Fatal Sabiá virus infection in a yellow fever suspected case, São Paulo, 2019

28 January 2020

Summary

- We have detected Sabiá virus (SABV) genome fragments from two patients suspected for yellow fever virus infection with haemorrhagic fever: Patient 19, a previously undetected case, and Patient 17, a recently reported case. SABV was identified in the patients using an untargeted metagenomic approach;
- Patient 19 was a male, 63 years old, resided in a rural area in the Assis municipality, Sao Paulo state, and reported no travel outside Assis. He was admitted on December 10, 2019, at Hospital das Clinicas, São Paulo, with drowsiness, lethargy, gingivorrhagia, fatigue, nausea, hypoxia, fever and myalgia. Patient 19 died on December 12, 2019;
- Genetic analysis of recovered genome sequences show that SABV strains are genetically diverse (≥10% nucleotide distance) at the nucleoprotein gene compared to each other, and compared to the reference virus sequence from 1990;
- The broad and previously unreported genetic diversity in SABV species revealed by this report poses challenges for existing molecular diagnostic tools and indicates the need to re-evaluate retrospective and prospective cases of haemorrhagic fever with negative yellow fever;
- In-house PCR and nested PCR detection protocols were designed to test SABV strains and all contacts from the most recent patient presenting with SABV; all contacts tested negative.

Clinical and epidemiological context

We have identified viral fragments of the Sabiá virus (SABV), Arenaviridae family, in 63 years old male (Patient 19), admitted on December 10, 2019, to Hospital das Clínicas, São Paulo, Brazil (HC-SP). The patient lived in a rural area in the municipality of Assis and reported no travel over the two weeks before onset of symptoms.

A sample of Patient 19 was taken on December 10, 2019, at the time of admission at HC-SP for the patient's participation in the SOFFA study (ethical approval obtained at the participating site and national research ethics committee, CAAE 82673018. 6.1001.0068). As per the eligibility criteria of the study, Patient 19 had suspected yellow fever infection with symptoms lasting 9 days and had a negative plasma yellow fever virus (YFV) real-time polymerase chain reaction (RT-PCR) reaction. Symptoms included drowsiness, lethargy, gingivorrhagia, fatigue, nausea, hypoxia, fever and

myalgia. Patient 19 died on December 12, 2019, 2 days after arriving to HC-SP and septicemia was considered the cause of death in the death certificate.

Before case from Patient 19 reported here, only three naturally occurring cases of arenavirus infection caused by SABV were detected in Brazil. Probable location of infection, date of onset of symptoms and date of death from naturally occurring SABV cases known to date can be found in **Table 1**.

Table 1. Epidemiological characteristics of detected SABV cases in São Paulo. PMI = Probable municipality of infection. SP = São Paulo. Cases from Dec 2019 were most likely infected in Assis (420 km from Cotia) and Sorocaba/Eldorado (65/210 km from Cotia). Distance between Assis and Sorocaba or Eldorado is 356/430 km. All patients had a fatal clinical outcome. No genomic report exists of strain SPH185338. *Under Investigation.

Strain	Sex, age	PMI in SP state	Occupation	Onset symptoms	Admission HC-SP	Death
SPH114202	Female, 25	Cotia	Agricultural engineer	1 Jan 1990	12 Jan 1990	16 Jan 1990
SPH185338	Male, 32	Espírito Santo do Pinhal	Coffee-grain machine operator	17 May 1999	21 May 1999	28 May 1999
Patient 17	Male, 53	Sorocaba/Eldorado	*	30 Dec 2019	06 Jan 2020	11 Jan 2020
Patient 19 (this report)	Male, 63	Assis	Rural farmer	1 Dec 2019	10 Dec 2019	12 Dec 2020

Laboratory investigation using metagenomic sequencing

Due to the recent case of arenavirus reported in Sao Paulo (Patient 17), the Institute of Tropical Medicine - Faculty of Medicine of the University of São Paulo (IMT-FMUSP) conducted a detailed investigation on yellow fever negative samples initially included in the SOFFA study. Of the 68 patients included in the study, a total of 8 patients were PCR negative for YFV and were analysed using an unbiased metagenomic approach. Patients' samples were collected in EDTA tubes. Plasma was separated in two aliquots and centrifuged at 10,000 g for 5 minutes. RNA was extracted from 200 uL of the sample supernatant using Qiagen RNA mini kits in a P3 Laboratory at IMT-FMUSP.

Sequencing was conducted using the MinION platform from Oxford Nanopore Technologies (ONT) at the IMT-FMUSP as described in CADDE Protocols with small modifications, specifically the metagenomic amplicon fragments were prepared for sequencing using the genomic sequencing kit LSK109 without barcoding. Each sample was subjected to two full technical replicates (RNA extraction, cDNA synthesis, PCR, library generation) before running on separate, fresh ONT

flowcells (one per replicate) to help rule out the possibility of between-sample or barcode contamination.

Two patients yielded reads related to arenavirus: Patients 17 and 19. SABV was detected in all four flowcells. The genome of SABV virus consists of two negative-stranded RNA molecules, L and S, of about 7.5 kb and 3.5 kb in length, respectively. Thus, reads were mapped against the SABV reference genome (L and S segments, Accession numbers NC_006317.1 and NC_006313.1) with minimap2 with parameters '-a' (SAM output) and '-x map-ont' (nanopore mapping mode), '-w1' and '-k11' (increased sensitivity). Sequencing and alignment statistics for each sample are summarized in **Table 2.**

Table 2. Sequencing and alignment statistics for the two SABV Patients analysed in this report.

Strain	Total no. reads	Aligned reads (unique)	Aligned bases
Patient 17	14,423,445	17,320 (0.12%)	1,830,030
Patient 19	6,424,328	132 (0.002%)	42,622

Genome statistics were obtained from samtools and the Tablet viewer. To recover consensus sequences from the S and L genomic fragments, we called variants detected with Medaka for regions of the genome covered with at least 5 reads. Coverage of the two sequences generated here is summarized in **Table 3** below.

Table 3. Sequencing coverage statistics for the two SABV patients analysed in this report.

Strain	Genome segment	Sequence coverage >1X (%)	Sequence coverage ≥ 5X (%)	Mean coverage depth (X)
Patient 17	S	100	100	407
	L	91.6	80.1	117
Patient 19	S	74.24	57.7	7.48
	L	34	14.2	2.57

Genetic diversity of SABV circulating in Sao Paulo

Table 4 summarizes nucleotide and amino acid identity for the L segment, S segment and NP. The sequences recovered from Patients 17 and 19, although still of draft quality due to low read counts in certain regions along the genome fragments, show convincing evidence that the two strains recovered in our study are genetically distinct. We find that at the nucleoprotein amino acid sequence level (n=605 amino acids), Patient 17 shares 93.05% of identity compared to reference strain SPH114202, while Patient 19 shares 96.83% amino acid identity to the reference strain. Patients 17 and 19 share 98.19% amino acid identity at the nucleoprotein.

We find that nucleotide identity between the three SABV strains varies between 88.67-89.06% within the nucleoprotein (n=1815 base pairs). The within-species diversity of SABV is similar to what has been observed in other arenaviruses, for example the Guanarito arenavirus in Venezuela and the Lassa virus in West Africa.

Table 4. Nucleotide and amino acid identity among SABV strains from naturally occurring infections. * We are currently conducting more thorough analyses to investigate low amino acid identity of Patient 17 L segment to reference strain.

Strain	Genome segment	Identity (%) to given strain (nt/aa)
Patient 17	L	Patient 19: 84.53% / 90.27% SPH114202: 73.59% / 68.2%*
	S	Patient 19: 73.06% / 76.97% SPH114202: 89.00% / 95.16%
	nucleoprotein	Patient 19: 89.06% / 98.19% SPH114202: 88.68% / 93.05%
Patient 19	L	SPH114202: 87.75% / 94.39%
	S	SPH114202: 74.44% / 77.74%
	nucleoprotein	SPH114202: 88.16% / 93.05%

Overall, given that the 2019 virus strains share high amino acid identity with the SABV reference strain at the nucleoprotein (this one of the criteria for within-species classification for the *Arenavirus* genus) (**Table 4**), our analysis indicates that strains recovered from Patients 17 and 19 both belong to the SABV species. Importantly, these findings shed new light into an earlier reporting from January 22, 2020, where it was claimed that the arenavirus recovered from Patient 17 did not belong to a defined virus species.

Phylogenetic analysis of L and S segments of the Arenavirus genus

Arenavirus can be divided into two groups: the New World (also named Tacaribe serocomplex) and the Old World (also named LCM/Lassa complex). From the 25 known New World arenaviruses species endemic in the Americas, 4 virus species can be found in Brazil: Cupixi virus discovered in 1965, Amapari virus, discovered in 1965, Flexal virus, discovered in 1977, and Sabiá virus, discovered in 1993. To investigate in more detail the diversity of SABV in relation to the other arenavirus, we next conducted separate phylogenetic analyses of L segment and the S segment of the SABV strains and other Arenavirus species from New World and Old World groups. Phylogenies were estimated using nt and amino acid (aa) alignments using a general time-reversible nucleotide substitution model and a blosum62 amino acid exchange matrix for nt and aa-bases phylogenies, respectively, with PhyML available through seaview's graphical user interface. An approximate likelihood ratio test was used to evaluate branch support.

Figure 1 shows ML phylogenetic trees estimated for the nucleotide alignments of the L segment and the S segments of 34 Arenavirus species. The New World group can be divided into three main clades, named clades A to C. All of the viruses causing hemorrhagic fever (HF) are members of clade B, which includes Junin (L segment: Accession numbers: AY358022 and S segment: AY358023), Machupo (L segment: AY358021 and S segment: AY129248.1), Guanarito (L segment: AY358024.2, S segment: AY129247), and Sabia virus reference strain SPH114202 (L segment: AY358026; S segment: U41071). These cause Argentinian HF, Bolivian HF, Venezuelan HF, and four fatal HF cases in Brazil (including the one reported here).

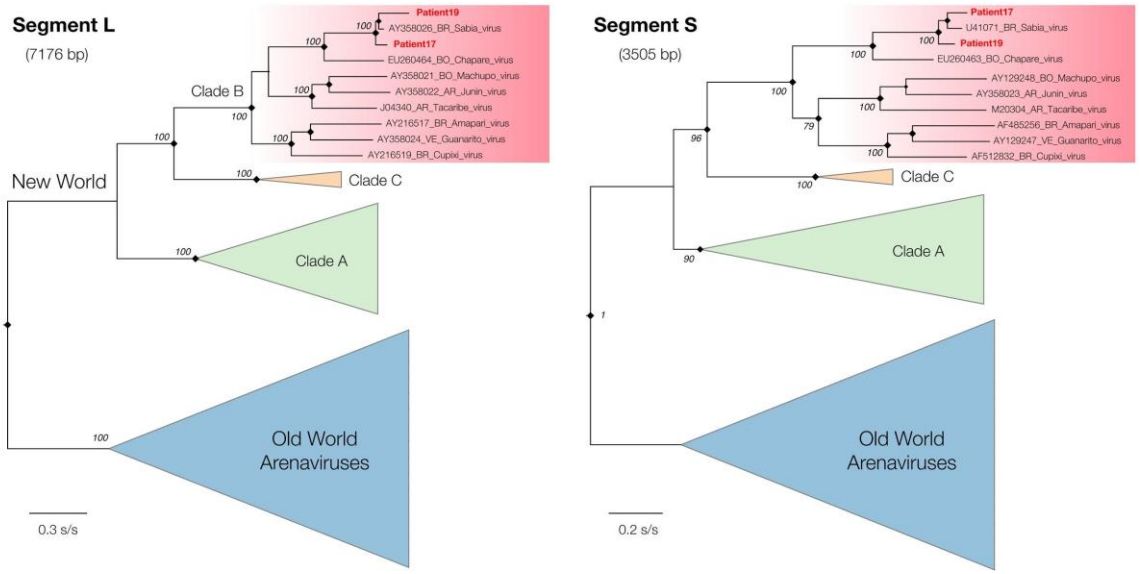

Figure 1. **Maximum likelihood phylogenetic trees of Arenavirus species.** SABV strains are highlighted by a box coloured with a red gradient. Tips of phylogeny indicate accession number of virus strains included in the analysis. Virus species and sampling locations for Clade B viruses are shown. Sequences from Patient 17 and Patient 19 are highlighted in red and fall with maximum statistical support in a clade with the reference SPH114202 SABV strain (see also **Table 1**). The phylogeny on the left is based on a nucleotide (nt) alignment of the L segment and the phylogeny on the right is based on an

alignment of the S segment. Scale bar is in units of nucleotide substitutions per site. bp = base pairs. BR = Brazil, AR = Argentina, BO = Bolivia, VE = Venezuela.

As expected, strains from Patients 19, 17 and SPH114202 cluster closely in a well-supported subclade of clade B New World Arenaviruses with maximum statistical support (bootstrap support = 100%). By visually comparing the topologies of segments L and S of New World Clade B arenaviruses, our analysis suggests no evidence of reassortment of SABV with other Arenavirus species. However, ongoing analyses are being conducted to identify possible within-species recombination, one of the key features of arenavirus evolution in nature. Currently the reservoir species for SABV remains unknown. It would be important to conduct investigative work to identify the reservoir species of SABV and its geographical distribution.

Taken together, genomic and epidemiological data suggest that Patient 19 is not part of a potential human to human transmission cluster involving Patient 17 that reported SABV infection from Sorocaba/Eldorado. Instead, these results suggests two different introductions from a natural reservoir, possibly rodents. Additional research should focus on investigating rodent reservoir species of SABV in or around probable location of infection (**Table 1**).

Genetic diversity of SABV challenges commonly used diagnostic tools

The onset of symptoms of 3 out of 4 naturally occurring SABV cases (**Table 1**) coincides with the season of yellow fever transmission in Sao Paulo (December - January). Moreover, Patient 17 and Patient 19 were both suspected of yellow fever infection. It is likely that several other infections may have gone unnoticed. The high within-species nucleotide diversity of Sabia virus reported here may challenge commonly used lab diagnostic tools to detect SABV strains circulating in Brazil. To investigate whether PCR primer nucleotide sequences from existing assays would detect the new sequences from the 2019 cases in Sao Paulo state, we summarized the matches and mismatches between SABV strains (reference NC_006317 and new sequences) and existing PCR primer sequences from Bowen et al. that is typically used in public health laboratories in Brazil (**Figure 2**).

Alignment of the primers from this assay indicates that the new Patient 17 strain is more divergent to the 1696R primer than the SABV reference sequence genome (**Figure 2**). The presence of two additional mismatch sites to this primer raises the possibility that currently used primers may not be able amplify newly identified SABV strains. Unfortunately, we currently lack the genomic sequence of Patient 19 at these sites (see also **Table 3**; please note that we are currently working to generate more a more complete genome sequence for Patient 19).

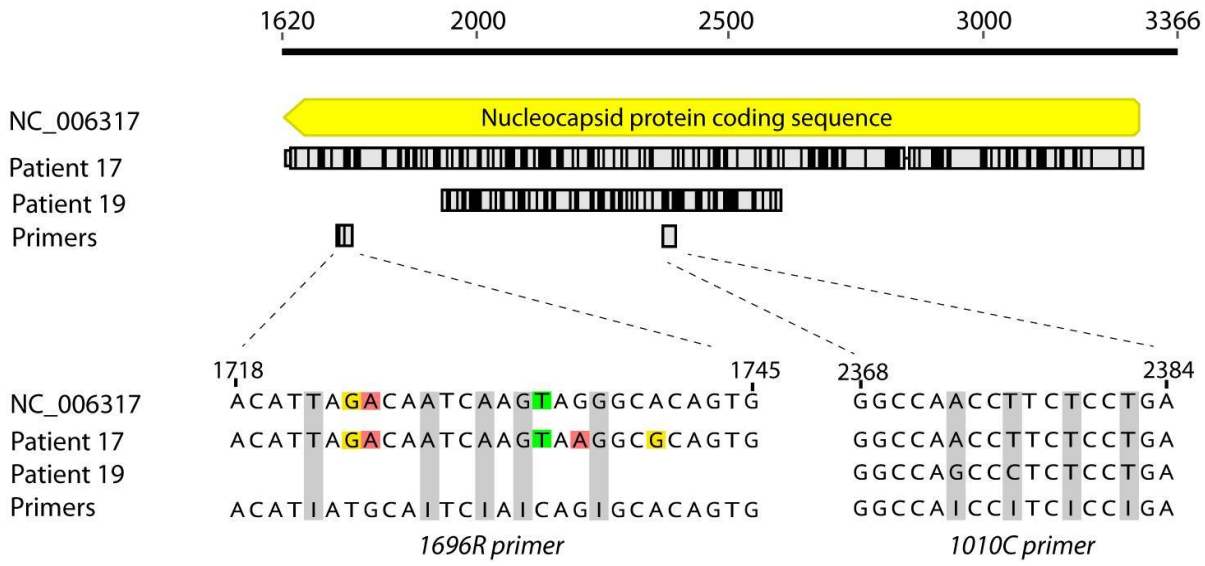

Figure 2. Comparison of matches and mismatches between SABV and existing diagnostic tools. Genomic locations are given relative to the start of the S segment of the SABV reference genome (NC_006317). Upper bars show the available sequence of the new strains relative to the SABV reference strain and location of primers, with variants coloured in black. Lower part shows the primer and SABV sequences at primer binding sites. Colours here indicate mismatches to the primers, and pale grey bars indicate inosine sites.

PCR primers to detect circulating SABV strains

SABV is highly infectious and lethal and may spread through aerosols. Person-to-person transmission can occur with direct contact with the blood or other body fluids of infected individuals. Airborne transmission of arenaviruses has been reported and contact with infected medical material can also be associated with transmission. In an attempt to design an effective PCR approach to investigate virus infection and possible contamination to close contacts of Patient 17 and Patient 19, we next designed and evaluated the performance of six new PCR primers to detect the virus strains recovered here. A list of the primer sequences and additional details is shown in **Table 5**.

All primer sets were ordered on Jan 21, 2020. Upon arrival on Jan 23, 2020, primers were pre-diluted in 1X TE buffer (Sigma) to 100 µM. The extracted RNA from Patient 17, Patient 19 and an additional negative control was converted into cDNA using Random Hexamers (Invitrogen) and ProtoScript II First Strand cDNA Synthesis Kit (New England BioLabs) following instructions. The PCR reactions were prepared according to the conditions described in **Tables 6 and 7**, changing the Tm according to the set of primers used.

Table 5. Description of the primer sets used for Conventional PCR validation. Tm = primer melting temperature.

Primer sets	Name	Sequence 5’ - 3’	N° of bases	Amplicon Length	Tm
1	Pos_29F	GTCACGCTTAAATCTTTGATTGC	23	350 bp	55°C
	Pos_381R	ACAGACACCTCAAGACACCA	20		
2	Pos_1642_2F	AGTGTGAAGAGCAATGCATCAG	22	200 bp	52°C
	Pos_1842R	AATCTGGTCAGATTTTGGTCACC	23		
3	All_3F	TCACGCTTAAATCTTTGATTGCG	23	121 bp	65°C
	All_3R	GCAGCAATYAAGCTCACTGC	20		
4	Pos_2F	CGCACCGGGGATCCTAGGC	19	380 bp	55°C
	Pos_381R	ACAGACACCTCAAGACACCA	20		
5	All_2F	GCAGTGAGCTTRATTGCTGC	20	723 bp	60°C
	All_2R	YACTGCATCAGTTATTGTCC	20		
6	Pos_171F	AACCTGTGGAAGAGTGGCCT	20	200 bp	52°C
	Pos_381R	ACAGACACCTCAAGACACCA	20		

Table 6. PCR performing reagents and conditions for the new SABV PCR detection protocols.

Component	Volume in 25 uL reaction
5x Q5 reaction buffer	5 uL
10 mM dNTPs	0.5 uL
Q5 DNA polymerase	0.25 uL
Nuclease-free water	Up to 25 uL
Primers 10 uL	1.25 uL forward primer 1.25 uL reverse primer
cDNA	2.5 uL

Table 7. PCR cycling conditions for the new SABV PCR detection protocols.

Temperature	Time	No. cycles
98 ^{0}C	1 min	1x
98 ^{0}C	30 s	35 x
Variable	30 s	
72 ^{0}C	1 min	
72 ^{0}C	2 min	1x
4 ^{0}C	Hold	

We next prepared a gel using the E-gel EX Agarose 2% (Invitrogen). To load the PCR products, 5 µl of the products were mixed with 15 µl of NFW and 20 µl of E-gel Sizing DNA ladder (Invitrogen) was used in the first band. The gel was placed into the E-gel equipment and the run was performed until the bands were distinguishable by transillumination. The specific bands of the correct size were observed according to expected amplicon lengths (**Table 5**) to all the reactions, except for the set 4 in which no bands were detected (**Figure 3**). No bands were detected in the negative controls, indicating that there was no contamination.

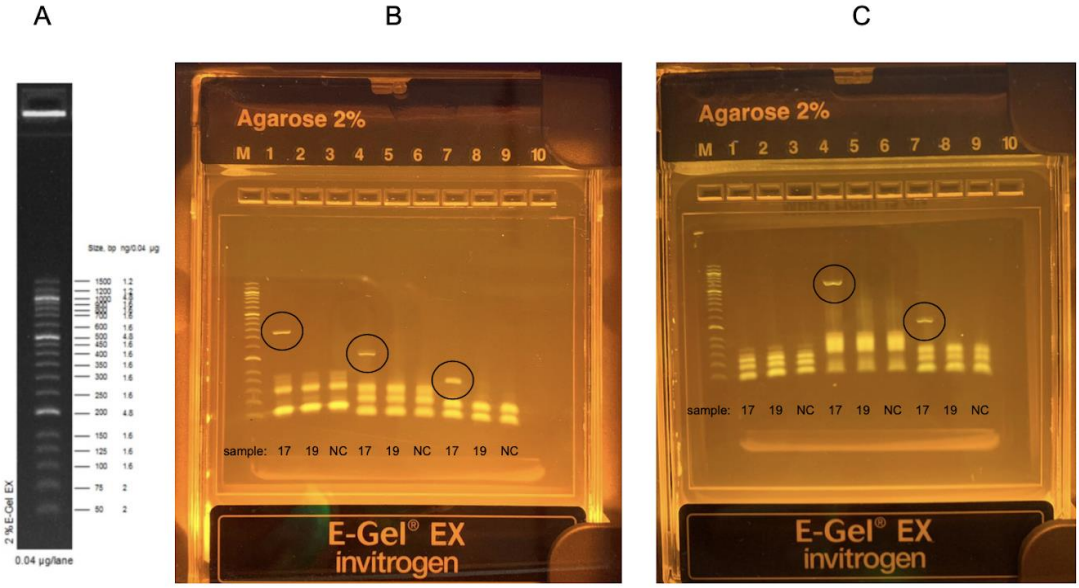

Figure 3. Results of the PCR validation for the different SABV schemes. Panel A: E-gel Sizing DNA leader template; Panel B: the bands highlighted by a circle refer to the sets 1, 2 and 3 respectively; Panel C: the bands highlighted by a circle refers to the sets 5 and 6, respectively. No bands were observed for set 4.

While 5 of the 6 sets of primers tested here were able to capture SABV in Patient 17 (the most recent case), we were not able to detect SABV in Patient 19 (oldest cases from early Dec 2019). [Noteworthy, to date no known contact of Patient 19 developed symptoms.] This is probably due to the low viremia of SABV in Patient 19 as suggested by the low fraction of SABV reads (0.12% in Patient 17 versus 0.002% for Patient 19) (**Table 2**) and the high within-species diversity of arenaviruses.

We then selected primer sets 1 and 3 and tested all patients who had direct contact with Patient 17 or with his blood using primer sets 1 and 3 (**Table 5**, **Figure 3B**) at the Hospital das Clínicas and at IMT-FMUSP. Reassuringly, all close contacts of Patient 17 (and two negative controls) tested negative for SABV while the positive control tested positive (**not shown**).

Nested PCR to detect SABV in samples with low viremia

We next designed two nested PCR primer sets in an attempt to detect the virus in Patient 19 (oldest cases from early Dec 2019) and the virus in the Patient 17 (newest case) in the same reaction. A nested PCR approach is a more sensitive reaction because it employs a second stage of amplification with a pair of internal primers (inner) to those used in the first stage (outer). A list of the primer sequences is shown in **Table 8.**

Table 8. Description of the primer sets used for nested PCR validation. Tm = primer melting temperature.

Primer sets	Name	Sequence 5' - 3'	N° of bases	Amplicon Length	Tm
1	Pos_29F_outer	GTCACGCTTAAATCTTTGATTGC	23	350 bp	55°C
	Pos_381R_outer	ACAGACACCTCAAGACACCA	20		
	Pos_171F_inner	AACCTGTGGAAGAGTGGCCT	20	200 bp	52°C
	Pos_381R_inner	ACAGACACCTCAAGACACCA	20		
2	S_outer_1_f	TCAGTGCAGGGACAGATCCA	20	494 bp	69°C
	S_outer_1_r	TCCCTGAGAAGAGGGCTCAG	20		
	S_inner_1_f	TACAACCCCTGGAGACCTCA	20	112 bp	68°C

	S_inner_1_r	TCAGGAGGTGTGTACCTGGG	20		

For this reaction, the extracted RNA from Patient 19 and an additional negative control was converted into cDNA as described above. Conventional PCR approach and the PCR reactions were prepared according to the conditions described in **Tables 9 and 10**, changing the Tm according to the set of primers used.

Table 9. Nested PCR performing reagents and conditions for the new detection protocols. Note that the reaction is done twice, once for the outer PCR reaction, and then for the inner PCR reaction.

Component	**Volume in 25 uL reaction**
5x Q5 reaction buffer	5 uL
10 mM dNTPs	0.5 uL
Q5 DNA polymerase	0.25 uL
Nuclease-free water	Up to 25 uL
Primers 10 uL	1 uL forward primer and 1 uL reverse primer
cDNA	2.5 uL

Table 10. PCR cycling conditions for the SABV Nested PCR detection protocol. Note that the reaction is done twice, once for the outer PCR reaction, and then for the inner PCR reaction.

Temperature	**Time**	**No. cycles**
98 ^{0}C	1 min	1x
98 ^{0}C	30 s	35 x
Variable	30 s	
72 ^{0}C	1 min	
72 ^{0}C	2 min	1x
4 ^{0}C	Hold	

Reassuringly, our nested PCR allowed the amplification of the sample from Patient 19 even with a low viremia (**Figure 4**). A new gel was prepared as before and the specific bands of the correct size were observed according to the Table 8 (Amplicon Length) to the reactions (**Figure 4**).

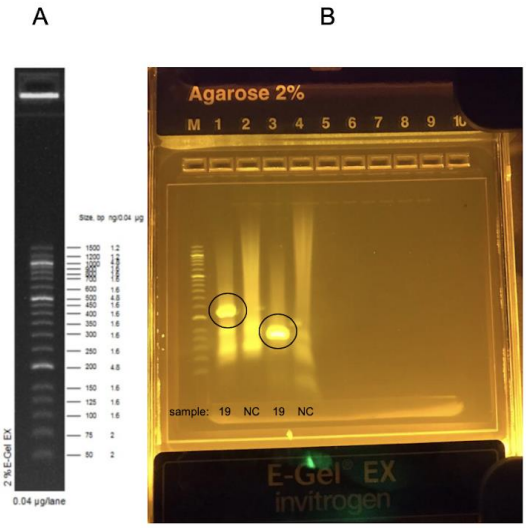

Figure 4. Results of the nested PCR validation for the different Sabiá virus schemes. Panel A: E-gel Sizing DNA leader template; Panel B: the bands highlighted by a circle refer to the sets 1 and 2.

Conclusions

In conclusion, our metagenomic approach was able to rapidly identify the third naturally occurring case of Sabiá virus in a patient with yellow fever-like symptoms. The case was first notified to Centro de Vigilancia Epidemiologica “Prof. Alexandre Vranjac”, Secretary of Health of São Paulo state, on January 21, 2020. Subsequently, we conducted a more extensive characterization of the virus and of the tools commonly used to identify SABV infection in Brazil. We found that the currently used PCR tools may not be able to detect circulating SABV strains due to the virus’ genetic diversity in primer binding sites. Using the genomic data generated by our team, we then designed a PCR scheme to test Patient’s 17 close contacts for SABV. All contacts tested negative (January 27, 2020). We then designed a more sensitive nested PCR protocol that successfully detected SABV in a low virus titer sample from Patient 19. The lower viremia in SABV in Patient 19 compared to

Patient 17 most likely results from the longer period between onset of symptoms and sample collection (9 days for Patient 19 versus 6 days for Patient 17). The framework described here, from metagenomic testing to virus sequencing, development and validation of standard and nested PCR laboratory diagnostic tools for divergent strains and variable viremia, was conducted in a total of 10 days. Serological and molecular testing of patients from Sorocaba/Eldorado and Assis and, more generally, from yellow fever negative cases with hemorrhagic symptoms, should allow a better understanding of SABV' full spectrum of SABV disease, its pathogenicity and transmissibility.

Data availability

The L and S fragments of the strains generated in this study are available for download following the hyperlinks: Patient 17 L segment.fas, Patient 17 S segment.fas, Patient 19 L segment.fas, Patient 19 S segment.fas. Please feel free to download, share, use, and analyze this data. We ask that you communicate with us if you wish to publish results that use these data in a journal. If you have any other questions, please contact us directly.

Contact

Professor Ester Cerdeira Sabino (sabinoec@gmail.com)

Instituto Medicina Tropical

Faculdade Medicina Universidade de São Paulo (FMUSP)

São Paulo, Brazil

Professor Anna Sara Levin (anna@usp.br)

Coordinator of the Group of Hospital Infection Control at Hospital das Clinicas of FMUS

Associate Professor of the Department of Infectious and Parasitic Diseases of FMUSP

Universidade of São Paulo, Brazil

Professor Nuno Rodrigues Faria (nuno.faria@zoo.ox.ac.uk)

Associate Professor, Department of Zoology, University of Oxford

Oxford, United Kingdom

Professor Nick Loman (nick@loman.net)

University of Birmingham

Birmingham, United Kingdom