeal sampling and provided a useful tool for monitoring abnormal febrile illness and tracking respiratory pathogen outbreaks (e.g. the influenza A/(H1N1)pdm09 and SARS-CoV-2 viruses)¹⁵. Select hospitals in the country screen patients reporting with respiratory symptoms and fever for influenzaviruses, respiratory syncytial virus (RSV), and, since 2020, SARS-CoV-2 by testing nasopharyngeal swabs with real-time quantitative polymerase chain reaction (RT-qPCR). During outbreak investigations, patients are screened for a larger suite of common pathogens with a 14-virus multiplex RT-qPCR panel that includes strains of parainfluenza viruses, coronaviruses, rhinoviruses, influenza viruses, adenoviruses, and bocaviruses⁷. However, 25% of patients reporting to Malagasy health centers with influenza-like illness between 2008 and 2009⁷ and 80% of symptomatic SARS-CoV-2 negative patients in hospitals and communities between 2020 and 2022¹⁶ remained undiagnosed after RT-qPCR screening for these 14 viruses.

Metagenomic Next Generation Sequencing (mNGS) enables identification of pathogens of unknown origin, by sequencing all nucleic acid in a single sample, allowing for *de novo* genome assembly of both novel and previously described microbes of viral, bacterial, fungal, or eukaryotic origin that may be

associated with febrile illness. As an unbiased method, mNGS is an effective surveillance tool for non-specific conditions such as febrile disease with respiratory symptoms. The utility of mNGS in identifying etiologic agents of disease has been previously demonstrated in studies of febrile illness in low-resource settings in Cambodia¹⁷ and Uganda^{18,19}, to investigate the etiology of undiagnosed dengue-like symptoms in Nicaragua²⁰, and to provide diagnosis of fever of unknown cause in travelers to the United Kingdom²¹ and the Netherlands²². As a tool, mNGS has led to the identification of novel human viruses^{23–25}, including those of zoonotic origin. Molecular epidemiology played a critical role in the COVID-19 pandemic, and led to a global expansion of sequencing capacity, although regional disparities in sequence reporting persist²⁶. In low and middle income countries (LMICs), mNGS platforms for pathogen surveillance offer unique benefits and face additional hurdles^{26,27}. A pan-pathogenic detection tool, mNGS is useful when samples are limited in quantity and has the potential to enhance biological surveillance in regions where cross-species transmission risk is high. However, establishing the laboratory and bioinformatics pipeline necessary to carry out mNGS is an ongoing challenge as it requires advanced training, costly equipment and reagents, and reliable cold-chain, electricity, and internet infrastructure.

Here, we report on five years of sampling with subsequent mNGS sequencing and analysis of undiagnosed febrile patients reporting with respiratory symptoms to public hospitals in Madagascar. We find evidence of viral infection in over nearly two thirds of cases and generate consensus genomes of respiratory viruses known to circulate in Madagascar, including human metapneumoviruses, rhinoviruses, seasonal coronaviruses, and orthopneumoviruses (RSV). We report factors associated with viral detection, contextualize the identified viruses, and assess primer compatibility between the RT-qPCR surveillance panel and the sequences. Finally, we conduct phylogeographic analyses to estimate the number of introductions of RSV subtype A, the only pathogen detected in this study with sufficient frequency to support a broader analysis.

METHODS

Study population and sample collection. Patients reporting to public hospitals in Madagascar with fever between 2014 and 2019 were screened according to syndromic surveillance guidelines. Nasopharyngeal swabs were collected from febrile patients with cough and were transported to the IPM laboratory in Antananarivo where they underwent RNA extraction and screening by real-time RT-PCR for 14 routine human respiratory viruses: human parainfluenzaviruses 1, 2, and 3, human coronaviruses OC43, 229E, HKU1, and NL63, human rhinovirus (HRV), influenzaviruses A and B, human metapneumovirus, RSV, human adenovirus, and human bocavirus (see **Table S1** for details). Nasopharyngeal samples analyzed in this study screened negative on this pathogen panel and later underwent mNGS. Eleven nasopharyngeal swabs from individuals who screened positive for RSV, HRV, and influenzavirus were included as positive controls, and several healthy individuals provided nasopharyngeal samples for negative controls.

mNGS sample processing, sequencing, and bioinformatics. Following sample collection, RNA extraction was conducted using the Zymo Quick DNA/RNA Microprep Plus Kit (Zymo Research, Irvine, CA) according to the manufacturer's instructions and including the reaction for DNase digestion. mNGS library preparation was completed using the NEBNext Ultra RNA II Library Prep Kit (New England BioLabs, Beverly, MA, USA) also per manufacturer instructions. Libraries were pooled and assessed for quality with electrophoresis (High-Sensitivity DNA Kit and Agilent Bioanalyzer; Agilent Technologies, Santa Clara, CA, USA) and small-scale sequencing (2 x 146 bp) on an Illumina iSeq machine (Illumina, San Diego, CA, USA) to permit equimolar pooling of individual libraries that proceeded to large-scale paired end mNGS (2 x 146 bp) on an Illumina NextSeq sequencing machine (Illumina, San Diego, CA, USA). Using the Chan Zuckerberg Infectious Disease (CZID) bioinformatics pipeline²⁸, raw sequence reads were progressively filtered for host reads, read quality, and duplicates (see **Figure S1**). *De novo* genome construction was conducted on the resulting filtered non-host reads and microbial taxa were identified based on BLAST nucleotide and protein comparison to NCBI databases²⁹. Where possible, consensus genomes were generated using the CZID pipeline. Complete and nearly complete consensus sequences were added to GenBank (accession numbers: PX467117-PX467146) and raw data was uploaded to the NCBI Sequence Read Archive (SRA) under BioProject number PRJNA1328692.

mNGS analysis. We sought to identify potential etiologies of febrile disease. We followed previous methods¹⁷ to consider identified viruses to be significant infections potentially causative of febrile disease if reads generated from mNGS produced a matched alignment of at least 50 base pairs in length, with a depth of at least 10 reads per million, and an e-value of 1e-10 or lower. We further required that reads mapping to target taxa generated a z-score of at least 1 following comparison with the background microbiome from two healthy human controls. Both e-values and z-scores were generated using the CZID pipeline²⁸.

Simple statistics. We analyzed patient demographic characteristics and compared characteristics between samples with and without evidence of viral infection using Chi squared and Fisher's exact tests for categorical variables and the Wilcoxon rank-sum test for differences in age. We conducted simple linear regression to evaluate association between the number of patients with and without evidence of viral infection sampled in each month during the study interval. All analyses were conducted using R version 4.4.1.

Microbial diversity. We assessed microbial diversity for each sample by calculating Shannon diversity indices for each sample in R version 4.4.1 using the Vegan package (<https://vegandevs.github.io/vegan/>). We compared samples with and without evidence of viral infection as well as to samples from two healthy controls. Differences in the mean Shannon diversity indices between groups were evaluated by ANOVA.

Analysis of primer match. To identify compatibility between sequences assembled from Malagasy patients to the primer set included in the 14-pathogen surveillance screening, we downloaded all primer sequences (described in previously published work⁷). We aligned consensus genomes from this study

with these primer sequences using GeneiousPrime and calculated percent pairwise identity and number of divergent base pairs between the primer set and each consensus genome corresponding to a pathogen included in the surveillance panel.

Sequence analysis. We first conducted manual BLASTn comparison of each consensus genome with the NCBI database. Then, we conducted two main phylogenetic analyses: 1) full-genome maximum likelihood phylogenies for the family or genus of the viruses for which consensus genomes were obtained and 2) characterization of potential sources of introduction of RSV subtype A lineages, the only pathogen identified in our data with sufficient frequency for such an approach.

In the first phylogenetic analysis, for each of the assembled consensus genomes, we combined our sequences with all reference sequences available in NCBI Virus corresponding to the same viral family, sub-family or genus of the target sequences of interest. As such, we combined our two Madagascar *Coronaviridae* sequences with 74 reference sequences (Taxid: 11118) and the reference sequence for porcine torovirus (accession number: NC_022787.1) as an outgroup. We combined the sole Madagascar *Parvovirinae* sequence with 48 reference sequences (Taxid: 40119) and the reference sequence for Danaus plexippus iteravirus (accession number: NC_023842.1) as an outgroup. To make the *Paramyxoviridae* dataset, we combined the unique Madagascar sequence with 92 reference sequences (Taxid: 11158) and included the reference sequence for Sun Coast virus (accession number: NC_025345.1) as an outgroup. We combined 15 Madagascar *Pneumoviridae* sequences with 11 reference sequences (Taxid: 11244) and the Sun Coast virus (accession number: NC_025345.1) as an outgroup. We finally combined 11 Madagascar *Enterovirus* sequences with 31 reference sequences (Taxid: 12059) and the reference sequence for porcine sapelovirus (accession number NC_003987.1) as an outgroup. Accession numbers for all reference sequences included in all phylogenies are listed in **Table S2**.

Sequences from each of these datasets were aligned using default parameters in the online software MAFFT v7.511³⁰ and visually verified in AliView³¹. Alignments were tested for the best fit nucleotide substitution model using ModelTest-NG³². Phylogenetic trees were inferred by maximum likelihood using the best-fitting model from ModelTest-NG, in all cases a general time reversible model with gamma distributed rates and invariant sites (GTR + I+G4), in RAxML³³.

To contextualize RSV subtype A lineages in Madagascar with concurrently circulating global strains, we generated a comprehensive sequence dataset by compiling all the RSV-A consensus genomes obtained from this study with all publicly available complete and near-complete RSV-A genomes with available collection location and date corresponding to anytime during or before the same sampling interval (2014–2019) from NCBI Virus and GISAID³⁴ databases. The reference sequence of RSV subtype B (accession number NC_001781.1) was included as an outgroup. We then subsampled the dataset to a maximum four sequences per country per month during the interval of interest using a Nextstrain pipeline³⁵. Model selection and phylogenetic tree reconstruction were both conducted using IQ-TREE 2.0.7³⁶.

All tree visualizations were conducted using ggtree³⁷ in R version 4.4.1.

RESULTS

Clinical, demographic, and geographic characteristics. We analyzed 102 nasopharyngeal swabs collected from febrile patients in Antananarivo and Mahajanga provinces across Madagascar between 2014 and 2019 (Table 1), 11 of which were previously diagnosed samples by RT-qPCR included as positive controls. Two samples were collected from healthy controls in 2019. Febrile patients ranged in age from infants to adults but skewed towards pediatric patients (median age: 2 years, IQR: 0–29.75 years). The median patient age was younger among individuals with evidence of viral infection (with viral infection: 1, IQR: 0–5.8 years; without viral infection: 31 years, IQR: 4.5–55.8 years, Wilcoxon rank-sum, $p < 0.001$), a pattern that held when the analysis was repeated excluding RSV infections (Wilcoxon rank-sum, $p < 0.001$). Both males and females were included in the sample (63% male), and sex did not differ between virus-infected and uninfected patients (χ^2 , $p = 1$). Most patients providing nasopharyngeal swabs presented to hospitals in Antananarivo, the central province that includes the capital city; only one sample was provided from Mahajanga province (Fig. 1).

Table 1. Participant demographic characteristics stratified by evidence of viral infection. Evidence of viral infection was determined from analysis of the mNGS reads from nasopharyngeal samples of patients reporting with fever and respiratory symptoms to hospitals across Madagascar; a read depth of at least 10 reads per million and statistical support for pathogen identification were considered evidence of viral infection.

	Total N=102	Evidence of viral infection N=66	No evidence of viral infection N=36
Age (years)	N (%)	N (%)	N (%)
0-5	59 (57.8)	49 (74.2)	10 (27.8)
6-9	3 (2.9)	2 (3.0)	1 (2.8)
10-19	4 (3.9)	1 (1.5)	3 (8.3)
20-29	10 (9.8)	7 (10.6)	3 (8.3)
30-39	6 (5.8)	1 (1.5)	5 (13.9)
49-49	6 (5.8)	2 (3.0)	4 (11.1)
50-59	5 (4.9)	2 (3.0)	3 (8.3)
60+	9 (8.8)	2 (3.0)	7 (19.4)
Sex			
Male	64 (62.7)	41 (62.1)	23 (63.9)
Female	38 (37.3)	25 (37.8)	13 (36.1)
Location			
Antananarivo	101 (99.0)	66 (100)	35 (97.2)
Mahajanga	1 (0.9)	0 (0)	1 (2.8)

Microbes detected by mNGS. A range of bacteria, eukaryotes, and viruses were identified in the nasopharyngeal swabs (Fig. 2). Although healthy controls had the highest average Shannon diversity index relative to both febrile samples with and without evidence of viral infection, and the mean Shannon

diversity index was lowest among samples with evidence of viral infection, differences between groups were not significant (ANOVA, $p = 0.06$).

Viruses detected by mNGS. Evidence of infection with known pathogenic viruses was identified in 66 (64.7%) nasopharyngeal samples, including from nine viral families (Table 2). Evidence of multiple virus coinfection was found in 16 (24%) samples. Most viral coinfections ($n = 12$) included two viruses, three samples had coinfections with three viruses, and one sample had evidence of infection with five viruses (bocaparvovirus, mastadenovirus, betapolyomavirus, cytomegalovirus, and alphacoronavirus) (see **Table S3** for further details).

Table 2

Viruses identified by mNGS in nasopharyngeal samples that meet the threshold for evidence of viral infection. Samples were collected in public hospitals across Madagascar between 2014 and 2019, screened negative on multiplex RT-qPCR, and underwent sequencing.

Virus Species	Family	Number of Samples*	Number of Whole Genomes
Betapolyomavirus	<i>Polyomaviridae</i>	1	-
Bocaparvovirus	<i>Parvoviridae</i>	2	1
Coronavirus	<i>Coronaviridae</i>	6	2
Betacoronavirus		5	2
Alphacoronavirus		1	-
Cytomegalovirus	<i>Orthoherpesviridae</i>	8	-
Enterovirus	<i>Picornaviridae</i>	20	10
Rhinovirus A		5	4
Rhinovirus B		1	1
Rhinovirus C		5	5
Other enteroviruses		9	-
Influenza virus	<i>Orthomyxoviridae</i>	3	-
Alphainfluenzavirus		2	
Betainfluenzavirus		1	
Mastadenovirus	<i>Adenoviridae</i>	12	-
Metapneumovirus	<i>Pneumoviridae</i>	8	3
Orthopneumovirus	<i>Pneumoviridae</i>	24	12
Parainfluenza virus	<i>Paramyxoviridae</i>	2	1
Human Parainfluenzavirus 4		1	1
Human Parainfluenzavirus 1		1	-
Rotavirus	<i>Reoviridae</i>	1	-
Simplexvirus	<i>Orthoherpesviridae</i>	1	-
* Does not sum to n = 66 samples with evidence of viral infection because 16 individuals were co-infected			

Eleven positive controls were included in this analysis, comprising two patients with influenza A(H3N2) virus detection by RT-qPCR, eight patients diagnosed with RSV, and one with RSV-HRV coinfection. Concordance between mNGS and RT-qPCR methods was high (9/11 samples, 81.8%). Among the two positive controls initially diagnosed with A(H3N2) virus, mNGS identified evidence of alphainfluenzavirus infection in only one. Among the 8 positive controls initially diagnosed with RSV, mNGS identified evidence of RSV infection in all cases. In the RSV-HRV coinfecting positive control, mNGS identified evidence of RSV but not HRV infection.

Thirty consensus genomes were obtained from 29 of 66 samples with evidence of viral infection; consensus genomes corresponded to 12 (40%) orthopneumoviruses (respiratory syncytial virus, RSV), 11 (36.7%) rhinoviruses (HRV), 3 (10%) metapneumoviruses (HMPV), 2 (6.7%) betacoronaviruses, and 1 (3.3%) each of bocavirus (HBoV) and parainfluenza virus (HPIV). Within each virus family, consensus genomes represented a broad virus diversity, including both RSV subtypes A and B; HRV species A, B, and C; human betacoronavirus species HKU1 and OC43; and HMPV subtypes 1 and 2 (Fig. 3a-e). All these viruses except HPIV-4 were included in the 14-pathogen surveillance panel, although primer match varied among the sequences (**Figure S2**; virus-specific mismatches described in the following section).

We describe the identified viruses in further detail:

Pneumoviridae. Multiple known human pathogens are members of the family *Pneumoviridae*, such as human metapneumovirus (HMPV) and human orthopneumovirus (RSV), both of which were included in the RT-qPCR surveillance panel. Between 2015 and 2017, 8 human metapneumovirus infections were identified by mNGS in Madagascar, yielding three whole genome HMPV sequences, two HMPV strain A and one HMPV strain B. BLAST comparison identified that HMPV-A sequences were closest to sequences circulating in China and Australia in the early 2010s and the HMPV-B sequence most resembled sequences circulating in Europe and east Asia in the same period. Although human metapneumovirus was included in the surveillance panel, both HMPV-A sequences had two nucleotide mismatches to the primer set (**Figure S2**).

Twenty-four RSV infections were identified in the mNGS data, twelve of which resulted in whole genome consensus sequences, including ten of subtype A and two of subtype B. No complete RSV sequence had perfect compatibility with the primer set used in RT-qPCR screening, with mismatches of at least one nucleotide per sequence (**Figure S2**). Phylogenetic analysis of the RSV-A sequences indicated two distinct clusters of cases, collected in March 2015 (n = 5) and 2019 (n = 4) (Fig. 4b). The 2015 sequences clustered with a previously published sequence from Madagascar collected in early 2016, and, at a more basal node, with sequences collected in Europe (Fig. 4c). The 2019 sequences were related most closely to sequences circulating in Australia and the Americas between 2017 and 2019 (Fig. 4d). One 2015 Madagascar sequence identified in this study did not cluster with any other Madagascar sequences, but rather with sequences recovered from Morocco and Australia in subsequent years (2016, 2019) (Fig. 4e).

Picornaviridae. The family *Picornaviridae* includes enteroviruses, many of which are known to cause human infection with a range of symptoms. Human rhinovirus was the only enterovirus included in the

multiplex PCR and one positive control was initially diagnosed with an HRV-RSV coinfection. However, the HRV infection diagnosed by RT-qPCR was not detected by mNGS. Between 2014 and 2019, eleven rhinovirus infections were identified in the mNGS data, comprising five HRV-A sequences, five HRV-C sequences, and one HRV-B sequence. Primer matching was low for the HRV-B sequence, with 5 nucleotide differences, and ranged more widely for HRV-A (between 0 and 6 mismatches) and HRV-C (3 nucleotide discrepancies) (**Figure S4**). HRV sequences do not appear to be closely related; among HRV-A, the Madagascar sequences were most closely related to strains A51, 21, 54, 94, and 101 while HRV-C sequences resembled types 3, 22, 24, and 27. Additionally, although no consensus genome was obtained, evidence of infection with enterovirus C99 and with coxsackie virus A was identified in one patient each (see **Table S3**).

Coronaviridae. The family *Coronaviridae* comprises many known human pathogens, including seasonal human coronaviruses as well as zoonotic coronaviruses with pandemic potential (e.g. SARS-CoV, MERS-CoV, SARS-CoV-2). Evidence of infection with both betacoronaviruses, including HCoV-OC43 and HCoV-HKU1, and the alphacoronavirus HCoV-NL63 was identified in six patients, yielding two consensus genomes of HCoV-OC43 (see **Table S3**). All human coronaviruses known to date were included in the surveillance panel. PCR primers perfectly aligned with one of the HCoV-OC43 full genomes identified (accession: PX467129) and mismatched by one nucleotide with the other (accession: PX467121) (**Figure S2**).

Parvovirinae. *Parvovirinae* includes known pathogens human bocaviruses 1, 2, 3, and 4. By mNGS, one human bocaparvovirus was identified in Madagascar from 2015. Although human bocavirus was included in the RT-qPCR screening, it was only detected by mNGS and there were no mismatches between the primers and the sequence.

Paramyxoviridae. The viral family *Paramyxoviridae* also encompasses many known human pathogens, including measles virus, mumps, and human parainfluenza viruses. Of the four types of HPIV, only types HPIV-1, HPIV-2, and HPIV-3 were included in the multiplex RT-qPCR screening. Despite its inclusion in the surveillance panel, HPIV-1 was not detected in a PCR-negative sample from 2017, but was identified by mNGS, although a full genome sequence was not obtained. One full-genome human parainfluenza virus 4 was identified from a sample collected in 2016; this type of HPIV was not included in the surveillance panel. By BLAST comparison, the HPIV-4 sequence was most similar to viruses from South America and Europe sampled between 2013 and 2016.

Polyomaviridae. Evidence of infection with human polyomavirus 4 (betapolyomavirus) was identified in one patient whose samples contained multiple co-detected viruses. This virus, first described in 2007 among hospitalized patients with pneumonia in Australia³⁸, has not previously been reported in Madagascar.

Orthoherpesviridae. *Orthoherpesviridae* is a family of DNA viruses capable of latent infection and subsequent reactivation, causing a wide range of diseases. No virus of this family was included in the multiplex RT-qPCR; however, two different viruses were detected by mNGS. One sample showed

evidence of infection with herpes simplexvirus 1 (HSV-1) and eight samples with cytomegalovirus, although full genome coverage was not achieved for any of these viruses. The HSV-1 infection and all cases of cytomegalovirus were detected alongside at least one other virus.

Reoviridae. No *Reoviridae*, a viral family that includes many common causes of enteric and respiratory disease, were included in the surveillance screening. Using mNGS, a rotavirus was detected in one sample that also had evidence of infection with orthopneumovirus and mastadenovirus.

Adenoviridae. Human adenovirus was included in the multiplex RT-qPCR screening but not identified by PCR in any samples assayed. Nonetheless, using mNGS, we identified evidence of infection with mastadenovirus in twelve patients but were unable to obtain a full genome.

DISCUSSION

We identified several viral pathogens potentially causing respiratory febrile illness in Madagascar that were not diagnosed at the time of initial hospital visit. All viruses for which we obtained consensus genomes were known human pathogens and most viruses identified in this study had been previously detected in Madagascar, including RSV^{9,39,40}, and influenza virus^{6,7,40}, both of which are responsible for high burdens of respiratory disease and hospitalization in children under five. Additionally, metapneumovirus^{16,41}, bocaparvovirus^{7,41}, seasonal coronaviruses⁷, parainfluenza virus^{7,8}, adenoviruses⁴², rotavirus⁴², and rhinovirus⁶ and other enteroviruses, including enterovirus C99⁴³ have all been identified in Madagascar during the last several decades. We present, to our knowledge, the first evidence of infection with human polyomavirus 4 in Madagascar. Despite prior documentation of their circulation in the country, neither cytomegalovirus, EV-C99 nor HPIV-4 were included in the 14-pathogen surveillance panel used for original screening. Rotavirus, which more frequently results in gastrointestinal disease manifestations but has been associated with respiratory symptoms⁴⁴, and herpes simplex virus 1, a virus which often establishes a latent infection not typically associated with respiratory symptoms but can be detected in instances of coinfection with another respiratory pathogen⁴⁵, a pattern replicated here, were additionally not included in the 14-pathogen original surveillance screening platform.

Consistent with global patterns of respiratory virus infections and previous estimates of the respiratory virus burden in Madagascar^{7,46}, respiratory virus infections were detected at higher frequency among younger patients, a pattern which held even after excluding RSV infection, which is strongly associated with patient ages < 2 years in Madagascar⁹. We identified no pattern of viral infection by sex, also consistent with previous research in the country^{7,47}. Codetection of respiratory pathogens has also been previously reported at a similar rate in Malagasy patients⁸. Finally, the broad diversity of respiratory viruses identified in this study is in line with previous work from Madagascar that found a number of different pathogens which underpin care seeking for respiratory symptoms^{8,16,48}. Studies in neighboring

Reunion⁴⁹ island and other countries in sub-Saharan Africa likewise report detection of numerous types of viruses^{50,51} and wide intra-virus diversity⁵² associated with respiratory illness in hospitals and clinics.

Despite a survey of microbial diversity in the remaining nasopharyngeal samples that did not yield evidence of viral infection (n = 36, 35.3%), we were unable to identify likely causal pathogens. Because many bacteria are not obligately pathogenic, without a more nuanced background sample of Malagasy populations to characterize the expected presence of nasopharyngeal microbial taxa, we cannot confidently conclude that the presence of a pathogen or dominance in the sequencing reads corresponds to a febrile illness. While previous work has characterized human respiratory tract microbiomes, microbial profiles of nasopharyngeal microbiomes can vary widely⁵³. Further, the respiratory microbiome has been shown to change throughout an individual's life⁵⁴ as well as in response to a number of behavioral or environmental factors^{54,55}. Previous studies have addressed the inherent variation in respiratory microbiomes by establishing thresholds for attributing respiratory illness to bacterial infection (i.e. in comparison to robust background selections of microbiome composition^{17,56}) or associated particular biotypes of healthy and infected respiratory microbiomes⁵⁷. However, this study collected nasopharyngeal samples from only two healthy controls, and we were unable to establish statistically significant changes in bacterial presence in nasopharyngeal samples. By contrast, respiratory viruses identified in this study are largely not considered potential commensal organisms and we conclude their presence in a nasopharyngeal sample corresponds to recent infection.

Using a global dataset of RSV subtype A sequences circulating concurrently with our Madagascar sequences, we identified evidence of both global interconnectivity and local transmission. The phylogenetic clustering of five RSV sequences collected in early 2015 with a previously published sequence from Madagascar suggests ongoing transmission within the country, in contrast with the divergent 2019 sequence cluster, which appears to have arrived to Madagascar via a different international transmission event. While thousands of human orthopneumovirus sequences have been made public from upwards of twenty African nations, complete or nearly complete RSV sequences of any subtype collected before 2020 are publicly reported in NCBI and GISAID databases from only nine countries. The COVID-19 pandemic resulted in a global expansion in sequencing technology⁵⁸, and since 2020, sequencing output of RSV has increased, with additional countries reporting thousands more sequences. However, limited sampling density of global RSV cases during our study interval restricts inference corresponding to the exact dynamics of RSV introductions to Madagascar. Expanded sequencing pipelines and public reporting of RSV will permit analysis of source and directionality of transmission.

Many of the viral pathogens identified by mNGS techniques were included in the multiplex RT-qPCR panel for screening respiratory infections at Malagasy hospitals. It is possible that a mismatch between the specific primers involved in the RT-qPCR and the genomic composition of circulating pathogens at the time of study resulted in the surveillance system missing detection of these infections; such mismatches have previously resulted in false negatives in screenings for RNA viruses including RSV⁵⁹

and influenza⁶⁰. Further, the substantial within-taxa diversity of viral genomes could result in reduced sensitivity of PCR methods. Pairwise identity between primers and sequences varied among the sequences identified in this study, with most virus sequences diverging from the primer sequences. However, our findings largely underscore the importance of multi-pathogenic surveillance for respiratory pathogens and provide support for continued broad surveillance as the majority of viruses identified by mNGS were represented in the 14-virus multiplex RT-qPCR screening panel. Four of the pathogens identified via mNGS that are not included in this panel, cytomegalovirus, EVC-99, HPIV-4 and rotavirus, have also been previously identified, albeit at low frequency, in Madagascar^{8,43}.

Evidence of viral infection was identified in approximately two thirds of patient samples included in this study with many samples still lacking diagnosis. Previous studies using mNGS to identify causes of undiagnosed illness have reported a range of pathogen detection rates, with several also identifying pathogens in close to 60% of samples^{17,19}, although studies frequently report pathogen identification in < 50% of samples^{18,21,61}, with higher detection rates possible⁶². While it is likely that some still-undiagnosed respiratory illness was caused by bacterial or eukaryotic infections, it is also possible that viral infections were unrecognized, even by mNGS. Contrary to previous findings that viral infection reduces nasopharyngeal microbiome diversity^{56,63}, we did not find a statistically significant difference in diversity indices between samples with and without evidence of viral infection, potentially suggesting that our thresholds for determining respiratory infection were too narrow, or that the mNGS did not capture all viral infections within the cohort. If samples were collected after the onset of infection and illness, false-negative results are possible as viruses may no longer be detectable either through PCR or sequencing at a later timepoint in an individual's infection, particularly given that many of the viruses identified in this study are RNA viruses, which degrade rapidly. Previous qualitative work focused on malarial fever in Madagascar suggests that care-seeking is frequently delayed until either the sick person does not respond to home treatment or the illness becomes severe⁶⁴. Such delays could reduce the probability of virus detection from febrile disease. In previous work using mNGS to study SARS-CoV-2, deep sequencing was not always able to identify genetic material in samples that had tested positive via PCR, with detection most successful during early stage disease⁵⁶. Additionally, lack of maintenance of cold chain risks degradation of RNA in samples and resultingly limited detection by molecular methods, a possibility that cannot be ruled out here. It is also possible that those reporting to a hospital with fever and respiratory symptoms could have an infection that is not primarily shed in respiratory excreta. Previous work suggests investigation of blood, serum, and fecal samples may yield more complete insight into currently undiagnosed febrile patients presenting in Madagascar hospitals; blood samples from individuals with fever enrolled between 2011 to 2013 in Madagascar analyzed with qPCR identified bacterial, eukaryotic, and viral pathogens, including *Klebsiella pneumoniae*, *Candida*, and dengue, enterovirus, and cytomegalovirus, but were only able to “diagnose” a pathogen in 25% of samples⁶⁵. Future work leveraging metagenomic sequencing methods on a wider range of samples with associated metadata may begin to bridge this remaining gap in pathogen detection and provide deeper insight into fever of unknown cause in Madagascar.

CONCLUSION

A variety of respiratory viruses contribute to some, but not all, undiagnosed healthcare visits for fever and cough in Madagascar. While the 14-pathogen multiplex PCR test used for pathogen surveillance cannot detect additional circulating viruses and primer compatibility may have reduced the sensitivity of PCR tests to the viruses identified by mNGS, the surveillance platform for respiratory infections in Madagascar appears well-suited to the current burden of disease. However, mNGS analysis of undiagnosed fevers provided a more comprehensive overview of the diversity of viruses resulting in clinic visits for febrile illness. Integrating mNGS with a shorter delay into disease surveillance efforts could provide a public health benefit in diagnosing pathogens not included in existing screening panels as well as a method for identifying novel or zoonotic pathogens emerging in human populations. Molecular epidemiology using mNGS provides a useful strategy to investigate the etiology of undiagnosed respiratory fevers and expanded sequencing capacity linked more closely with sample collection could yield further insight.

Abbreviations

mNGS

metagenomic Next Generation Sequencing

RT-qPCR

reverse transcription quantitative polymerase chain reaction

RSV

respiratory syncytial virus

RSV-A

respiratory syncytial virus subtype A

IPM

Institut Pasteur de Madagascar

LMICs

lower- and middle-income countries

HRV

human rhinovirus

HMPV

human metapneumovirus

HBoV

human bocavirus

HPIV

human parainfluenza virus

HSV

herpes simplex virus

Declarations

Ethics approval and consent to participate:

This project was reviewed and received approval from the Ministry of Health in Madagascar (Authorization N 068-MSANP/CE) and the Institutional Review Board of the University of California Berkeley. The project was conducted in accordance with the Declaration of Helsinki. Participants provided written informed consent and participants under the age of 18 provided informed assent while consent was obtained from parents or legal guardians.

Consent for publication

Not applicable.

Availability of data and materials

The genomic datasets generated and analyzed during the current study are available in the Sequence Read Archive and on GenBank.

Competing interests

The authors report no relevant competing interests regarding the publication of this article.

Funding

Research was funded by the Bill and Melinda Gates Foundation (GCE/ID OPP1211841 to CEB and J-MH), the Adolph C. and Mary Sprague Miller Institute for Basic Research in Science (postdoctoral fellowship to CEB), the Department of Education Graduate Assistance in Areas of National Need (P200A210054 fellowship funding to SL), and the Chan Zuckerberg Biohub.

Author's contributions:

CEB and J-MH conceived of the project and acquired the funding. Initial PCR screening of respiratory samples was conducted by NHR, VR, and THR. mNGS sequencing was conducted by HCR, THR, CEB, and VA. SL conducted the statistical and phylogenetic analyses and wrote the original draft of the manuscript. All authors edited and approved the manuscript.

Acknowledgements:

The authors gratefully acknowledge Joseph L. DeRisi, Maira Phelps, and Cristina M. Tato of the Chan Zuckerberg Biohub (CZB) for logistical support and Angela Detweiler, Norma Neff, and Michelle Tan of the CZB for sequencing support. We acknowledge the Virology Unit of the Institut Pasteur de Madagascar for supporting the project. We additionally acknowledge all data contributors and their submitting laboratories for generating the genetic sequence and metadata and sharing via the GISAID

Initiative, on which this research is partially based. This work was completed in part with resources provided by the University of Chicago's Research Computing Cluster.

References

1. Sarfo JO et al (2023) Acute lower respiratory infections among children under five in Sub-Saharan Africa: a scoping review of prevalence and risk factors. *BMC Pediatr* 23:225
2. Bender RG et al (2024) Global, regional, and national incidence and mortality burden of non-COVID-19 lower respiratory infections and aetiologies, 1990–2021: a systematic analysis from the Global Burden of Disease Study 2021. *Lancet Infect Dis* 24:974–1002
3. Frigati L et al (2025) Respiratory infections in low and middle-income countries. *Paediatr Respir Rev* 54:43–51
4. Okomo U, Idoko OT, Kampmann B (2020) The burden of viral respiratory infections in young children in low-resource settings. *Lancet Global Health* 8:e454–e455
5. Rajatonirina S et al (2014) Early-warning health and process indicators for sentinel surveillance in Madagascar 2007–2011. *Online J Public Health Inf* 6:e197
6. Guillebaud J et al (2018) Study on causes of fever in primary healthcare center uncovers pathogens of public health concern in Madagascar. *PLoS Negl Trop Dis* 12:e0006642
7. Razanajatovo NH et al (2011) Viral Etiology of Influenza-Like Illnesses in Antananarivo, Madagascar, July 2008 to June 2009. *PLoS ONE* 6:e17579
8. Hoffmann J et al (2012) Viral and Atypical Bacterial Etiology of Acute Respiratory Infections in Children under 5 Years Old Living in a Rural Tropical Area of Madagascar. *PLoS ONE* 7:e43666
9. Randriambolamanantsoa TH et al (2024) Identifying climatic drivers of respiratory syncytial virus (RSV) seasonality in Antananarivo, Madagascar, 2011–2021: a sentinel surveillance study. *bmjph* 2
10. Randrianasolo L et al (2010) Sentinel surveillance system for early outbreak detection in Madagascar. *BMC Public Health* 10:31
11. Alonso WJ et al (2015) Influenza seasonality in Madagascar: the mysterious African free-runner. *Influenza Other Respir Viruses* 9:101–109
12. Heraud J-M, Razanajatovo NH, Viboud C (2019) Global circulation of respiratory viruses: from local observations to global predictions. *Lancet Global Health* 7:e982–e983
13. Guillebaud J, Héraud J, Razanajatovo NH, Livinski AA, Alonso WJ (2017) Both hemispheric influenza vaccine recommendations would have missed near half of the circulating viruses in Madagascar. *Influenza Other Respir Viruses* 11:473–478
14. Rakotoarisoa A et al (2017) Evaluation of the influenza sentinel surveillance system in Madagascar, 2009–2014. *Bull World Health Organ* 95:375–381
15. Tejiokem MC et al (2024) African countries from the Pasteur Network reexamine their syndromic sentinel surveillance system associated with household contact within the AFROSCREEN program. *Front Public Health* 11

16. Razanajatovo NH et al (2023) Epidemiological Patterns of Seasonal Respiratory Viruses during the COVID-19 Pandemic in Madagascar, March 2020–May 2022. *Viruses* 15:12
17. Bohl JA et al (2022) Discovering disease-causing pathogens in resource-scarce Southeast Asia using a global metagenomic pathogen monitoring system. *PNAS* 11, e2115285119
18. Mwakibete L et al (2023) Metagenomic next-generation sequencing to characterize potential etiologies of non-malarial fever in a cohort living in a high malaria burden area of Uganda. *PLOS Global Public Health* 3:e0001675
19. Ramesh A et al (2019) Metagenomic next-generation sequencing of samples from pediatric febrile illness in Tororo, Uganda. *PLoS ONE* 14:e0218318
20. Yozwiak NL et al (2012) Virus Identification in Unknown Tropical Febrile Illness Cases Using Deep Sequencing. *PLoS Negl Trop Dis* 6:e1485
21. Jerome H et al (2019) Metagenomic next-generation sequencing aids the diagnosis of viral infections in febrile returning travellers. *J Infect* 79:383–388
22. Reyes A et al (2021) Viral metagenomic sequencing in a cohort of international travellers returning with febrile illness. *J Clin Virol* 143:104940
23. Albariño CG et al (2014) Novel Paramyxovirus Associated with Severe Acute Febrile Disease, South Sudan and Uganda, 2012. *Emerg Infect Dis* 20:211–216
24. Wu F et al (2020) A new coronavirus associated with human respiratory disease in China. *Nature* 579:265–269
25. Zhang X-A et al (2022) A Zoonotic Henipavirus in Febrile Patients in China. *N Engl J Med* 387:470–472
26. Yek C et al (2022) Metagenomic pathogen sequencing in resource-scarce settings: Lessons learned and the road ahead. *Front Epidemiol* 2
27. Helmy M, Awad M, Mosa KA (2016) Limited resources of genome sequencing in developing countries: Challenges and solutions. *Applied Translational Genomics* 9:15–19
28. Kalantar K et al (2020) IDseq—An open source cloud-based pipeline and analysis service for metagenomic pathogen detection and monitoring. *GigaScience* 9
29. Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ (1990) Basic local alignment search tool. *J Mol Biol* 215:403–410
30. Katoh K, Rozewicki J, Yamada K (2019) MAFFT online service: multiple sequence alignment, interactive sequence choice and visualization. *Brief Bioinform* 20:1160–1166
31. Larsson A (2014) AliView: a fast and lightweight alignment viewer and editor for large datasets. *Bioinformatics* 30:3276–3278
32. Darriba D et al (2020) ModelTest-NG: A New and Scalable Tool for the Selection of DNA and Protein Evolutionary Models. *Mol Biol Evol* 37:291–294
33. Kozlov AM, Darriba D, Flouri T, Morel B, Stamatakis A (2019) RAXML-NG: a fast, scalable and user-friendly tool for maximum likelihood phylogenetic inference. *Bioinformatics* 35:4453–4455

34. Elbe S, Buckland-Merrett G (2017) Data, disease and diplomacy: GISAID's innovative contribution to global health. *Global Challenges* 1:33–46
35. Hadfield J et al (2018) Nextstrain: real-time tracking of pathogen evolution. *Bioinformatics* 34:4121–4123
36. Nguyen L-T, Schmidt HA, von Haeseler A, Minh BQ (2015) IQ-TREE: A Fast and Effective Stochastic Algorithm for Estimating Maximum-Likelihood Phylogenies. *Mol Biol Evol* 32:268–274
37. Yu G, Smith DK, Zhu H, Guan Y, Lan TT-Y (2016) ggtree: an r package for visualization and annotation of phylogenetic trees with their covariates and other associated data. *Methods Ecol Evol* 8:28–36
38. Gaynor AM et al (2007) Identification of a Novel Polyomavirus from Patients with Acute Respiratory Tract Infections. *PLoS Pathog* 3:e64
39. Razanajatovo NH et al (2018) Epidemiology of severe acute respiratory infections from hospital-based surveillance in Madagascar, November 2010 to July 2013. *PLoS ONE* 13:e0205124
40. Razanajatovo NH et al (2022) Assessment of surveillance predictors for suspected respiratory syncytial virus, influenza and *Streptococcus pneumoniae* infections in children aged < 5 years in Madagascar. *IJID Reg* 2:82–89
41. Rajatonirina S et al (2013) Outcome Risk Factors during Respiratory Infections in a Paediatric Ward in Antananarivo, Madagascar 2010–2012. *PLoS ONE* 8:e72839
42. Randremanana R et al (2012) Case-Control Study of the Etiology of Infant Diarrheal Disease in 14 Districts in Madagascar. *PLoS ONE* 7:e44533
43. Razafindratsimandresy R et al (2022) Enterovirus detection in different regions of Madagascar reveals a higher abundance of enteroviruses of species C in areas where several outbreaks of vaccine-derived polioviruses occurred. *BMC Infect Dis* 22:821
44. Dian Z et al (2021) Rotavirus-related systemic diseases: clinical manifestation, evidence and pathogenesis. *Crit Rev Microbiol* 47:580–595
45. Pata R, Datar P The Diagnosis and Management of Herpes Simplex Pneumonia in the Critical Care Setting: A Comprehensive Review. *Cureus* 15, e43224
46. Randriambolamanantsoa TH et al (2024) Climatic drivers of seasonal dynamics for Respiratory Syncytial Virus (RSV) in Antananarivo, Madagascar, 2011–2021. 2024.02.03.24302203 Preprint at <https://doi.org/10.1101/2024.02.03.24302203>
47. Kyu HH et al (2022) Age–sex differences in the global burden of lower respiratory infections and risk factors, 1990–2019: results from the Global Burden of Disease Study 2019. *Lancet Infect Dis* 22:1626–1647
48. Rakoto-Andrianarivelo M et al (2005) High Frequency of Human Enterovirus Species C Circulation in Madagascar. *J Clin Microbiol* 43:242–249
49. M'nemosyme N et al (2025) Shifts in Respiratory Virus Epidemiology on Reunion Island From 2017 to 2023: Impact of COVID-19 Pandemic and Non-Pharmaceutical Interventions. *Influenza Other Respir Viruses* 19:e70075

50. Loevinsohn G et al (2021) Respiratory pathogen diversity and co-infections in rural Zambia. *Int J Infect Dis* 102:291–298
51. Ndiaye D et al (2024) Prevalence of Respiratory Pathogens in Nasopharyngeal Swabs of Febrile Patients with or without Respiratory Symptoms in the Niakhar Area of Rural Senegal. *Pathogens* 13:655
52. L’Huillier AG et al (2015) Molecular Epidemiology of Human Rhinoviruses and Enteroviruses Highlights Their Diversity in Sub-Saharan Africa. *Viruses* 7:6412–6423
53. Cremers AJ et al (2014) The adult nasopharyngeal microbiome as a determinant of pneumococcal acquisition. *Microbiome* 2:44
54. Kumpitsch C, Koskinen K, Schöpf V, Moissl-Eichinger C (2019) The microbiome of the upper respiratory tract in health and disease. *BMC Biol* 17:87
55. Ge Y, Tang G, Fu Y, Deng P, Yao R (2025) The impact of environmental factors on respiratory tract microbiome and respiratory system diseases. *Eur J Med Res* 30:236
56. Mostafa HH et al (2020) Metagenomic Next-Generation Sequencing of Nasopharyngeal Specimens Collected from Confirmed and Suspect COVID-19 Patients. *mBio* 11. 10.1128/mbio.01969-20
57. Kelly MS et al (2017) The Nasopharyngeal Microbiota of Children With Respiratory Infections in Botswana. *Pediatr Infect Dis J* 36:e211
58. Tegally H et al (2022) The evolving SARS-CoV-2 epidemic in Africa: Insights from rapidly expanding genomic surveillance. *Science* 378:eabq5358
59. Kamau E et al (2017) Recent sequence variation in probe binding site affected detection of respiratory syncytial virus group B by real-time RT-PCR. *J Clin Virol* 88:21–25
60. Klungthong C et al (2010) The impact of primer and probe-template mismatches on the sensitivity of pandemic influenza A/H1N1/2009 virus detection by real-time RT-PCR. *J Clin Virol* 48:91–95
61. Saha S et al (2019) Unbiased Metagenomic Sequencing for Pediatric Meningitis in Bangladesh Reveals Neuroinvasive Chikungunya Virus Outbreak and Other Unrealized Pathogens. *mBio* 10. 10.1128/mbio.02877-19
62. Baillie VL, Madhi SA, Ah Yong V, Olwage CP (2023) Metagenomic sequencing of post-mortem tissue samples for the identification of pathogens associated with neonatal deaths. *Nat Commun* 14:5373
63. Edouard S et al (2018) The nasopharyngeal microbiota in patients with viral respiratory tract infections is enriched in bacterial pathogens. *Eur J Clin Microbiol Infect Dis* 37:1725–1733
64. Favero R et al (2022) Experiences and perceptions of care-seeking for febrile illness among caregivers, pregnant women, and health providers in eight districts of Madagascar. *Malar J* 21:212
65. Marks F et al (2021) Pathogens That Cause Acute Febrile Illness Among Children and Adolescents in Burkina Faso, Madagascar, and Sudan. *Clin Infect Dis* 73:1338–1345

Figures

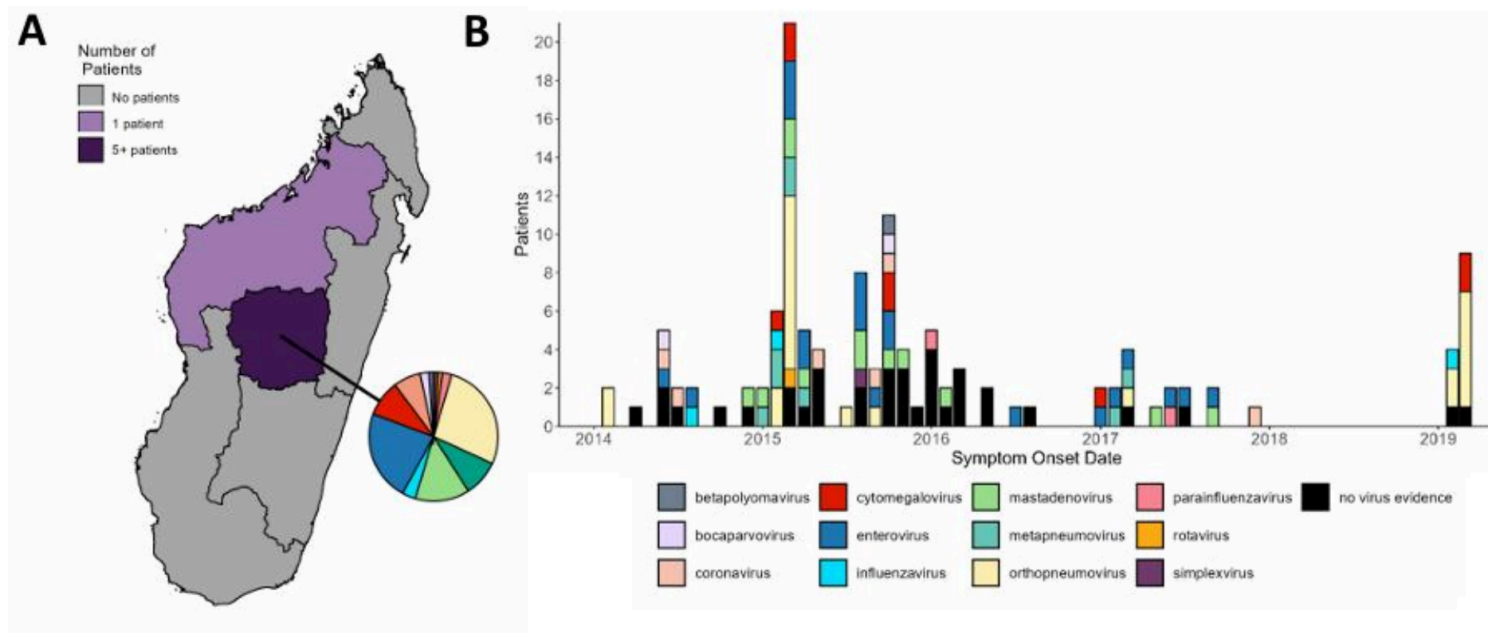


Figure 1

A) Province-level map of Madagascar displaying locations where patient samples were collected. Background color corresponds to the number of individuals sampled in each location. Pie charts indicate the type of viruses identified in the patient population of the three geographic regions in which patient samples yielded evidence of viral infection. Provinces where no samples had evidence of viral infection do not have pie charts. B) Time series of all nasopharyngeal sample collections between 2014 and 2019, with samples that yielded evidence of viral infection indicated in color.

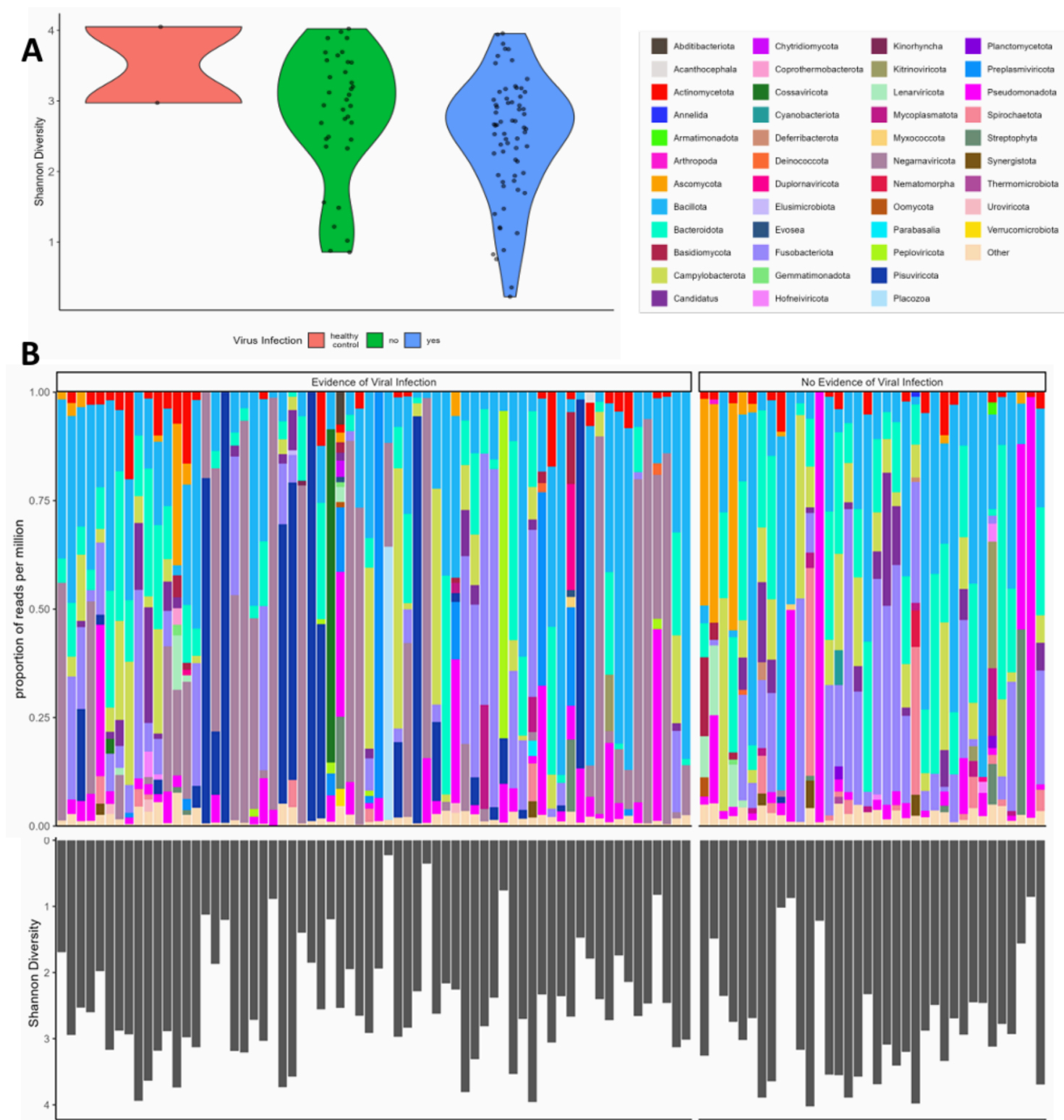


Figure 2

A) Shannon diversity distributions for healthy controls included in the analysis, and samples in which evidence of viral infection was and was not found. B) The relative proportion of reads per million contributed by each microbial class for each sample, stratified by whether or not evidence of viral infection was detected. Classes of microbes that contributed less than 1% of the relative proportion of

total reads per million were reclassified as “Other.” The Shannon diversity index of each sample is displayed beneath each column.

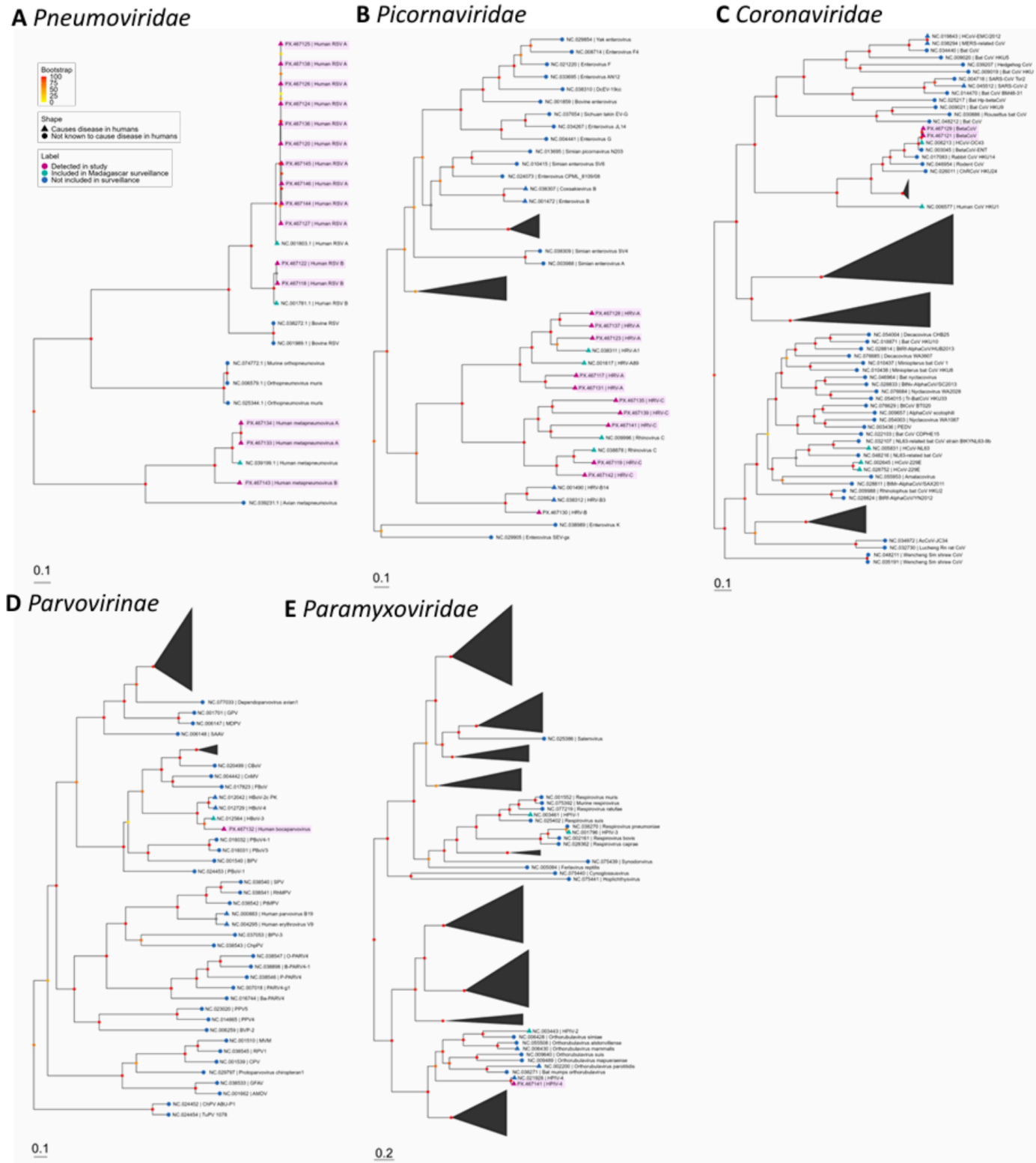


Figure 3

Phylogenetic characterization of the consensus genomes identified via mNGS including full genome nucleotide trees for A) *Pneumoviridae*, B) *Picornaviridae*, C) *Coronaviridae*, D) *Parvovirinae*, and E) *Paramyxoviridae*. Contextual sequences were obtained from NCBI’s GenBank database, comprising

reference sequences of the family of viruses identified in the Madagascar study. Maximum likelihood reconstruction of phylogenetic trees was conducted using GTR+I+G4 nucleotide substitution models. Node color indicates whether viruses were detected by mNGS, included in the original multiplex-PCR screening protocol, or not included in the original surveillance. Text highlights also indicate viruses detected by mNGS. Node shape indicates whether the virus is known to cause disease in humans.

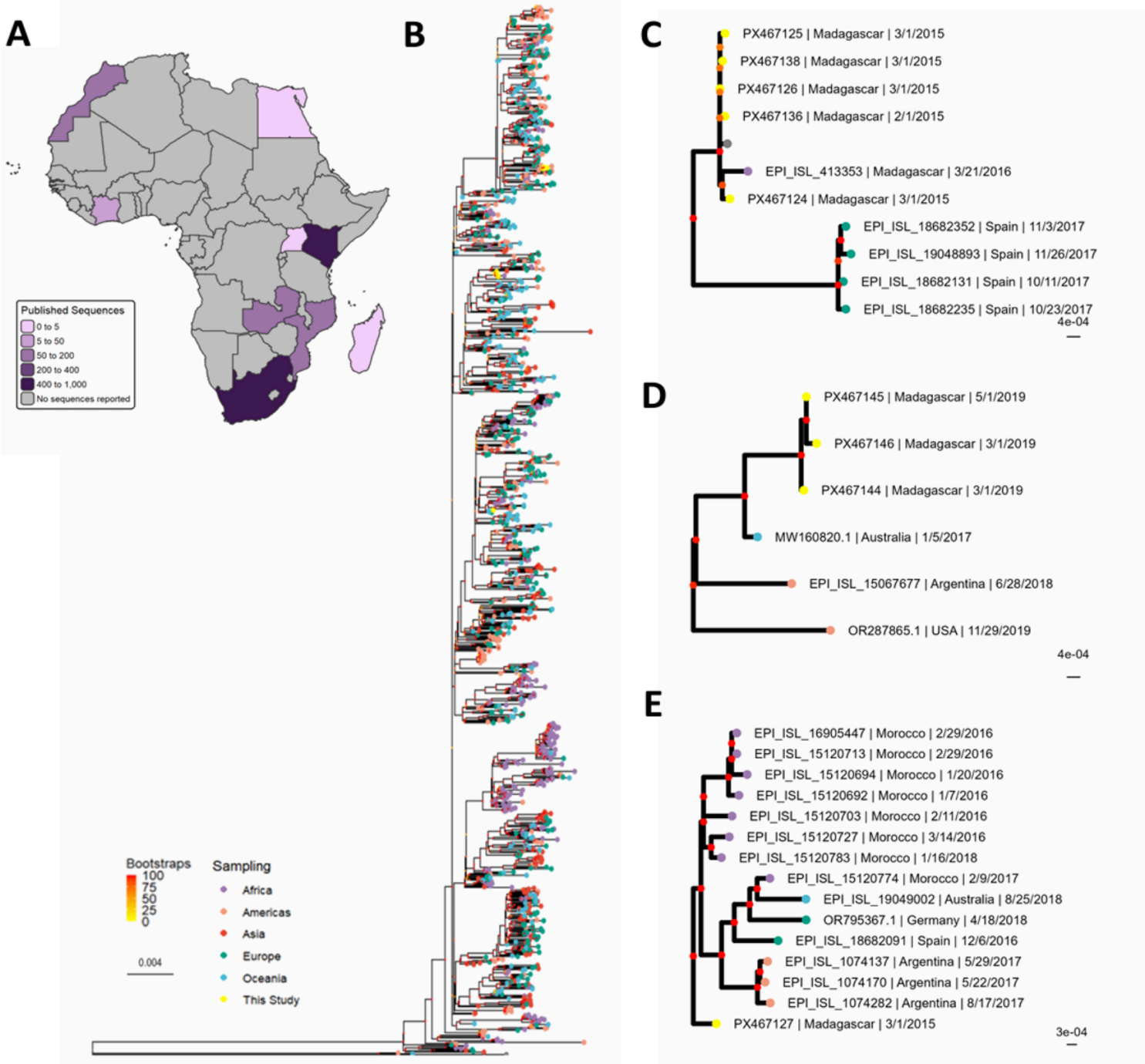


Figure 4

A) Map of all whole genome African RSV sequences available on GenBank/GISAID with collection date between 2013 and 2019. B) Maximum likelihood phylogeny (GTR+I+G4) of whole genome nucleotide RSV subtype A sequences (n=1294) color coded by continent of origin. RSV-A sequences were obtained

from public databases and subsampled. The reference sequence for RSV subtype B was used as an outgroup. C) Subset of part B phylogeny showing a cluster of Madagascar sequences from this study collected in 2015. D) Subset of the part B phylogeny including cluster of Madagascar sequences from this study collected in 2019. E) Subset of part B phylogeny including Madagascar sequence from this study collected in 2015.