

RESEARCH

Open Access



Virome analyses of *Hevea brasiliensis* using small RNA deep sequencing and PCR techniques reveal the presence of a potential new virus

Paula L. C. Fonseca¹, Fernanda Badotti², Tatiana F. P. de Oliveira^{1,3}, Antônio Fonseca³, Aline B. M. Vaz^{1,4}, Luiz M. R. Tomé¹, Jônatas S. Abrahão¹, João T. Marques⁵, Giliane S. Trindade¹, Priscila Chaverri^{6,7}, Eric R. G. R. Aguiar^{5,8*} and Aristóteles Góes-Neto^{1*}

Abstract

Background: *Hevea brasiliensis* is an important commercial crop due to the high quality of the latex it produces; however, little is known about viral infections in this plant. The only virus described to infect *H. brasiliensis* until now is a *Carlavirus*, which was described more than 30 years ago. Virus-derived small interfering RNA (vsRNAs) are the product of the plant's antiviral defense triggered by dsRNA viral intermediates generated, during the replication cycle. These vsRNAs are complementary to viral genomes and have been widely used to identify and characterize viruses in plants.

Methods: In the present study, we investigated the virome of leaf and sapwood samples from native *H. brasiliensis* trees collected in two geographic areas in the Brazilian Amazon. Small RNA (sRNA) deep sequencing and bioinformatic tools were used to assemble, identify and characterize viral contigs. Subsequently, PCR amplification techniques were performed to experimentally verify the presence of the viral sequences. Finally, the phylogenetic relationship of the putative new virus with related viral genomes was analyzed.

Results: Our strategy allowed the identification of 32 contigs with high similarity to viral reference genomes, from which 23 exhibited homology to viruses of the *Tymoviridae* family. The reads showed a predominant size distribution at 21 nt derived from both strands, which was consistent with the vsRNAs profile. The presence and genome position of the viral contigs were experimentally confirmed using droplet digital PCR amplifications. A 1913 aa long fragment was obtained and used to infer the phylogenetic relationship of the putative new virus, which indicated that it is taxonomically related to the Grapevine fleck virus, genus *Maculavirus*. The putative new virus was named *Hevea brasiliensis virus* (HBrV) in reference to its host.

Conclusion: The methodological strategy applied here proved to be efficient in detecting and confirming the presence of new viral sequences on a 'very difficult to manage' sample. This is the second time that viral sequences, that could be ascribed as a putative novel virus, associated to the rubber tree has been identified.

Keywords: *Hevea brasiliensis*, Virome, RNA deep sequencing, ddPCR, HBrV

* Correspondence: ericgdp@gmail.com; arigoesneto@pq.cnpq.br

⁵Department of Biochemistry and Immunology, Instituto de Ciências Biológicas, Universidade Federal de Minas Gerais (UFMG), Belo Horizonte, MG 31270-901, Brazil

¹Department of Microbiology, Instituto de Ciências Biológicas, Universidade Federal de Minas Gerais (UFMG), Belo Horizonte, MG 31270-901, Brazil

Full list of author information is available at the end of the article



Background

Hevea brasiliensis (the rubber tree) is described as source of higher quality latex and rubber among all known plant species. The latex produced by this plant has a set of features such as elasticity, abrasion and impact resistance, heat dispersion and malleability, which makes it the ideal feedstock for building products in many different areas, such as engineering, medical and pharmaceutical industries [1, 2]. To date, no synthetic rubber obtained from petroleum has shown similar properties, thus making *H. brasiliensis* latex irreplaceable [3]. Natural rubber is an indispensable commodity used to manufacture more than 50,000 products, accounting for over US\$16.5 billion annually in global exportation [4].

Currently, more than 90% of natural rubber production occurs in Asia, mainly in Malaysia, Thailand and Indonesia, even though the rubber tree originated in the Brazilian Amazon [5]. Commercial rubber plantations in Brazil and other countries in Latin America have failed due to the South American Leaf Blight (SALB) disease [6], caused by the fungus *Pseudocercospora ulei* [7]. Apart from SALB, *H. brasiliensis* is also known to be susceptible to many other diseases caused by fungi [8], such as anthracnose caused by *Colletotrichum gloeosporioides* [9] or the powdery mildew caused by *Oidium hevea* [10]. Virus have also been described to infect rubber trees. However, the knowledge about viral pathogens is restricted to one species belonging to the genus *Carlavirus*, which is associated with the Leaf Disease of Viral Origin [11].

In response to viral infections, eukaryotic organisms have developed different strategies to protect their genomes. RNA interference (RNAi) is an important regulatory mechanism to induce the silencing of self and non-self RNA through sequence-specific homologous interactions [12]. In this process, small RNAs (sRNAs) are generated through the processing of long double-stranded RNA (dsRNA) precursors by an RNase III-like enzyme (DICER). The produced short sequences are loaded onto the argonaute enzyme (AGO) to generate the RNA-induced silencing complex (RISC), which will target complementary regions of mRNA, leading to translation inhibition and mRNA destabilization [12–14]. In plants, this RNA interference pathway is named Post-Transcriptional Gene Silencing (PTGS) [15], and small interfering RNAs (siRNAs) are the most important sRNAs in the plant's response against viral infections [16].

Therefore, the sequencing of sRNAs is a useful tool for detecting phytoviruses, also in asymptomatic and infected plants [17, 18]. The deep sequencing of virus small interfering RNAs (vsiRNAs) can be used to reconstitute the sequences from which they originated and can also be indicative of an infection without requiring

direct detection of the virus. This strategy has been successfully applied to identify and characterize viruses in plants, fungi and animals [19–21].

Here, we investigated the virome of asymptomatic native *H. brasiliensis* trees using sRNA deep sequencing, bioinformatics tools and PCR techniques. This strategy allowed to identify viral sequences, which after their assembly, were found to be phylogenetically related to Grapevine fleck virus (GFkV), genus *Maculavirus*. Our findings expand the knowledge about the *H. brasiliensis* virome and suggest that other studies are necessary to explore the viral biodiversity in this plant.

Methods

Study areas and sample collection

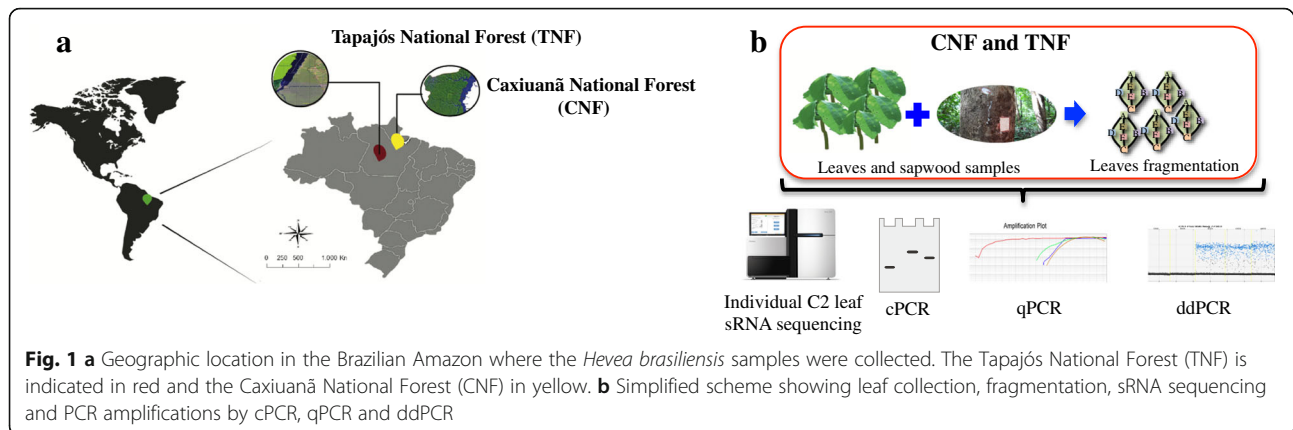
This study was carried out in two different areas in the state of Pará, Brazil: Caxiuanã National Forest - CNF (01°37'S – 02°15'S; 51°19'W – 51°58'W) and Tapajós National Forest - TNF (02°45'S–04°15'S; 54°45'W–55°30'W) (Fig. 1a). Native *H. brasiliensis* species prevail in these areas, which are characterized by a short dry season and excessive precipitation, with a mean annual temperature of 25.9 °C and 82% humidity. The permissions for the study were obtained from the *Sistema de Autorização e Informação em Biodiversidade* (SISBIO), authentication code 31162617.

Leaf and sapwood samples were randomly collected from five adult individuals in CNF (C1–C5) and from 10 individuals in TNF (T1 – T10), from which five were adults and five were plantlets. All collected samples were originated from asymptomatic plants with the leaves showing homogenous green coloration without any kind of discoloration, wilting, or necrotic lesions. Five fragments from each leaf (Fig. 1b) and five pieces (5 mm each, 100 mg in total) of sapwood fragments (3–6 cm periderm) from each individual were placed in sterile cryotubes containing 1.0 mL of RNAlater® Stabilization Solution (Thermo Fisher Scientific, Carlsbad, CA, USA), transported to the laboratory in an ice box, and stored at – 80 °C [22].

RNA extraction and small RNA deep sequencing

Leaf fragments from CNF were ground with liquid nitrogen and 500 mg aliquots were used for RNA extraction using TRIzol Reagent® (Thermo Fisher Scientific, Carlsbad, CA, USA) following the manufacturer's recommendations. The quality and quantity of RNA were evaluated using spectrophotometry (NanoDrop ND-1000, NanoDrop Technologies, Wilmington, DE, USA) and automated electrophoresis systems (2100 Bioanalyzer, Agilent RNA 6000 Nano Kit, Agilent Technologies, Waldbronn, DE).

RNA samples were stored with 30 µL of RNA protection reagent (OMEGA bio-tek, Norcross, GA, USA).



The samples were prepared for sequencing with the NEXTflex Small RNA – Seq Kit V3 and sRNAs were selected by size (15–35 nt) in denaturing SDS-PAGE electrophoresis (Bio Scientific Corp, Austin, TX, USA) and sequenced using Illumina HiSeq (Illumina, San Diego, CA, USA).

Bioinformatic analyses

Pre-processing of sRNA libraries and virome analyses were performed as described by [23]. Briefly, raw sequences were submitted to quality filters and adaptor removal. Sequences with low Phred quality (< 20), ambiguous nucleotides and/or a length shorter than 15 nt were eliminated. The remaining sequences were mapped and compared to all bacterial, fungal and *H. brasiliensis* reference sequences available on the National Center for Biotechnology Information - NCBI (<https://www.ncbi.nlm.nih.gov/>) using Bowtie, allowing one mismatch [24]. The *H. brasiliensis* genome was downloaded from the Genome Online Database – GOLD (<https://gold.jgi.doe.gov/>).

Sequences that did not present similarities with bacteria, fungi or the host were used for contig assembly and subsequent analyses. Assembled contigs greater than 50 nt were characterized based on sequence similarity and pattern-based strategies. The identification of conserved domains was performed using HMMER [25]. The sRNA size profile was calculated as the frequency of each sRNA read size on the reference genome, considering each polarity separately. The density of sRNAs was calculated as the number of times sRNA reads covered each nucleotide on the reference sequence genome. The Z-score was used to normalize the sRNA size profile and to plot heat maps for each contig or reference sequence using the R with ggplots package [26]. Pearson correlation (confidence interval > 95%) of the Z score values was used to estimate the relationship between sRNA profiles from different contigs and reference sequences. Similarities between sRNA profiles were

computed using hierarchical clustering with UPGMA as the linkage criterion. Groups of sequences with more than one element with at least 0.8 of Pearson correlation between each other were assigned to clusters. The sRNA size profile coverage density was calculated using *in-house* Perl scripts and plotted using the R program with the package ggplot2 [23].

To predict the genome organization of the viral assembled contigs, we developed a strategy based on sequence similarity searches. First, we compared the contigs against the non-redundant GenBank database (NR) using Blast (minimum *e*-value of $1e^{-5}$) to select the closest references. Once the genome viral references were selected (Additional file 1: Figure S1), the contigs positioning in each genome were calculated as the total number of contigs mapping into the reference. A ‘partial score’ was generated for each contig against each reference and stored (Additional file 1: Figure S1), subsequently a ‘cumulative score’ was generated (Additional file 1: Figure S1) considering the sum of all the ‘partial scores’.

Detection of viral RNA by qPCR and ddPCR

To confirm the pipeline used for contig ordering, conventional PCR (cPCR), quantitative PCR (qPCR) and digital droplet PCR (ddPCR) amplification techniques were performed using oligonucleotides designed to amplify four larger fragments (Additional file 2: Table S1).

A total of 200 ng of total RNA were used in Reverse Transcriptase Reaction (RT-PCR) with random primers (500 ng/μL) (Thermo Fisher Scientific, Carlsbad, CA, USA) and M-MLV Reverse Transcriptase (Promega, Madison, WI, USA), following the protocols described by [27]. Conventional PCR was performed using 1.5 μL of each designed primer (10 pmol/μL), 4.0 μL (200 ng) of cDNA using KAPA Taq PCR Kit (KAPA BioSystems, Wilmington, MA, USA). The amplification reaction was performed using the annealing temperature at 56 °C. The cDNA samples were also used to amplify the *H.*

brasiliensis actin constitutive gene by qPCR to evaluate the extraction efficiency [28].

The fragment amplification by qPCR was performed using TaqMan Universal Master Mix II (Thermo Fisher Scientific, Carlsbad, CA, USA) in a solution containing 2.0 μ L (100 ng) of cDNA in a reaction volume of 8.5 μ L containing 5.0 μ L of TaqMan Universal Master Mix II (Thermo Fisher Scientific, Carlsbad, CA, USA) and 1.5 μ L of primers (10 pmol/ μ L) with probes (5 pmol/ μ L), and 2.0 μ L of DNase/RNase-free water. Amplification reactions were performed according to the TaqMan recommendations (Thermo Fisher Scientific, Carlsbad, CA, USA) with an annealing temperature of 60 °C. After amplification, the qPCR products were cleaned up with a GenElute PCR Clean-up Kit (Sigma-Aldrich, St. Louis, MO, USA) and sequenced using traditional Sanger technology. The ddPCR was carried out using ddPCR Supermix for Probes (BIO-RAD, Hercules, CA, USA), 2.0 μ L (100 ng) of cDNA in a reaction volume of 18.0 μ L containing 12.5 μ L of 2X ddPCR Supermix for Probes (BIO-RAD, Hercules, CA, USA), 1.5 μ L of primers (10 pmol/ μ L) with probes (5 pmol/ μ L), and 4 μ L of DNase/RNase-free water. The droplet was transferred to a semi-skirted 96-well PCR plate (Eppendorf, Hamburg, DE), which was sealed and subjected to amplification in a Px2 Thermal Cycler (Thermo Electron Corporation, Foster City, CA, USA). Following PCR amplification, the plate was placed in a QX200 droplet reader (BIO-RAD, Hercules, CA, USA). The amplicon quantity was evaluated using QuantaSoft version 1.7 (BIO-RAD, Hercules, CA, USA) by determining the threshold value and the number of positive copies of the target in a 20 μ L reaction.

Phylogenetic analysis

The viral contigs were translated into amino acids, concatenated and aligned with reference sequences of the *Tymoviridae*, *Alphaflexiviridae*, *Gammaflexiviridae* and *Betaflexiviridae* families using MAFFT version 7 using the G-INS-i criteria. Viral sequences of the referred families were obtained from NCBI protein databases under the following accessions: *Tymoviridae* family: Grapevine fleck virus (NP_542612); Citrus sudden death-associated virus (YP_224218); Mayze rayado fino virus (AAK52838); Oat blue dwarf virus (NP_044447); Anagryis vein yellowing virus (YP_002308578); Diascia yellow mottle virus (YP_002048673); Eggplant mosaic virus (NP_040968); Nemesia ring necrosis virus (YP_002308442); Ononis yellow mosaic tymovirus (NP_041257); Plantago mottle virus (YP_002308445); Turnip yellow mosaic virus (AAB2649); *Alphaflexiviridae* family: Shallot virus X (NP_620648); Botrytis virus X (AAL17722); Lolium latent virus (YP_001718499); Indian citrus ringspot

virus (NP_203553); Potato virus X (YP_002332929); *Sclerotinia sclerotiorum* debilitation-associated RNA virus (YP_325662); *Betaflexiviridae* family: Apple stem grooving virus (NP_044335); Cherry virus A (NP_620106); Aconitum latent virus (NP_116487); Citrus leaf blotch virus (NP_624333); Apple stem pitting virus (NP_604464); Grapevine rupestris stem pitting-associated virus (NP_047281); Peach chlorotic mottle virus (YP_001497153); Apple chlorotic leaf spot virus (NP_040551); Peach mosaic virus (YP_002308565); Grapevine virus A (NP_619662) and *Gammaflexiviridae* family: Botrytis virus F (NP_068549).

Phylogenetic analyses were carried out in the Geneious 9 and MEGA 7 software using distance-based (distance matrix) and character-based (maximum likelihood) methods, respectively. Mean distances and the neighbor-joining algorithm were used for distance analyses, and the best-fit model of protein evolution, previously selected in ProtTest 3.2 with Akaike Information Criterion (AIC), was used for maximum likelihood analyses [29, 30]. Clade robustness was assessed using bootstrap proportions (1000 replicates). The trees were edited using FigTree 4.0 (<http://tree.bio.ed.ac.uk/software/figtree/>) and mid-term rooted.

Results

Virome analyses

To investigate the virome of *H. brasiliensis*, one leaf sample obtained from one tree was selected to construct two sRNA libraries based on RNA quality. The two libraries were pooled, totalizing 21,446,061 reads. Reads with low quality or containing ambiguous bases were excluded, which resulted in 18,591,249 reads left. The pre-processing of the data resulted in a decrease of only ~13% of the reads, highlighting the representativeness and quality of our libraries. We enriched reminiscent reads for viral sequences by filtering out reads matching bacterial, fungal and host genomes. The filter step reduced the number of reads to 1,737,014, which were then used for contigs assemblage and characterization of the putative viral genome (Additional file 3: Table S2).

A total of 110 non-redundant contigs were obtained and characterized through sequence similarity searches against NCBI databases. From these, 32 contigs showed significant similarity to the viral reference genomes, from which 23 exhibited homology at the protein level to positive sense single stranded RNA (ssRNA) viruses of *Tymoviridae* family. The remaining contigs that did not show similarity to the viral reference databases presented small RNA size distribution with predominance of 20–22 nt in length derived from both strands (Fig. 2, Additional file 4: Figure S2). We still detected contigs showing similarity with plant, bacteria and fungi, which

likely represent sequences that are not present in the reference genomes deposited in NCBI database.

Characterization of the viral genome

The investigation of the molecular characteristics of the sRNAs mapped into assembled contigs showed that most of the sequences with similarity to Plantae, Bacteria and Fungi have a broad size distribution and originated mostly from positive sense strand RNAs. On the other hand, small RNAs mapped into contigs exhibiting similarity to viral sequences showed a size distribution between 20 and 22 nt derived from both strands, which was consistent with the signature of sRNAs derived from vsiRNA pathway (Fig. 2).

The assembled contigs showed similarities to eleven different viral genomes of the *Tymoviridae* family. In addition, 13 out of 32 contigs presented conserved domains found in viruses of the *Tymoviridae* family (Fig. 3a). Contig sequences and details about the sequence similarity searches are shown in Additional file 5: Table S3 and Additional file 6: Table S4, respectively.

Detection of viral RNA by qPCR and ddPCR

To confirm the presence and validate the predicted position of viral contigs, we designed oligonucleotides to amplify four fragments (Fig. 3b and red arrows in Additional file 4: Figure S2). Quantitative PCR (qPCR) was performed and successfully amplified three out of the four fragments (Fig. 3c), while ddPCR allowed the detection and quantification of the four viral fragments with RNA copies ranging from 25,125 to 104,250.

The presence of viral sequences was evaluated in all the *H. brasiliensis* leaves and sapwood samples collected in TNF and CNF areas using conventional PCR (cPCR), qPCR and ddPCR techniques. However, apart from the leaf sample from CNE, no signal of viral contigs was detected in the samples.

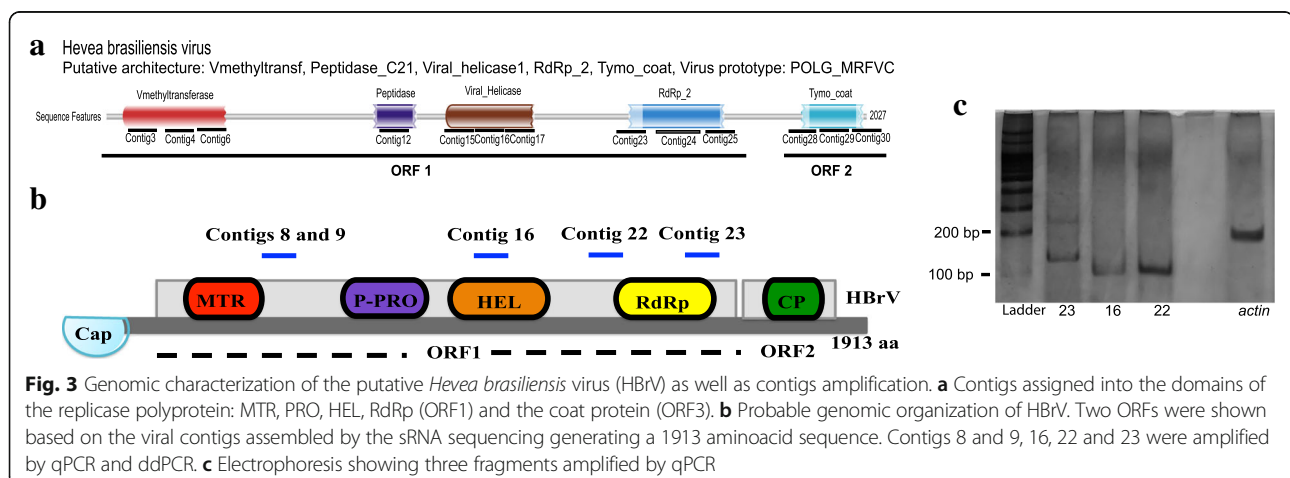
Phylogenetic analyses

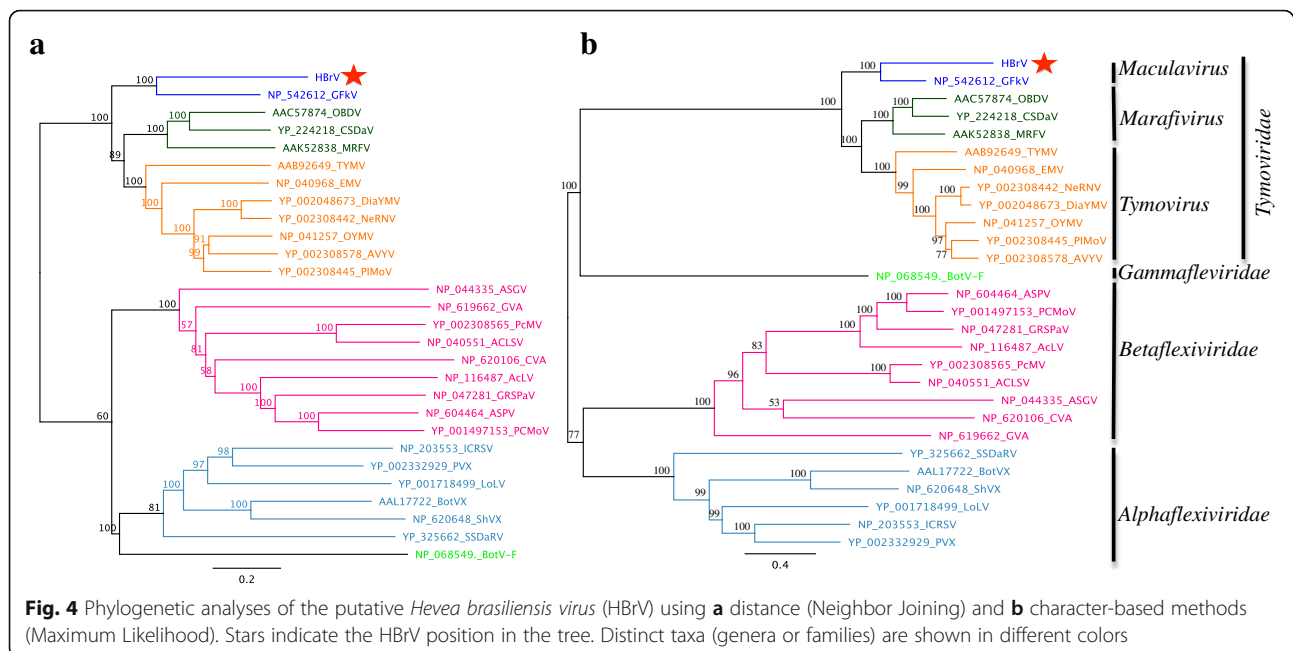
Based on the contig ordering strategy, we generated a 1913-aa long concatenated sequence that was used to infer the phylogenetic relationships of the putative virus. Using the distance-based method with the Neighbor-joining algorithm, it consistently clustered with the Grapevine fleck virus (GFkV) *Maculavirus* with strong statistical support (bootstrap = 100%). In addition, the putative virus was also phylogenetically close to *Marafivirus* and *Tymovirus* (bootstrap = 77.9%) (Fig. 4a). Maximum likelihood character-based phylogeny exhibited a similar topology, with the contig and GFkV *Maculavirus* clustering together (bootstrap = 100%). This analysis also revealed that the contig was related to *Marafivirus* and *Tymovirus* (bootstrap = 82%) (Fig. 4b). Altogether, distance- and character-based phylogenies indicate that the putative new virus belongs to a species closely related to *Maculavirus* and was named *Hevea brasiliensis* virus (HBrV) in reference to its host.

Discussion

Viral dsRNA is used by the plant's immune system to produce virus small interfering RNAs (vsiRNAs) 20–22 nucleotides in length [12, 13]. The vsiRNA class is usually the most abundant in plants infected with RNA virus and have been widely used as indirect evidence of the presence of viral infections since they are produced through recognition of viral dsRNA generated during the virus replication cycle [17, 23]. Therefore, our results indicate that the putative new virus HBrV is replicating in *H. brasiliensis* leaf, since the predominant size of reads entering in the viral contigs assembly was 20–22 nt in both strands.

Metagenomic studies usually generate fragmented sequences, which hamper the characterization of genome structure and phylogenetic analyses [31, 32]. To overcome this limitation, we developed an in silico strategy





based on sequence similarity searches to define the order of viral contigs and their genome structure. Our analyses revealed that 32 contigs have high similarity to viral reference sequences, and 13 of them were assigned to five conserved protein domains found in viruses from the *Tymoviridae* family: methyltransferase (Mtr), RNA helicase (Hel), papain-like cysteine protease (P-Pro), RNA-dependent RNA polymerase (RdRp), and the coat gene.

The ordering and characterization of the viral contigs allowed the design of specific oligonucleotides to experimentally confirm the presence of the viral RNA sequences in our sample via PCR amplifications. Conventional PCR was not suitable to detect their presence, while using qPCR three out of four contigs were amplified. Small amounts of target molecules in a qPCR reaction could decrease amplification efficiency due to the reduced probability of binding oligonucleotides in the template [33]. Digital droplet PCR (ddPCR) allowed the amplification and absolute quantification of the four fragments covering five contigs. The improved performance of ddPCR is probably explained by the massive sample partitioning, high resistance to inhibitory effects of different matrices and the direct quantification that dispenses calibration curves [34, 35].

The use of *in silico* similarity searches associated with experimental confirmation of the RNA viral sequences allowed us to generate a 1913 aa long fragment, which was used to estimate the phylogenetic relationships of the putative virus. Distance and character-based phylogeny methods indicate that HBrV clusters with the

Grapevine fleck virus, genus *Maculavirus* (*Tymoviridae* family).

Genomes of viruses belonging to the *Tymoviridae* family differ in the number of ORFs (from 1 to 4) and in the distribution of genes, but they all encode a large polyprotein essential for viral replication [36]. The genome of *Maculavirus* contains four ORFs, the longest (ORF1) encoding a replication polyprotein constituted by four domains: a methyltransferase (Mtr); a papain-like cysteine protease (P-Pro); an RNA helicase (Hel); and an RNA-dependent RNA polymerase (RdRp), which are conserved among all the ssRNA viruses [36, 37]. In the leaf sample evaluated in this study, a high number of contigs with similarity to the RdRp region was amplified by ddPCR, which may indicate that the putative new virus is active and replicating.

Although hundreds of plant viruses have been described over the past 120 years, only one species (*Carlavirus*, order Tymovirales) was found to infect native *H. brasiliensis* trees in Amazonia [11]. The virus occurrence was reported in leaves of stunted seedlings possessing inter-veinal chlorosis [11]. The genus *Maculavirus*, together with *Marafivirus* and *Tymovirus*, compose the *Tymoviridae* family (order Tymovirales). Members of the *Tymoviridae* family are usually associated with eudicotyledonous plants, which include *H. brasiliensis*, except for some *Marafivirus* that infect monocots (Poaceae). Infections by maculaviruses are known to be phloem-limited [36, 38, 39]. Our findings corroborate these data, since we only detected viral sequences in the leaves and not in the xylem tissue (sapwood).

Conclusions

Small RNA deep sequencing combined with bioinformatic tools and in vitro amplifications proved to be a robust strategy for the accurate identification of viral sequences in *H. brasiliensis* leaf samples. Using this approach, we detected viral sequences associated to the rubber tree that could be ascribed to a putative novel virus, which was named *Hevea brasiliensis* virus (HBrV). As far as we know, this is the second time that viral sequences were identified in this plant; the first, *Carla-virus*, was described more than 30 years ago. Our results highlight the necessity of new studies to identify viruses associated with this economically important, but poorly studied plant.

Additional files

Additional file 1: Figure S1. Overview of the strategy used for the in silico contig ordering. Contigs were anchored in the genome of related viruses based on sequence similarity searches. The position of contigs and the score associated to each position was stored for each reference genome assessed (partial score). After the evaluation of all references, the final contig position was defined by the highest cumulative score obtained through sum of partial scores. (PDF 422 kb)

Additional file 2: Table S1. Oligonucleotides designed for the amplification of viral contigs. (DOCX 59 kb)

Additional file 3: Table S2. Summary metrics of the sequencing and from the contigs assembled in sample C2. (DOCX 49 kb)

Additional file 4: Figure S3. Schematic representation of the Grapevine fleck virus (GFKV) organization. ORF1 (upper box) codes for the replication associated polyprotein (RP) containing the domains of methyltransferase (MTR); papain-like protease (PRO); helicase (HEL); RNA-dependent RNA polymerase (POL); and the coat protein (CP), ORF2 (lower box) encodes the putative movement protein (MP). Bars represent the contigs assembled according to the sequence similarity searches. Red arrows represent the oligonucleotides designed to amplify four fragments covering contigs regions across the genome. (PDF 399 kb)

Additional file 5: Table S2. DNA sequences corresponding to the viral contigs identified in the *Hevea brasiliensis* sample. (DOCX 119 kb)

Additional file 6: Table S4. Identification and characterization of the assembled contigs based on NCBI database searches. Similarity and coverage percentages, accession numbers, contig position and sizes are listed. (DOCX 92 kb)

Abbreviations

AGO: Argonaute; AIC: Akaike Information Criterion; CNF: Caxiuanã National Forest; cPCR: Conventional PCR; ddPCR: Droplet Digital PCR; dsRNA: Double-stranded RNA; GFKV: Grapevine fleck virus; GOLD: Genome Online Database; HBrV: *Hevea brasiliensis* virus; Hel: Helicase; Mtr: Methyltransferase; NCBI: National Center for Biotechnology Information; P-Pro: Papain-like cysteine protease; PTGS: Post-Transcriptional Gene Silencing; qPCR: Quantitative PCR; RdRp: RNA dependent RNA polymerase; RISC: RNA-induced silencing complex; SALB: South American Leaf Blight; siRNA: Small interfering RNA; sRNA: Small RNA; TNF: Tapajós National Forest; vsiRNA: Virus small interfering RNA

Acknowledgements

The authors would like to thank the Graduate Programs of Bioinformatics (<http://www.pgbioinfo.icb.ufmg.br/>) and Microbiology (<http://www.microbiologia.icb.ufmg.br/pos/>) of the Universidade Federal de Minas Gerais (UFMG). We also thank Dener Eduardo Bortolini for helping to review the figures.

Funding

This work was funded by the National Academy of Sciences (NAS) Sub-Grant Number: PGA-2000003475 (2013/2016), Fundação de Amparo à Pesquisa do Estado da Bahia Grant INT008/2014, Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES), and Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq).

Availability of data and materials

Small RNA libraries sequenced in this study were deposited at Short Read Archive (SRA) under accessions: SRR6326570 and SRR6133884.

Authors' contributions

Conceived and designed experiments: PLCF, ERGRA, FB, AF, JSA, JTM, GST, PC, AG-N. Collected and processed samples: PLCF, TFPO, ABMV, LMRT, AG-N. Analysed the data: PLCF, ERGRA, FB, JTM, AG-N. Wrote the manuscript: PLCF, ERGRA, FB, AG-N. All authors read and approved the final manuscript.

Ethics approval and consent to participate

Not applicable.

Consent for publication

Not applicable.

Competing interests

The authors declare that they do not have competing interests.

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Author details

¹Department of Microbiology, Instituto de Ciências Biológicas, Universidade Federal de Minas Gerais (UFMG), Belo Horizonte, MG 31270-901, Brazil. ²Department of Chemistry, Centro Federal de Educação Tecnológica de Minas Gerais (CEFET-MG), Belo Horizonte, MG 30421-169, Brazil. ³LANAGRO/MG –Laboratório Nacional da Agricultura, Ministério da Agricultura (MAPA), Pedro Leopoldo, MG 33600-000, Brazil. ⁴Faculdade de Minas (FAMINAS), Belo Horizonte, MG 31744-007, Brazil. ⁵Department of Biochemistry and Immunology, Instituto de Ciências Biológicas, Universidade Federal de Minas Gerais (UFMG), Belo Horizonte, MG 31270-901, Brazil. ⁶Department of Plant Science and Landscape Architecture, University of Maryland, College Park, MD 20742, USA. ⁷Escuela de Biología, Universidad de Costa Rica, San Pedro, San José 11501-2060, Costa Rica. ⁸Instituto de Ciências da Saúde, Universidade Federal da Bahia (UFBA), Salvador, BA ,40110-100, Brazil.

Received: 16 August 2018 Accepted: 16 November 2018

Published online: 26 November 2018

References

- Cornish K. Similarities and differences in rubber biochemistry among plant species. *Phytochemistry*. 2001;57:1123–34.
- Makita Y, Ng KK, Singham GV, Kawashima M, Hirakawa H, Sato S, Othman AS, Matsui M. Large-scale collection of full-length cDNA and transcriptome analysis in *Hevea brasiliensis*. *DNA Res*. 2017;24(2):159–67.
- Aoki Y, Takahashi S, Takayama D, Ogata Y, Sakurai N, Suzuki H, Asawatreratanakul K, Witisuwannakul D, Witisuwannakul R, Shibata D, Koyama T, Nakayama T. Identification of laticifer-specific genes and their promoters regions from a natural rubber producing plant *Hevea brasiliensis*. *Plant Sci*. 2014;225:1–8.
- Instituto Agrônomo (IAC): Centro de seringueiras e sistemas agroflorestais. <http://www.iac.sp.gov.br/> (2008). Accessed 10 October 2017.
- Van Beilen JB, Poirier Y. Establishment of new crops for the production of natural rubber. *Trends Biotechnol*. 2007;25(11):522–9.
- Lieberei R. South American leaf blight of the rubber tree (*Hevea* spp.): new steps in plant domestication using physiological features and molecular markers. *Ann Bot*. 2007;100:1125–42.
- Hora Júnior BT, Macedo DM, Barreto RW, Evans HC, Mattos CRR, Maffia LA, Mizubuti ESG. Erasing the past: a new identity for the Damoclean pathogen causing south American leaf blight of rubber. *PLoS One*. 2010;9(8):e104750.

8. Kimati H, Amorim L, Rezende JAM, Filho AB, Camargo LEA. Manual de fitopatologia. In: Volume 2: Doenças das plantas cultivadas. São Paulo: Editora Agronômica Ceres; 1997.
9. Liu XJ, Yang YT, Leng HQ. Identification of species and forms of *Colletotrichum gloeosporioides* in rubber growing regions in South China. *Chin J Trop Crops*. 1987;8:93–101.
10. Limlaisang S, Kom-un S, Furtado EL, Liew KW, Salleh B, Sato Y, Takamatsu S. Molecular phylogenetic and morphological analyses of *Oidium heveae*, a powdery mildew of rubber tree. *Mycoscience*. 2005;46:220–6.
11. Gama MICS, Kitajima EW, Ávila AC, Lin MT. Um Carlavirus em seringueira (*Hevea brasiliensis*). *Fitopatol Bras*. 1983;3:621.
12. Romay G, Bragard C. Antiviral defenses in plants through genome editing. *Front Microbiol*. 2017;8:47.
13. Zhang K, Raboanatahiry N, Zhu B, Li M. Progress in genome editing technology and its application in plants. *Front Plant Sci*. 2017;8:177.
14. Aguiar ERGR, Olmo RP, Marques JT. Virus-derived small RNAs: molecular footprints of host-pathogen interactions. *Wires RNA*. 2016;7:824–37.
15. Zamore PD. Viewpoint: Ancient pathways programmed by small RNAs. *Science*. 2002;296:1265–9.
16. Sunkar R, Zhu J. Micro RNA. Short-interfering RNAs in plants. *J Integ Plant Biol*. 2007;49:817–26.
17. Kreuze JF, Perez A, Untiveros M, Quispe D, Fuentes S, Barker I, Simon R. Complete viral genome sequence and discovery of novel viruses by deep sequencing of small RNAs: a generic method for diagnosis, discovery and sequencing of viruses. *Virology*. 2009;388:1–7.
18. Guleria P, Mahajan M, Bhardwaj J, Yadav SK. Plant small RNAs: biogenesis, mode of action and their roles in abiotic stresses. *Genomics Proteomics Bioinformatics*. 2011;9(6):183–99.
19. Kreuze J. siRNA deep sequencing and assembly: piecing together viral infections. In: Gullino ML, Bonants PJM, editors. *Detection and diagnostics of plant pathogens, 21 plant pathology in the 21st century*. Nova lorque: Springer; 2014. p. 21–38.
20. Loconsole G, Saldarelli P, Doddapaneni H, Savino V, Martelli GP, Saponari M. Identification of a single-stranded DNA virus associated with citrus chlorotic dwarf disease, a new member in the family Geminiviridae. *Virology*. 2012; 432:162–72.
21. Vainio EJ, Juvansuu J, Streng J, Rajamaki ML, Hantula J, Valkonen JP. Diagnosis and discovery of fungal viruses using deep sequencing of small RNAs. *J Gen Virol*. 2015;96:714–25.
22. Yockteng R, Almeida AMR, Yee S, AndreT, Colin H, Specht CD. A method for extracting high-quality RNA from diverse plants for next-generation sequencing and gene expression analyses. *Appl Plant Sci*. 2013;1(12): 1300070.
23. Aguiar ERGR, Olmo RP, Paro S, Ferreira FV, de Faria IJDS, Todjro YMH, et al. Sequence-independent characterization of viruses based on the pattern of viral small RNAs produced by the host. *Nucleic Acids Res*. 2015; 43(13):6191–06.
24. Langmead B, Trapnell C, Pop M, Salzberg SL. Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biol*. 2009;10(3):R25.
25. Finn R, Clements J, Arndt W, Miller B, Wheeler T, Schreiber F, Bateman A, Eddy S. HMMER web server: 2015 update. *Nucleic Acids Res*. 2015;43:W30–8.
26. Wickham H, York S. ggplot2: elegant graphics for data analysis. New York: Springer-Verlag; 2009.
27. Al Rwahnih M, Daubert S, Úrbez-Torres J, Cordero F, Rowhani A. Deep sequencing evidence from single grapevine plants reveals a virome dominated by mycoviruses. *Arch Virol*. 2010;156:397–03.
28. Duan C, Rio M, Leclercq J, Bonnot GO, Montoro P. Gene expression pattern in response to wounding, methyl jasmonate and ethylene in the bark of *Hevea brasiliensis*. *Tree Physiol*. 2012;30:1349–59.
29. Akaike H. A new look at the statistical model identification. *IEEE Trans Autom Control*. 1974;19:716–23.
30. Abascal F, Zardoya R, Posada D. ProtTest: selection of best-fit models of protein evolution. *Bioinformatics*. 2005;21:2104–5.
31. Studholme D. Deep sequencing of small RNAs in plants: applied bioinformatics. *Brief Funct Genomics*. 2011;1:71–85.
32. Rose R, Constantinides B, Tapinos A, Robertson D, Prospero M. Challenges in the analysis of viral metagenomes. *Virus Evol*. 2016;2:vev022.
33. Zhao Y, Xia Q, Yin Y, Wang Z. Comparison of droplet digital PCR and quantitative PCR assays for quantitative detection of *Xanthomonas citri* Subsp. *citri*. *PLoS One*. 2016;11:e0159004.
34. Baker M. Digital PCR hits its stride. *Nat Methods*. 2012;9(6):541–4.
35. Huggett JF, Cowen S, Foy CA. Considerations for digital PCR as an accurate molecular diagnostic tool. *Clin Chem*. 2015;1(1):79–88.
36. King AM, Lefkowitz E, Adams MJ, Carstens EB. Virus taxonomy: ninth report of the international committee on taxonomy of viruses. Elsevier: International Union of Microbiological Societies Virology Division; 2012.
37. Simmonds P, Adams M, Benkó M, Breitbart M, Brister J, Carstens E, Davison A, Delwart E, Gorbalenya A, Harrach B. Consensus statement: virus taxonomy in the age of metagenomics. *Nat Rev Microbiol*. 2017;15:161–8.
38. Sabanadzovic S, Saldarelli P, Martelli G, Ghanem-Sabanadzovic N. Complete nucleotide sequence and genome organization of grapevine fleck virus. *J Gen Virol*. 2001;82:2009–15.
39. Katsuma S, Tanaka S, Omuro N, Takabuchi L, Daimon T, Imanishi S, Yamashita S, Iwanaga M, Mita K, Maeda S, Kobayashi M, Shimada T. Novel macula-like virus identified in *Bombyx mori* cultured cells. *J Virol*. 2005;79: 5577–84.

Ready to submit your research? Choose BMC and benefit from:

- fast, convenient online submission
- thorough peer review by experienced researchers in your field
- rapid publication on acceptance
- support for research data, including large and complex data types
- gold Open Access which fosters wider collaboration and increased citations
- maximum visibility for your research: over 100M website views per year

At BMC, research is always in progress.

Learn more [biomedcentral.com/submissions](https://www.biomedcentral.com/submissions)

