BMC Plant Biology

RESEARCH Open Access

Integrative multi-omics approach identified emerging viral threat in *Theobroma cacao* plants

Paula Luize Camargos Fonseca[1†], Mirian S. Santos[2†], Jonatha dos Santos Silva[2†], Juliana N. Armache[3], Carlos Henrique de Carvalho Neto[2], Renata S. Nascimento[2], João Pedro Nunes Santos[2], Gleice R. R. Pires[4], Isaque João da Silva de Faria[3], Yuliana Mora Ocampo[2], Juliana Freitas-Astua[5], Pedro L. Ramos-Gonzaléz[6], Juliano O. Santana[2], Valéria F. Fernandes[2], João P. R. Marques[7], Aline D. Tassi[5], George A. Sodré[8], Luciene Cristina Gastalho Campos Luiz[2], Wilson B. Luiz[2], Carlos P. Pirovani[2], Anibal R. Oliveira[2] and Eric R. G. R. Aguiar[2*]

Abstract

Background *Theobroma cacao* seeds are the source of chocolate and other valuable products. Pathogens can damage cocoa beans, reducing yield and quality. This study examined viruses linked to chlorosis and leaf deformation in cocoa plants in Bahia, Brazil, using an integrative approach.

Results Macroscopic and microscopic analysis of symptomatic *T. cacao* leaves revealed chlorosis and histological alterations consistent with viral infection, including cell and nuclear hypertrophy. Deep sequencing of small RNA (sRNA) from these tissues identified virus-derived sRNA (vsiRNAs) profile, which suggests that viral transcripts activate the antiviral RNA interference pathways. The vsiRNAs were used to assemble two viral sequences: one related to geminivirids and the other resembling a strain of the recently described Cacao swollen shoot Ghana S virus, a putative member of the family *Caulimoviridae* commonly found as an Endogenous Viral Element (EVE) in symptomless plants. The geminivirid was tentatively designated *Citlodavirus theobromae,* while *the* badnavirus isolate was named Brazilian cacao swollen shoot Ghana S virus Ilhéus. Mass spectrometry detected multiple viral peptides from coding genes from both viral genomes, supporting the potential occurrence of viral translation and replication. Notably, symptoms were observed only in association with the geminivirids. Although viral DNA from both viruses was present in symptomatic and asymptomatic leaves of the same plant, geminivirid RNA was exclusively detected in symptomatic tissues, suggesting a correlation between the citlodavirustranscript accumulation and symptom expression.

Conclusion Our findings highlight the emergence of a new viral geminivirid threat that could affect cocoa production in South America.

Keywords *Theobroma cacao*, Virome, *Geminiviridae*, *Badnavirus*, Multi-omics integration

[†]Paula Luize Camargos Fonseca, Mirian S. Santos and Jonatha dos Santos Silva contributed equally to this work.

*Correspondence:
Eric R. G. R. Aguiar
ericgdp@gmail.com

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

Background

The plant species *Theobroma cacao* L., native to the Amazon basin in South America, is an important neotropical crop that belongs to the family *Malvaceae*, subfamily *Sterculioideae* [48]. There are at least 22 different *Theobroma* species, but *T. cacao* is the primary source of cocoa [6]. Different varieties of *T. cacao* are used to produce cocoa and chocolate. The main ones include Criollo, which was utilized by the Aztec and Maya civilizations. Criollo accounts for only 5–10% of chocolate production and is considered the most aromatic of the cocoa beans. The Forastero cultivar, which has many sub-varieties, is responsible for 80% of the chocolate production and is often referred to as Brazilian cacao. The Trinitario cultivar is a hybrid of Criollo and Forastero [19, 64].

Brazil, once a leading cocoa exporter, is currently the fifth-largest producer, contributing about 5% of global production, having produced approximately 300,000 tons in 2023 [16, 30]. The state of Bahia, located in the northeast region of Brazil, is the country's largest cocoa producer, yielding 125,000 tons of beans. The southern part of the state is a traditional cocoa-growing area [15]. Nevertheless, despite the economic importance of cocoa for the state, the occurrence of diseases remains one of the primary obstacles to increasing productivity in plantations across the country [51]. Many fungal pathogens have been identified infecting cocoa, reducing the productivity of the crop by affecting the fruits and seeds [59, 71]. However, few cases of viral pathogens affecting cocoa plants have been reported in Brazil. Recently, the cotton leafroll dwarf virus (species *Polerovirus CLDV*), which is known to infect cotton plants, was also found in cocoa plantations in the state of Bahia [55]. Among viruses that infect cocoa trees, the badnaviruses (genus *Badnavirus*, family *Caulimoviridae*) represent the predominant group, as they can infect both cultivated and wild plant hosts worldwide [14, 56]. Badnaviruses have a circular, double-stranded DNA (dsDNA) genome ranging from 7 to 9.2 kbp and contain three to seven open reading frames (ORFs) [72]. The genome is encapsidated in a non-enveloped bacilliform particle. Badnaviruses are transmitted by one or more species of mealybugs (Pseudococcidae) and, in some cases, by aphids, in a non-circulative, semi-persistent manner [56, 72]. To date, the International Committee on Taxonomy of Viruses (ICTV) recognizes 74 species of badnaviruses that infect cacao, of which at least eight are found in either West Africa or Asia and are believed to share African endemism [72]. Cacao mild mosaic virus (CaMMV, species *Badnavirus tessellotheobromae*) and cacao yellow vein banding virus (CYVBV, species *Badnavirus venatheobromae*) are the only two recognized to date whose members are found in the tropical Americas [49]. Recently, two isolates of CaMMV were identified in Bahia, Brazil (CaMMV-BR321 and CaMMV-BR322), along with one isolate in Puerto Rico (CaMMV-PR) [56]. This virus was later also detected in cocoa trees in Bahia [42]. Furthermore, sequences related to the badnavirus cacao swollen shoot S virus (CSSV) were identified as Endogenous Viral Elements (EVEs) in most of the assessed clones available in the germplasm bank at the CIRAD in Montpellier, France [50, 73]. Although few of these sequences are consistent with almost complete CDS representing viral genomes, no symptoms were observed in any of the evaluated varieties [50]. EVEs are fragments or near-to-complete viral genomes integrated into the host genome. Once integrated, they can be fixed in the population through natural selection or drift. EVEs are found in different types of organisms and because of their widespread presence, they allow us to explore the evolutionary relationship between viruses and their hosts [33, 61].

Many viral pathogens remain undiscovered [78]. These yet-to-be-discovered pathogens can cause numerous diseases, posing a serious threat to food security in developing countries and potentially leading to significant economic losses in important crops [60]. Various techniques for viral identification and detection can be employed. High-throughput sequencing (HTS) offers a non-biased improved identification of viruses infecting a plant, that is based on existing genomic data, and enables the assembly of partial or complete viral genomes [63].

One approach to HTS involves sequencing and analysis of small RNAs (sRNAs) generated by the plant host. These sRNAs are produced by the RNA interference pathways, which play key roles in regulating gene expression and restricting the replication of viruses and mobile genetic elements [35, 45]. Among the different classes of sRNAs, virus-derived small interfering RNAs (vsiRNAs) are of particular interest. vsiRNAs originate from viral RNA during infection and constitute part of the plant`s antiviral defense mechanism. Upon recognition of viral RNA in the cytoplasm, RNA-dependent RNA polymerase 6 (RDR6) synthesizes complementary strands, forming double-stranded RNAs (dsRNAs), which are then cleaved into 21–24 nucleotide fragments by Dicer enzymes [31, 45]. These dsRNAs are then loaded into the RNA-induced silencing complex (RISC), triggering post-transcriptional viral gene silencing [31, 45, 76].

Due to the biogenesis mechanism of these vsiRNA molecules, it has been successfully applied for the identification of viral sequences in many organisms [21, 24, 77]. In conjunction with sRNA deep sequencing, proteomics can also serve as an important strategy for virus identification and characterization, as it allows for the exploration of differential protein expression and the presence of viral-encoded proteins [68]. Here, we employed an integrative multiomics strategy to identify and characterize

a mixed viral infection associated with symptomatic *T. cacao* plants.

Methods

An overview of the different approaches employed in our study for the identification and characterization of Theobroma cacao viruses can be found in Figure S1. Details about each section are described below. The experiments performed in this study were conducted using different plant samples.

Plant material, RNA extraction

Leaves from *Theobroma* cacao plants of the clones PC-209, CCN51, CEPEC2002, and BN34—showing symptoms such as chlorosis and leaf deformation, as well as asymptomatic leaves (used as negative control)—were collected at the Executive Committee of the Cocoa Farming Plan (CEPLAC) (14°45′35.8"S, 39°13′49.09"W), in Ilhéus, Bahia, Brazil. Total RNA was extracted using the TRIzol reagent (Thermo Fisher Scientific, USA). The quality and quantity of the extracted RNA were measured using a NanoDrop spectrophotometer (Thermo Fisher Scientific), with 260/280 ratios ranging from 1.95 to 2.15.

Light microscopy

Fragments of leaf tissue from PC-209 genotype –both chlorotic and asymptomatic samples—were fixed in a solution of 2% glutaraldehyde and 2.5% paraformaldehyde in 0.05 M cacodylate buffer (pH 7.2) for 24 h at room temperature. To enhance the fixative penetration, the tissues were subjected to vacuum (approximately 0.1 mm Hg) for 10–15 min. Subsequently, the samples were dehydrated and infiltrated with historesin, as described previously [43]. After polymerization, cross-sections of 5–8 µm were made using the Leica RM2355 rotary microtome and stained with 0.05% toluidine blue in phosphate-citrate buffer (pH 4.5) [65]. Phenolic compounds were detected using 1% ferric chloride [22]. Tissue sections were mounted on Entellan® and analyzed using a Zeiss Axioskop light microscope. The images were captured using 20 × and 100 × objective lenses, corresponding to total magnifications of 200 × and 1000 ×, respectively.

Transmission electron microscopy

Leaf tissue samples were cut into small sections (approximately 1 mm^2) and immediately fixed in 2.5% glutaraldehyde in 0.1 M phosphate buffer (pH 7.2) at 4 °C for 24 h. After fixation, the samples were rinsed in the same buffer and post-fixed in 1% osmium tetroxide for two hours at room temperature. The tissues were then dehydrated through a graded ethanol series (30% to 100%) and embedded in epoxy resin. Ultrathin Sects. (70–90 nm) were obtained using an ultramicrotome, collected on copper grids, and stained with uranyl acetate followed by lead citrate. The sections were examined using a transmission electron microscope operating at an accelerating voltage of 80–120 kV.

Small RNA deep sequencing and bioinformatics analysis

The sRNA library from leaves of the PC-209 genotype was prepared using approximately 100 ng of the total RNA extract. sRNA fraction was enriched using the Magnetic Bead Purification Module and the library was constructed using the Ion Total RNA-Seq Kit v2 (Thermo Fisher Scientific). The protocol was partially developed using the Ion Chef System and the library was sequenced on the Ion GeneStudio S5 (Thermo Fisher Scientific, USA).

Raw sequenced reads were submitted to a quality filter and adaptor trimming using *Cutadapt* v1.12 [44], discarding sequences with low Phred quality (< 20), ambiguous nucleotides, and/or with length shorter than 15 nt. Pre-processed reads were submitted to a pipeline for viral sequence identification previously published by our group [3]. Briefly, pre-processed sequences were mapped to the reference *T. cacao* genome (GCA_000208745.2) using *Bowtie* v1.1.2 [36], allowing one mismatch. Sequences that did not match the host genome were compared to all bacterial reference sequences available on the National Center for Biotechnology Information (NCBI) (https://www.ncbi.nlm.nih.gov/) using *Bow tie*, allowing one mismatch. Unmapped reads were used for contig assembly using *SPAdes* 4.0.0 [10]. Assembled contigs larger than 200 nt were analyzed by sequence-similarity searches using the Blast tool [5] against NCBI NT and NR databases. Contigs that displayed matching regions with viral sequences were submitted to redundancy removal and contig extension using *CAP3* [28]. To stretch non-redundant viral contigs, and possibly merge some of them, a second round of assembly was done using *SPAdes*, providing the parameter –trusted-contigs and giving as input the previously assembled contigs that matched viral sequences in public databases, according to a previously published methodology [29]. Finally, the resulting contigs from the second round of assembly were submitted to sequences-similarity searches and assessed for the presence of conserved domains using *CDBlast* [79] and *Phmmer* (hmmer.org) tools.

The profile of the sRNAs derived from putative viral contigs was assessed using in-house scripts developed using *Perl* v5.16.3, *BioPerl* library v1.6.924, and R v3.3.1. Plots were made in R using *ggplot2* v2.2.0 package. Three primer pairs were designed according to the presence of conserved domains identified in putative viral contigs using Primer3 [74]. The coordinates and approximate size of PCR amplicons are provided in Figure S2. Each primer pair was then tested in the *OligoAnalyzer tool* (IDT, http

s://www.idtdna.com/pages/tools/oligoanalyzer) to check the melting temperature, and the presence of hairpins, self- and heterodimers. We used the software *SnapGene* (www.snapgene.com) to visualize primer alignments and their flanking regions.

Pairwise comparison of viral sequences

To compare the viral sequences identified with those previously deposited in databases, we performed a nucleotide similarity analysis using BLAST on the NCBI platform, using the non-redundant nucleotide database (nr/nt). For each sequence, we analyzed the top six hits from distinct species, considering only the first hit in cases where multiple entries corresponded to different isolates of the same species (Table S1). Pairwise comparisons were made with the *BLAST* tool with the *blastn* subprogram. Identity and coverage values were used to assess pairwise comparisons between closely related species. Plots were made in R using *ggplot2* v2.2.0 package.

Phylogenetic analysis

Phylogenetic analyses for each of the putative viral sequences were performed to taxonomically assign the identified contigs. Sequence alignments were made in *MAFFT* tool (https://mafft.cbrc.jp/alignment/server/index.html) [34]. The best-fit model for amino acid or nucleotide was selected in *ModelTest-NG* [18] for each group. Phylogenetic inferences were performed using the Maximum likelihood method on the *CIPRES* Portal (https://www.phylo.org/) [46]. Clade robustness was assessed using 1000 replicates of bootstrap and the trees were edited using *iTOL* version 5.7 (https://itol.embl.de/) [38].

Table 1 Oligonucleotides used for viral sequences amplification

Primer name	Sequence (5′–3′)	Tm (°C)	GC content (%)	Amplicon size (nt)
Rep_Gemn—F	GCAGGAGATCGTAGTGGTTCG	60	57	~150
Rep_Gemn—R	ACATCATCAGAACGGCGACC	60	55	
Cap_Gemn—F	AGCCTTCGAGAAGAATGAGT	60	45	~600
Cap_Gemn—R	CTTGAGCAGGATCAGTTACAG	60	47.6	
Ctrl_Badn—F	AGCTAACATGGCTCCAGTGG	60	55	~150
Ctrl_Badn—R	GTTCCTGTGCCCTTGTTGTG	60	55	
Mid_Badn—F	CTAAACGACTGCAAGCGAGC	60	55	~1500
Mid_Badn—R	GCGAAGGCTTACTCGTGAGA	60	55	
Tub_Theo—F	TGCAACCATGAGTGGTGTTCA	60	60.4	~100
Tub_Theo—R	CAGACGAGGGAAGGGAATGA	60	58.8	
Surv_Badna—F	AACCTTCTCGTGTACGCTGG	56,2	55	150
Surv_badna—R	CCCGGCATTCTGAAGTGAT	53,7	52	

Gemn Geminivirid, *Rep* Replication-associated protein, *Cap* Capsid, *Badn* Badnavirus, *Ctrl* Control group (polyprotein gene), *Mid* Region between contigs (primers designed to close this gap)

Viral amplification and Sanger sequencing

Total RNA was isolated using a modified CTAB method [73] followed by purification by SV Total RNA Isolation System kit (Promega, USA) according to the manufacturer's instructions. Total RNA (500 ng) was converted into cDNA using GoScript™ Reverse Transcription System. Applied Biosystems QuantStudio® Real-Time PCR Systems was used for real time RT-PCR. GoTaq® qPCR Master Mix, SYBR® Green I, was used in RT-qPCR.

The presence of each identified virus in *T. cacao* samples was confirmed by PCR amplification using specific primers in other clone varieties. Leaves from the infected cocoa genotypes—*Colección Castro Naranjal*: CCN-51, *Centro de Pesquisa do Cacau*: CEPEC-2002 and BN34—were collected at CEPLAC, Ilhéus, Bahia, Brazil. Negative controls consisted of samples from an asymptomatic cocoa tree. Twenty-five milligrams of each sample were macerated in liquid nitrogen, and DNA was extracted using the CTAB protocol [20]. DNA integrity and concentration were analyzed on a 1.5% agarose gel stained with GelRed™ and quantified using a NanoDrop 2000 spectrophotometer (Thermo Fisher Scientific, USA). Amplifications were performed with GoTaq® Flexi DNA Polymerase kit (Promega) using specific primers (Table 1) following the manufacturer's recommendations. In total, four primer pairs were designed and tested to amplify badnavirus and geminivirid sequences from the plant material.

PCR products were analyzed by electrophoresis on a 1.5% agarose gel in TAE buffer stained with GelRed™. The 1Kb Plus DNA Ladder molecular weight marker was used, and the amplicons were visualized under UV light. The amplicons generated from the positive sample of the CCN-51 genotype were selected and Sanger sequenced (https://actgene.com.br/).

Detection and amplification of a Geminivirid in T. cacao samples

Leaf samples from symptomatic CCN51 variety were collected. The leaves were flash-frozen in liquid nitrogen and subsequently freeze-dried. Approximately 100 mg of the dried plant material was ground and subjected to DNA extraction using the CTAB method, as described previously [20]. Following extraction, samples were tested for the presence of the geminivirid by PCR using the GoTaq® G2 Hot Start Master Mix kit (Promega, USA), according to the manufacturer's instructions, with specific primers targeting the viral capsid (Cap) and replicase (Rep) regions (Table 2).

Once the presence of the virus was confirmed, one sample was selected for full genome amplification (3.7 kb), which was carried out in two steps. In the first step, a 3.049 kb fragment (~3 kb) was amplified using primers A and B (Table 1), incorporating restriction sites for

Table 2 Primer sequences used for detection and amplification of the viral genome

Primer	Sequence 5'–3'	Tm	Restriction enzyme	Amplicon size (bp)
Rep—F	GCAGGAGATCGTAGTGGTTCG	60°C	-	150
Rep—R	ACATCATCAGAACGGCGACC		-	
Cap—F	AGCCTTCGAGAAGAATGAGT	60°C	-	600
Cap—R	CTTGAGCAGGATCAGTTACAG		-	
Primer A—R	*ata***AAGCTT**ATGCTTCTCGCGTTAAATCC	60°C	HindIII	3.049
Primer B—F	*aa***CTGCAG**CTGTGAATCATGCTGGTGCT		PstI	
Primer C—F	*cg***GAATTC**TGAGGGATGGAATATCCTGTT	60°C	EcoRI	1.445
Primer D—R	CACTACCTGGGCGTCTATCTTT		-	

*The bolded region in the sequence indicates the recognition site for the restriction enzymes

HindIII and *PstI* enzymes. In the second step, a 1.445 kb fragment (~ 1.4 kb) was amplified using primers C and D (Table 1), incorporating a restriction site for the *EcoRI* enzyme. PCR assays were performed using the Platinum™ High Fidelity Taq DNA Polymerase kit, following the manufacturer's guidelines.

Amplified products were analyzed by electrophoresis on a 1% agarose gel. Bands corresponding to the expected sizes were excised and purified directly from the gel using the Wizard® SV Gel and PCR Clean-Up System kit (Thermo Fisher Scientific, USA), according to the manufacturer's instructions.

Proteomics assay

Protein extraction and quantification

Total protein extraction was performed following the protocol described previously [52]. Lyophilized leaf biomass (200 mg) from four *T. cacao* scion/rootstock combinations (CCN-51/CCN-51, CCN-51/BN-34, CCN-51/PS-1319, and CCN-51/PH-16) was used. Protein quantification was carried out using the 2D Quant Kit (GE Healthcare Life Sciences), according to the manufacturer's instructions.

Peptide digestion and desalting

Protein digestion followed the in-solution protocol developed by the National Biosciences Laboratory (LNBio, BRA) (https://lnbio.cnpem.br/wp-content/uploads/2012/11/protocolo_digestao_solucao.pdf; accessed on 12/03/2022). For each sample, 100 µg of protein was digested. The resulting tryptic peptides were desalted using C18 resin tips (10 µL; Agilent Bond Elut OMIX®) following the manufacturer's recommendations. Peptides were eluted in 50 µL of a solution containing 50% acetonitrile, 49.9% water, and 0.1% formic acid, and stored at 4 °C until mass spectrometry analysis.

Peptide separation and mass spectrometry (MS)

Peptide analysis was performed using an Agilent 1290 Infinity II HPLC system coupled to an Agilent 6545 LC/QTOF mass spectrometer. Each sample was analyzed in three technical replicates (5 µL injection each). Peptides were separated on a reversed-phase C18 column (AdvanceBio Peptide Mapping, 2.1 × 250 mm, Agilent) at 55 °C using a 20-min gradient with mobile phases A (H_2O + 0.1% formic acid) and B (acetonitrile + 0.1% formic acid). The gradient profile was as follows: 5–35% B (1–10 min), 35–70% B (11–14 min), 70–100% B (16–18 min), and 100% B (16–20 min), followed by 5 min for column re-equilibration.

Peptides were ionized by electrospray and analyzed in AutoMSMS mode, selecting up to 10 precursors per cycle. Selection criteria included a threshold of 1000 counts/spectrum, 100% purity stringency, 30% purity cutoff, and peptide isotopic model, with charge states 2, 3, > 3, and unknown. Collision energy was calculated using the formula: *(slope)*(m/z)/100 + offset, with slope and offset ranging from 3.1 to 5 and −4 to 10, respectively, depending on charge state. Ion source settings were gas temperature 325 °C, gas flow 13 L/min, capillary voltage 4000 V, skimmer voltage 56 V. Nitrogen was used as the collision gas. Instrument control and data acquisition were managed using Agilent MassHunter Acquisition software (see Table S2).

Peptide identification against protein database

Spectral data were analyzed in triplicate using Spectrum Mill software (Rev B.06.00.203 SP1; Broad Institute). Parameters for spectral extraction included: MS noise threshold of 10 counts, fixed carbamidomethylation (C), precursor mass range of 200–6000 Da, ± 60 s retention time tolerance, ± 1.4 m/z tolerance, and automatic precursor charge detection.

Searches were conducted against *T. cacao* protein sequences from UniProt (downloaded 25/03/2021) and a custom viral ORF database. Parameters included: up to 2 missed cleavages; fixed modification: carbamidomethylation (C); variable modifications: oxidized methionine (M), pyroglutamic acid (N-terminal Q), deamidation (N), and phosphorylation (S, T, Y); minimum matched peak intensity: 10%; precursor mass tolerance: ± 10 ppm. Identifications were filtered using a false discovery rate (FDR) threshold of < 1%. Results were exported in MPP APR format using protein–protein comparison mode.

Identification of differentially abundant proteins

Differential abundance analysis was performed using Mass Profiler Professional 15.1 (MPP; Agilent). Protein abundance was calculated as the mean peptide abundance per protein. Comparative analysis between proteins identified in the *T. cacao* database and those in the viral ORF database was conducted. Proteins were filtered based on peptide frequency, retaining only those detected in at least 66% of replicates under at least one condition.

Statistical significance was assessed using an unpaired t-test (treatment vs. control), with asymptotic p-values. Multiple testing correction was applied using the Benjamini–Hochberg method. Proteins were considered differentially abundant if they met both criteria: p-value < 0.05 and fold change > 1.5.

Results

T. cacao leaves show virus infection characteristics

We observed chlorotic spots unevenly distributed along the leaf blade in many adult individuals from at least nine different genotypes from the collection area (Fig. 1 and Figure S3). These symptoms were observed on leaves of different developmental stages and appeared as irregular spots, sometimes next to veins (Fig. 1). These leaves usually had uneven unilateral growth due to wrinkling and chlorotic spots that never progressed to necrosis. The analysis of cross-sections of chlorotic regions of the leaves revealed the occurrence of variations in the cells of the adaxial surface epidermis, palisade, and spongy parenchyma (Fig. 2A). Hypertrophy was observed in some of the epidermis cells on the adaxial surface of chlorotic leaves (Fig. 2B). Changes in the palisade parenchyma are characterized by hyperplasia of the palisade and spongy parenchyma cells, in addition to hypertrophy in the spongy parenchyma cells, leading to a reduction of intercellular spaces (Fig. 2B, C). Nuclei of infected spongy parenchyma cells showed a considerable increase in their size (Fig. 2D and E). Affected tissue also presented phenolic compound accumulation (black arrow) in both spongy and palisade parenchyma (Fig. 2G-H).

Identification of viral sequences using small RNA

Leaf chlorosis and microscopy inspections indicated the potential presence of viral infection in *T. cacao* individuals. Therefore, we performed sRNA sequencing to investigate the presence of viral agents infecting the cocoa plants. Deep sequencing of the library constructed from symptomatic *T. cacao* PC-209 genotype leaves (Fig. 1) generated 6,299,243 reads. After quality processing and host RNA subtraction, a total of 3,545,851 reads were used as input in the viral sequence identification pipeline. Contig assembly yielded 317 contigs, ranging from 50 nt to 3,889 nt, with an average length of 213 nt. Contigs

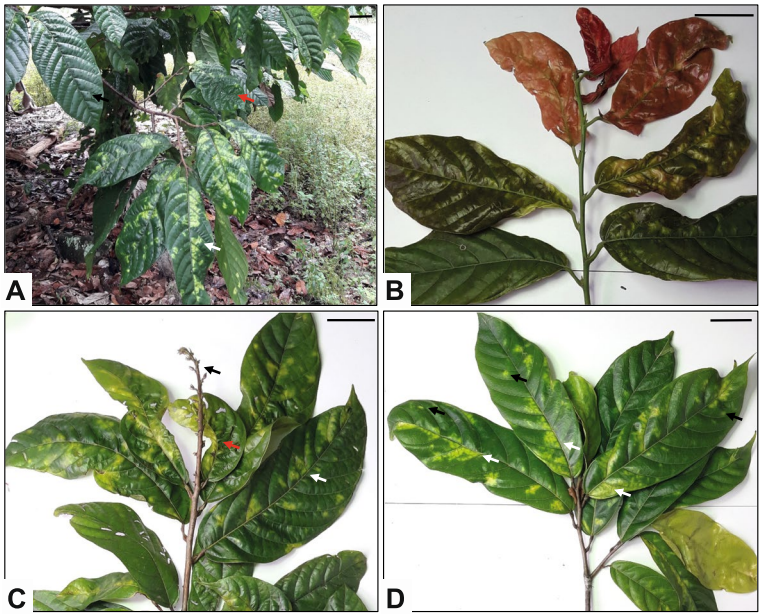

Fig. 1 *T. cacao* leaves exhibit virus-like symptoms of chlorotic spots and wrinkling. **A** Cacao tree of a genotype of the PC genotype, with arrows pointing to a healthy, asymptomatic leaf (black), and affected leaves, showing chlorosis in patches mainly distributed along the leaf midrib (white) and wrinkling (red). **B** New leaves of the genotype PC-209 exhibiting chlorosis and wrinkling. **C** Branch of the genotype TSH 1188, with arrows pointing to developed leaves exhibiting virus-like symptoms, such as chlorosis (white), uneven unilateral growth due to wrinkling (red), and the premature leaves falling, shortening of internodes and proliferation of lateral branches characteristic of the '*engurruñadera*' (black). **D** Branch of unidentified genotype originated from somatic embryogenesis, with arrows pointing to developed leaves with chlorosis as small spots (black) and in patches distributed along the leaf midrib (white), and distant from it (red). Scales: 50 mm

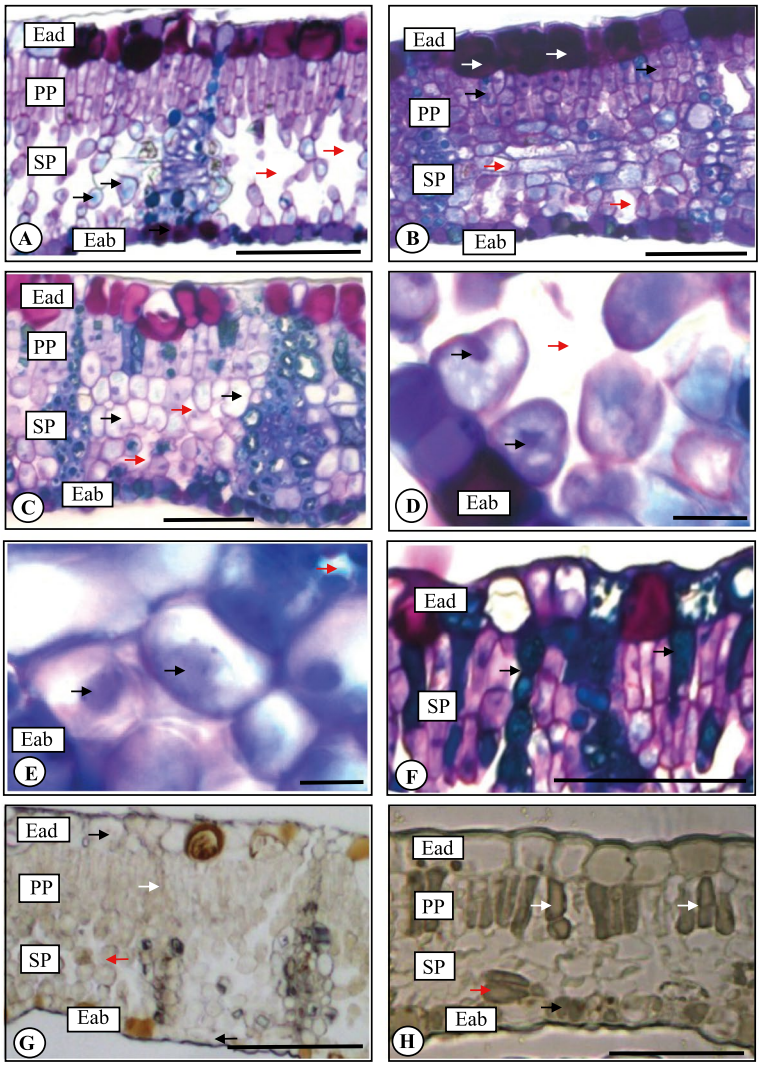

Fig. 2 Light micrographs of toluidine blue-stained (**A**–**D**) and ferric chloride-tested (**E**–**F**) cross sections of cacao leaves in regions healthy (**A**, **D**, **G**) and virus- like symptoms of chlorosis (**B**, **C**, **E**, **F**). A) Healthy region, with arrows pointing at the SP with irregularly shaped cells (black arrows) and intercellular spaces (red arrows). **B** Symptomatic region, with arrows pointing the Ead with hypertrophic cells (white arrows), PP with hyperplastic cells (black arrows) and SP with reduced, obliterated intercellular spaces (red arrows). **C** Symptomatic region, with arrows pointing the SP with hypertrophic cells (black arrows) and reduced intercellular spaces (red arrows). **D** Healthy region, with arrows pointing the SP with cell normal nuclei (black arrows), and with intercellular space (red arrow). **E)** Symptomatic region, with arrows pointing the SP with bulky nuclei in the cells (black arrows) and obliterated intercellular spaces (red arrow). **F** Symptomatic region, with arrows pointing the SP cells stained with toluidine blue (black arrows), evidencing the presence of phenolic compounds. **G** Healthy region, with arrows pointing the Ead and Eab (black arrows), PP (white arrow) and SP (red arrows) with little evidenced phenolic compounds. **H** Symptomatic region, with arrows pointing the Eab (black arrow), PP (white arrows) and SP (red arrow) with phenolic compound depositions. Genotypes: unidentified, originated from somatic embryogenesis (**A**, **C**–**E**), TSH 1188 (**B**, **G**, **H**) and PC 209 (**F**). Ead – Adaxial surface epidermis; Eab – Abaxial surface epidermis; PP—Palisade parenchyma; SP—Spongy parenchyma. Scales: A–C, F–H = 50 µm; D and E = 10 µm. Images A–C and F–H were taken at 1000 × magnification, while images D and E were taken at 200 × magnification

longer than 200 nt were subjected to sequence-similarity searches against public sequence databases, which identified six putative viral sequences. Half of the contigs (three contigs each and identity above 69%) showed significant similarity to viruses from the *Caulimoviridae* family, while the remaining sequences (identity above 76%) were related to sequences of viruses from family *Geminiviridae*. Table S3 shows the size, coverage, identity, and e-value of each contig against its best hit. We enhanced the assembly with a contig extension process, resulting in four viral contigs: three similar to caulimovirids and one related to geminivirids species. The details of each contig are described in Table S3, 2nd assembly.

Genome structure of T. cacao-associated viruses

The genome reconstruction using gap-driven oligonucleotides and Sanger sequencing resulted in two genomes of 8,117 and 3,718 nt, corresponding to tentative badnavirus

and geminivirid, respectively (Figures S2 and S4 and Table S4). The badnavirus sequence showed high identity (100%, Table S4) with cacao swollen shoot Ghana S virus, likely representing a new strain. Ghana S virus is one of several distinct badnaviruses that can cause cacao swollen shoot disease (CSSD), indicating that it belongs to a broader group of phylogenetically related, yet individually distinguishable, viruses that are all capable of causing the disease [7]. The identity matrix revealed that the genome assembled in this study shares 100% nucleotide identity with CSSD isolates, indicating a close genetic relationship (Table S4). Indeed, both the length and genomic structure are consistent with a described full-length badnavirus previously described [14], which typically ranges from six to nine kb, encoding two proteins: RNA-dependent RNA polymerase (RdRp) and a polyprotein (Fig. S2). We observed that the contig related to the reverse transcriptase gene associated with Long Terminal Repeats (RT-LTR) exhibited a high nucleotide sequence identity (over 99%) with the reference badnavirus genome. However, while the RT-LTR region showed only two mismatches and 100% coverage, the polyprotein segment (contig 7) accumulated 242 mismatches and 14 gaps, resulting in 46% coverage of the respective segment (Table S3).

On the other hand, the geminivirid-related sequence was fully assembled and encodes the canonical proteins expected for this viral family, including two hypothetical proteins, capsid, movement, and replication-associated proteins (Figure S2B). We identified an intronic region between the replication-associated proteins after annotating the viral contigs (Figures S2 and S4). This region, previously described in some geminivirids, is spliced during viral replication [47].

Phylogenetic characterization of newly identified viral sequences

After performing the complete assembly of both viral genomes, we evaluated their relationship with closely related viruses by analyzing their genomic information. The badnavirus-related sequence showed high nucleotide similarity (identity values range from 91.7 to 99.88%) to viruses of the cacao swollen shoot virus complex (CSSV) (Table S3), which has previously been associated with a severe disease in *T. cacao*, known as CSSD [1]. A previous study conducted a phylogenetic analysis using different CSSV ORF3 gene sequences (RT RNase H), identifying 12 subgroups based on differences in their sequence information [50]. Following this classification and using a similar analysis, we were able to assign the badnavirus identified in our study to Type VI CSSV (Fig. 3). As a result, we classified it as an isolate of cacao swollen shoot Ghana S virus (CSSGSV) and named it the Brazilian cacao swollen shoot Ghana S virus Ilhéus isolate to reflect its host origin.

The sequence related to geminivirids showed nucleotide identity ranging from 24 to 75.9% to previously described species (see blast result described in Table S3). Our phylogenetic reconstruction placed the assembled sequence within the *Citlodavirus* genus, alongside camellia chlorotic dwarf-associated virus (species *Citlodavirus camelliae*), citrus chlorotic dwarf-associated virus (*Citlodavirus citri*), passion fruit chlorotic mottle virus (*Citlodavirus passiflorae*), and Myrica rubra citlodavirus 1 (*Citlodavirus myricae*) (Fig. 4). The percentage of identity at nucleotide level with public sequences (~76%, ICTV criteria < 78%) indicates it is likely a new viral species (Figure S6 and Table S3). The novel viral genome identified in Brazil was named Theobroma citlodavirus (ThCTV), and we suggest *Citlodavirus theobromae* as the binomial name to the tentative new species.

Small RNA-mediated antiviral defense distinguishes viral sequences in Theobroma cacao

We leveraged our sRNA deep sequencing data to investigate the sRNAs produced by plant PTGS pathways. Our results revealed a distinct vsiRNA profile derived from the putative viral sequences, which is consistent with PTGS activation. This profile comprised a symmetric accumulation of 20–24 nt-long reads, without any clear enrichment of a specific 5' base (Fig. 5A).

The sRNAs profiles from the badnavirus CSSGSV displayed a distinct pattern from that observed for the citlodavirus ThCTV, demonstrating distinct molecular signatures. While CSSGSV showed an sRNA size distribution and base enrichment typical of exogenous RNAs targeted by the PTGS pathways, ThCTV exhibited a profile analogous to the processing of endogenous RNAs (Fig. 5A) [24]. However, it is important to highlight that ThCTV codes for a V2-like protein similar to the V2 protein of citrus chlorotic dwarf-associated virus that was shown to suppress plant RNA silencing [9, 37, 81].

The sRNA coverage also differed between sequences derived from the badnavirus and the citlodavirus. CSSGSV-derived sRNAs displayed symmetrical and uniform coverage along the segment in coding regions, whereas the contig related to the geminivirid showed uneven sRNA density, with high abundance at the 3' end of the genome, predominantly covering the replication-associated genes (Fig. 5B). Similar to other citlodaviruses, we also observed a lack of RNAs between the replication-associated proteins, suggesting that this region is likely excised after transcription [62, 67] (Fig. 5B).

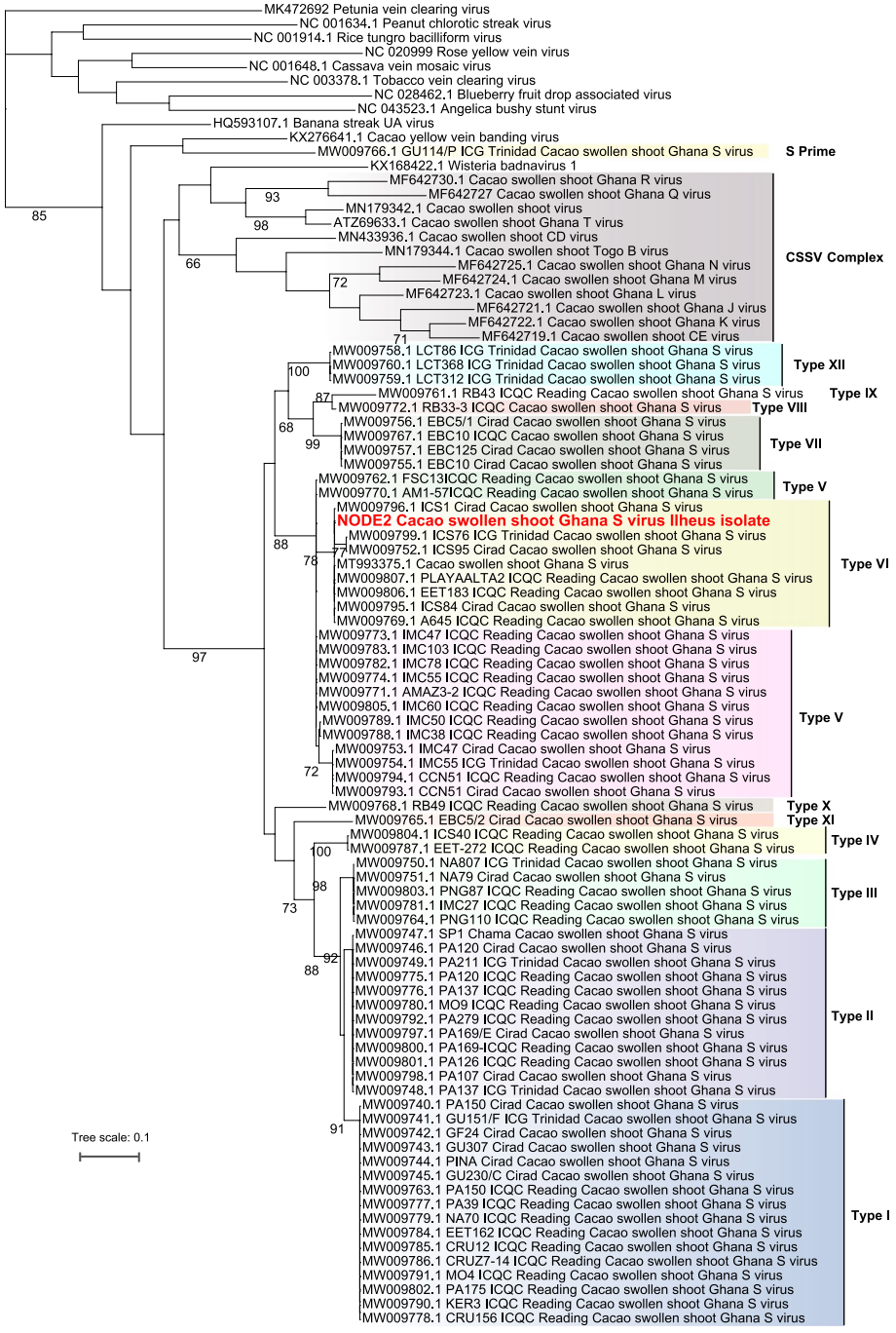

Fig. 3 Maximum likelihood tree analysis of viral sequence related to members of the family *Caulimoviridae*. The phylogeny used sequences of 92 cacao swollen shoot Ghana S virus (CSSGSV) publicly available, and the new strain identified in our study, based on nucleotide sequence of the ORF3 gene. Twelve subtypes of CSSGSV are shown. The badnavirus genome recovered in this study is in red letters with subtype VI

Identification of virus-derived peptides suggests active infection

Proteomics analysis was conducted on lyophilized leaf biomass from *T. cacao* scion/rootstock combinations (CCN-51/CCN-51, BN-34, PS-1319, and PH-16) to confirm the viruses detected by sRNA sequencing and evaluate viral protein translation and viability. Gel-free proteomics experiments allowed us to identify 105 (47 unique) virus-derived peptides that showed exclusive similarity to the viral proteins of the viruses identified in this study (Fig. 5C). The peptides ranged from 13 to 47 amino acids, with an average length of ~ 23 amino acids, and exhibited 100% identity to those deduced from the sequenced genomes (Table S2). Notably, in at least two

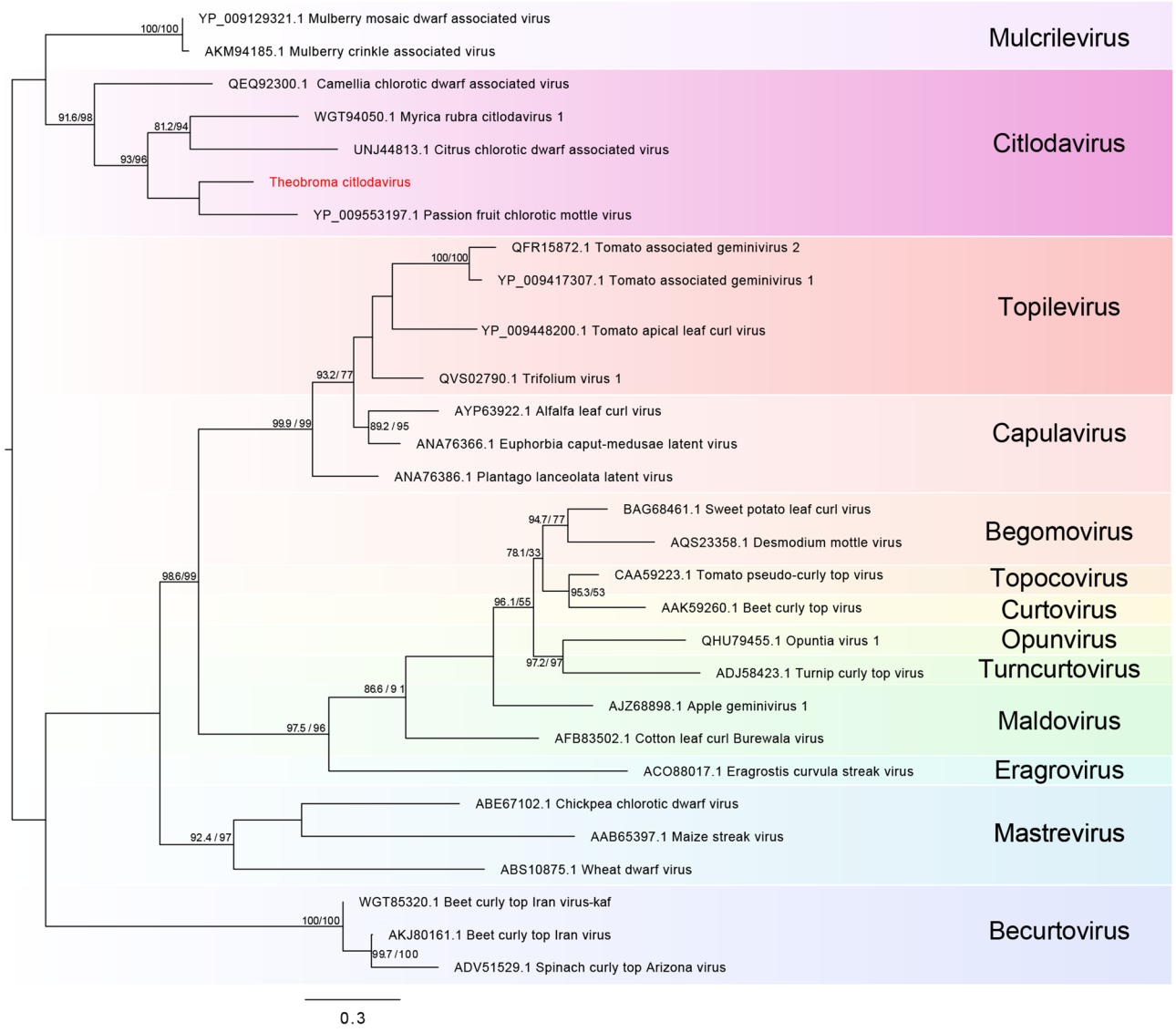

Fig. 4 Maximum likelihood tree of the viral sequence related to family *Geminiviridae*. The phylogenetic analysis of the new geminivirid was conducted based on the protein sequence of the replicase gene. Sequences from members from all genera of the *Geminiviridae* family were included

out of the four experiments, we were able to identify two or more distinct peptides for each of the viral genomes, reinforcing that these viral sequences are transcribed and translated into proteins (Fig. 5C and Table S2).

Chlorotic symptoms are linked to the citlodavirus presence

Since both viruses produce proteins and potentially establish active infections, we aimed to investigate whether both CSSGV and ThCTV are associated specifically with symptomatic leaves. Leaf samples from three *T. cacao* varieties—CCN51, CEPEC 2002, and BN34 — were analyzed under different conditions (Figs. 6 and S6). For CCN51, we examined symptomatic leaves, asymptomatic leaves from symptomatic plants, and asymptomatic leaves from entirely symptomless plants (Fig. 6). Geminivirid sequence was detected in both symptomatic and asymptomatic leaves from symptomatic plants; however, it was absent in leaves derived from asymptomatic individuals. In contrast, badnavirus sequences were detected in all samples tested, suggesting that the geminivirid presence is specifically associated with symptomatic individuals. To further evaluate this association, we analyzed other *T. cacao* varieties. Similar to CCN51, citlodaviruses were detected in both symptomatic and asymptomatic leaves from symptomatic plants of the CEPEC 2002 clone. Conversely, the geminivirid was not detected in leaves from a completely asymptomatic plant of the BN34 variety (Table S5). These results support a specific linkage between the presence of this virus and symptom expression in susceptible cacao varieties. Since

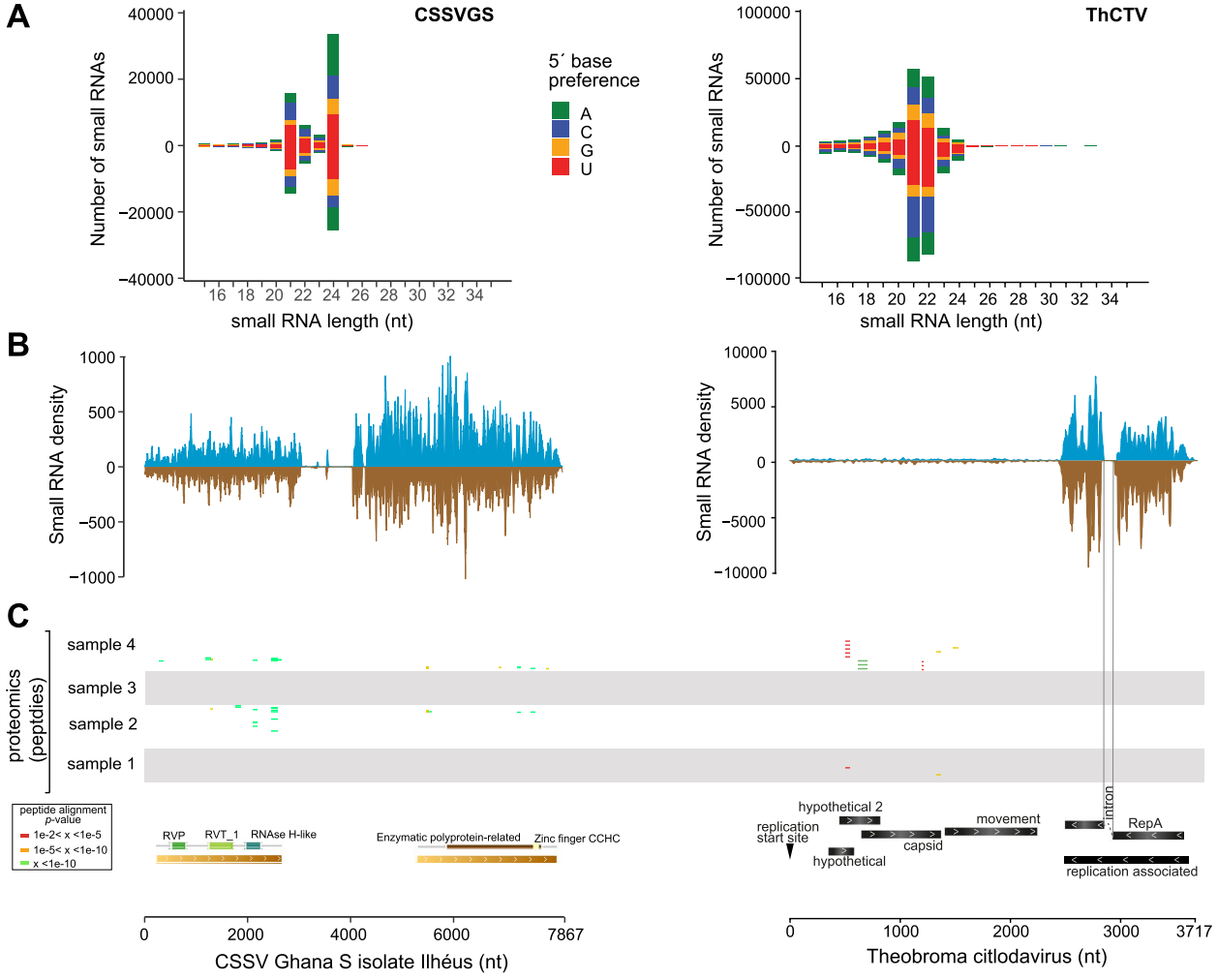

Fig. 5 Small RNA and peptides profile derived from *T. cacao*-associated virus. Small RNA size distribution and 5' base preference (**A**), distribution of small RNAs from viral genomes (**B**), and location of peptides identified (**C**) for CSSVGS (left panel and ThCTV (right panels). Positive and negative values in A and B reflect sequences that matched the sense and antisense strands of the viral genome, respectively

both symptomatic and non-symptomatic leaves from the same plant at the same development stage were initially found to harbor the virus, we next investigated whether chlorotic symptoms correlate with ThCTV RNA abundance. Viral RNA was detected exclusively in symptomatic leaves, whereas all asymptomatic leaf samples ($n = 3$) tested negative for viral presence (Fig. 6). This finding suggests a strong association between symptom manifestation and active viral replication.

Discussion

The state of Bahia, located in the northeastern region of Brazil, is the leading producer of cocoa in the country, with over 400,000 hectares of plantations distributed across 81 municipalities. Cocoa cultivation in Bahia is predominantly carried out under the 'cabruca' agroforestry system (accounting for 60% of total cultivation), which involves growing cocoa trees beneath the canopy of native Atlantic Forest species, typically along rivers, valleys, and hillsides. This system contributes significantly to the conservation of the Atlantic Forest biome. Cultivation is based on clonal cocoa varieties released and recommended by the *Comissão Executiva do Plano da Lavoura Cacaueira* (CEPLAC). These clones are selected for their high productivity, favorable fruit size and number, and resistance to witches' broom and other common cocoa diseases in the region [41]. Currently, 18 clonal varieties are officially recommended for planting, including CCN51, CEPEC2002, and BN34, which show notable resistance to witches' broom disease [4, 13]. In this study, we identified two DNA viruses infecting *T. cacao* belonging to two distinct families, *Caulimoviridae* and *Geminiviridae*. In both cases, we found evidence of their active replication by analyzing their sRNA profiles, which indicated that the PTGS antiviral response was triggered by the host cells, and presence of

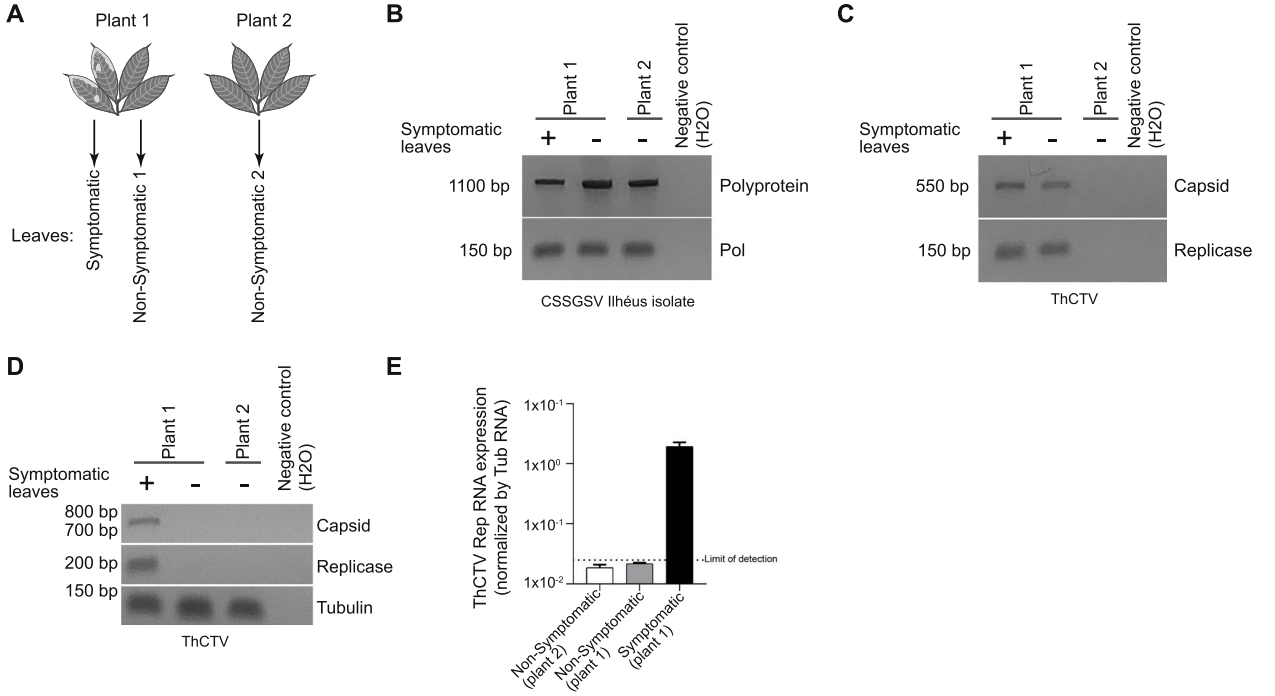

Fig. 6 Amplification of ThCTV and CSSGSV in symptomatic and asymptomatic *T. cacao* leaves. **A** Schematic representation of symptomatic and non-symptomatic leaves collected from *T. cacao*. **B**, **C** Agarose gels showing the amplification of the cacao swollen shoot Ghana S virus (CSSGSV) polyprotein and polymerase (Pol) genes (**B**) and capsid and replicase genes of Theobroma citlodavirus (ThCTV) (**C**) in plant with leaves with and without (-) chlorotic areas. **D** Agarose gel showing the cDNA amplification of capsid, replicase, and tubulin (endogenous control) genes in symptomatic (+) and asymptomatic (-) leaves (**E**) RNA levels of the ThCTV replicase gene normalized by tubulin gene in symptomatic and asymptomatic leaves of *T. cacao*

viral translational products was confirmed by proteomic assays. PTGS is a mechanism used by plants to silence the expression of both host and viral genes, thereby negatively impacting their translation. Our results indicate that both viruses are actively replicating in cacao cells and trigger an immune response [3]. As described in previous studies, the vsiRNA signature (nucleotide base distribution and peaks) is specific to each virus species. When the signatures differ, it is assumed that the virus profiles are from different viral species [2, 3]. According to our phylogenetic analysis, the virus belonging to the family *Caulimoviridae* was assigned to the genus *Badnavirus*. Caulimovirids can integrate into their host genome, and their presence may not be associated with viral infections. The genome identified in our study has not been recognized as a badnavirus species by ICTV yet. Indeed, the CSSGSV has recently been considered an Endogenous Viral Element (EVE) present in the genome of *T. cacao* [50]. More recently, a complete badnavirus-like genome integrated into the genome of cacao germplasm has been reported [73]. Nevertheless, EVEs derived from badnaviruses can give rise to episomal viral infections when the host is subjected to different stress conditions [53]. In some cases, systemic viral infections can occur due to induction by abiotic stresses, such as temperature changes, lack of water, and nutritional deficits [39].

In our study, the genome of the *cacao swollen shoot Ghana S virus* (CSSGSV) was assembled from a sRNA library, and its presence was confirmed in leaf tissues from multiple *T. cacao* genotypes through the detection of both viral-derived sRNAs and peptides by mass spectrometry. These findings suggest that the viral genome is transcriptionally and translationally active. However, despite this molecular evidence, we cannot conclusively determine whether the activity originates from an exogenous, replicating virus or from an Endogenous Viral Element (EVE) integrated into the host genome. Attempts to recover the circular viral genome using amplification with high fidelity enzyme were unsuccessful, and transmission electron microscopy (TEM) did not reveal the presence of viral particles—likely due to sample constraints. These limitations argue against the presence of a systemic, actively replicating virus and support the possibility that the badnavirus-like sequences correspond to transcriptionally active EVEs, as previously reported in *T. cacao* [50, 73]. Given that EVEs derived from badnaviruses can be transcriptionally activated under abiotic stress conditions [39, 53], further investigation—such as in situ hybridization or quantification of viral RNA and DNA across environmental conditions—is necessary to elucidate whether the observed transcription reflects latent EVE activation or active viral infection (episomal).

The second virus sequence identified in our study grouped with other species from the genus *Citlodavirus* within the family *Geminiviridae*. To support the taxonomic assignment of ThCTV to the genus *Citlodavirus*, we performed a comparative genomic analysis with representative geminivirids. The assembled genome exhibited the typical organization observed in members of genus Citlodavirus, including the presence of conserved ORFs encoding the replication-associated protein (Rep), coat protein (CP), and movement protein (MP), arranged in the characteristic bidirectional orientation. A conserved sequence motif at the origin of virion strand replication, such as the nonanucleotide sequence 5`-TAAT ATTAC-3` was identified. Other motifs involved in rolling circle replication (RCR), such as FLTY (motif I) and HLH (motif II) were identified in the Rep-associated genes, further supporting its classification. In addition, we identified putative viral suppressor of RNA silencing (VSR), which are known to play key roles in evading PTGS and promoting systemic infection. The V2 protein has also been implicated in facilitating viral movement and enhancing the spread of the virus throughout the plant [25, 40, 80]. Viruses in this family are classified into 15 different genera (https://ictv.global/vmr) based on their genomic organization, host range, and vector insects (such as leafhoppers, aphids, and whiteflies) [23]. Members of this group are important pathogens, causing diseases in economically relevant plant species worldwide, and are considered the viruses that most impact crops [23, 82]. However, although geminivirids have been identified in numerous plants, to date, no member of the family has been found to infect *T. cacao*. This report enhances our understanding of the host range of this viral family and expands the number of diseases that may be caused by its members.

In the cacao genotypes analyzed, both viral genomes—badnavirus and geminivirid—were detected. However, in asymptomatic plants, only the badnavirus was present (episomal badnavirus was not detected). This pattern suggests that the geminivirid (*ThCTV*) is the primary agent associated with the chlorotic symptoms observed. Although *ThCTV* likely drives symptom expression, the contribution of the badnavirus cannot be excluded, as members of this genus can cause similar symptoms in plants, such as chlorotic spots, mosaic patterns, leaf malformations, and curling [32, 55] —which closely resemble those observed in cacao plants from Ilhéus, Bahia, Brazil, with the exception of leaf curling. Moreover, mixed infections involving distinct viruses have been reported to exacerbate disease symptoms due to synergistic interactions, such as enhanced viral replication or suppression of host defenses [66, 70]. Although our study did not directly assess this possibility, we recognize that co-infection with CSSGSV and ThCTV could synergistically intensify symptoms. Future experimental approaches involving single and combined inoculations will be important to clarify the extent of such interactions.

One limitation of our study was the absence of sequencing data from asymptomatic leaves. Nonetheless, the amplification results obtained from leaves of different *T. cacao* genotypes consistently indicate that only ThCTV was detected under chlorotic conditions.

The viruses assembled and detected in our study may pose emerging hazards for cocoa cultivation in the tropical Americas. For instance, badnaviruses are economically important to cocoa plants worldwide [56], particularly in West Africa, where infection by CSSD leads to a decrease in seed quality and production [1, 8]. Badnaviruses are known for their potential for intra- and interspecific recombination, an important factor in generating genomic variability and, consequently, species diversity [57]. Furthermore, the establishment of mixed viral infections in host plants is crucial for the genetic recombination of different species and strains [12, 17, 26, 27, 54, 56]. Geminivirids are known to reduce the fruit and seed production of various crops. The new species identified in *T. cacao* may hinder the plant's development and compromise cocoa production. The prevalence, damage, and economic importance of this virus in cocoa crops are still unknown, and Kochs postulates need to be applied to confirm a new plant pathogen. Recently, there has been an increase in the number of new geminivirid species and genera described in both cultivated and non-cultivated plant hosts through metagenomics-based approaches [75].

The anatomical alterations observed in the leaf blades of infected tissues can be attributed to viral invasion and replication, as well as to the plant's defense responses aimed at limiting the spread of the infection to other tissues. We observed histopathological effects caused by the viruses, detected through molecular analysis of chlorotic leaves, such as hyperplasia and hypertrophy of the palisade and spongy parenchyma cells, as well as alterations in nuclear volume The anatomical alterations associated with geminivirids in *Althea rosea* (*Althea rosea enation virus* – [AREV]) and *Papaver somniferum* (*Ageratum enation virus* – [AEV]) were different from those reported in our study, with changes in vascular tissues (disorganized vascular bundles, highly lignified xylem cells, and collapsed phloem cells) in leaves and stems, along with cell death. However, hyperplasia in the leaf blade cells and hypertrophy of the cortical parenchyma cells of *P. somniferum* were observed [11, 69]. Regarding the phenolic compound accumulation, it has been observed that it is a regular defense against pathogens, as observed in the palisade parenchyma of grape leaves infected with *Phakopsora euvitis* [58].

In summary, we identified two viruses associated with cacao plants in Ilhéus, Bahia, Brazil. One of them is a new viral genome from the *Geminiviridae* family, and the other is a new strain of CSSGSV (genus *Badnavirus*). These findings are significant in the agricultural context since cocoa cultivation has relevant economic importance worldwide. Furthermore, the emergence and establishment of new pests, particularly viruses, can cause serious harm to crops, including the potential cessation of cultivation in certain regions and a decrease in production. Therefore, it is necessary to carry out new assays to identify which symptoms are caused by each species of virus and to understand the course of the symptoms to assess the real impact of those viruses on *T. cacao* production.

Supplementary Information

The online version contains supplementary material available at https://doi.org/10.1186/s12870-025-07472-z.

Supplementary Material 1: Figure S1.Schematic representation of the study methodology.Symptomatic and non-symptomatic leaves were collected from different ***Theobroma cacao*** varieties. After collection, the material was used for total RNA and protein extraction, as well as microscopy. Total RNA from leaves were extracted and small RNA sequencing were performed. PCR amplification was performedfor the presence of viral sequences.

Supplementary Material 2: Figure S2. Viral genomic structure. Representation of the badnavirus (A) and geminivirids (B) genomes. The genome organization of the badnavirus includes domains for RVP (retroviral protease), RVT_1 (reverse transcriptase), RNAse H-like, enzymatic polyprotein-related regions, and a Zinc finger CCHC motif (A). The genome structure of geminivirid include hypothetical proteins, capsid, movement, RepA, and replication-associated proteins. The replication start site is indicated with a black arrowhead, and an intron is also marked (B). Green arrowheads represent primer binding sites designed for amplification of viral regions. Three pairs of primers were designed and used. Badnavirus (A): One primer pair amplifies a 150 nt fragment in the polyprotein gene, a second primer pair amplifies a region in the middle of the genome of approximately 1,500 nt, and a third primer pair amplifies another polyprotein fragment of 150 nt. Citlodavirus (B): One primer pair amplifies the capsid gene with a 600 nt product, another pair amplifies the replication-associated protein region with a product of approximately 150 nt, and a third primer pair closes the viral genome, being present at both the end and the beginning of the sequence. Red arrows indicate prime pairs that were used for Sanger Sequencing and posterior closing of the viral genomes. Red arrows indicate the sense of amplification.

Supplementary Material 3: Figure S3. ***T. cacao*** leaves from different genotypes presenting chlorosis spots. White arrows indicate chlorosis spots in the leaves. The name of the variety is indicated above the leaf.

Supplementary Material 4: Figure S4. Circular genome of the geminivirid identified in the study. Protein-coding genes are indicated in purple and introns in orange. GC content is shown throughout the genome in black

Supplementary Material 5: Figure S5. Bubble plot indicating pairwise comparison of Theobroma citlodavirus and closely related species. Coverage is represented by circle length while identity is represented by color. Comparisons with identity equal or higher than 80% are represented by red borders

Supplementary Material 6: Figure S6. Amplification of ThCTV genomic regions in ***symptomatic and asymptomatic T. cacao leaves.*** Agarose gel showing the amplification bands for the plant tubulin and capsid and replicase genes of Theobroma citlodavirus ***(ThCTV) in T. cacao*** varieties BN34 and CEPEC 2002. For CEPEC, ***individuals with different symptomatic profiles were assessed.***

Supplementary Material 7: Figure S7. Uncropped gels images of badnavirus and citlodavirus amplification in CCN51 and CEPEC varieties

Supplementary Material 8: Table S1. Pairwise nucleotide comparison of ThCTV and closely related geminivirid species

Supplementary Material 9: Table S2. Peptides identified on leaves biomass of ***T. cacao*** genotypes derived from viral sequences identified in our study.

Supplementary Material 10: Table S3. Information of sequence similarity search of the assembled viral contigs from the geminivirid and caulimovirid against public databases.

Supplementary Material 11: Table S4. Identity matrix of badnavirus sequences used in the comparative analysis with the badnavirus genome assembled in this study

Supplementary Material 12: Table S5. Summary of viral detection (badnavirus and geminivirus) in symptomatic and asymptomatic cacao leaves across different genotypes.

Supplementary Material 13: Supplementary File 1.Sequences assembled in our study

Acknowledgements
We would like to thank all colleagues who helped in the construction of the work and discussion of the results.

Authors' contributions
PLCF, MSS, JSS, JNA, CHCN, RSN, JPNS, GRRP, IJSF, YMO, JFA, PLRG, JOS, VFF, WBL, CPP, ARO and ERGRA performed the laboratory and bioinformatic data analysis. ERGRA developed the study concept and supervised the project. All author read and approved the final version of the manuscript.

Funding
Our study was supported by CAPES (Financial code 001), CNPq and FAPEMIG (process code BPD-00820–22). JFA, CPP and ERGRA are reseach fellows from CNPq.

Data availability
The small RNA deep sequencing library produced in this study can be found in the NCBI SRA database under project number PRJNA947677. Viral sequences described in our study were deposited at NCBI Nucleotide database (https://www.ncbi.nlm.nih.gov/nucleotide/) under deposit numbers BankIt2956920 and BankIt2956979 and in Supplementary File 1. The full uncropped gels images are available in Supplementary Figure S7. The proteomics data were submitted in MassIVE Repository and are available under the accession https://doi.org/10.25345/C5TT4FW94 (https://massive.ucsd.edu/ProteoSAFe/dataset.jsp?accession=MSV000089048).

Declarations

Ethics approval and consent to participate
Not applicable.

Consent for publication
Not applicable.

Competing of interests
The authors declare no competing interests.

Author details
[1]Departamento de Genética, Instituto de Ciências Biológicas, Universidade Federal de Minas Gerais, Belo Horizonte, Minas Gerais 30270-901, Brazil
[2]Departamento de Engenharias e Computação, Universidade Estadual de Santa Cruz (UESC), Rodovia Jorge Amado Km 16, Ilhéus, Bahia 45662-900, Brazil
[3]Departamento de Bioquímica E Imunologia, Instituto de Ciências Biológicas, Universidade Federal de Minas Gerais, Belo Horizonte, Minas Gerais 30270-901, Brazil

[4]Escola Superior de Agricultura "Luiz de Queiroz" - Universidade de São Paulo, Piracicaba, São Paulo 13418-900, Brazil
[5]Embrapa Mandioca E Fruticultura, Cruz das Almas, Bahia 7–44370-000, Brazil
[6]Instituto Biológico de São Paulo, São Paulo, São Paulo 04014-900, Brazil
[7]Departamento de Ciências Básicas, Faculdade de Zootecnia E Engenharia de Alimentos, Universidade de São Paulo, Pirassununga, São Paulo 13635900, Brazil
[8]Departamento de Ciências Agrárias E Ambientais DCAA, Universidade Estadual de Santa Cruz (UESC), Rodovia Jorge Amado Km 16, Ilhéus, Bahia 45662-900, Brazil

Received: 20 May 2025 / Accepted: 22 September 2025
Published online: 21 October 2025

References

1. Abrokwah F, Dzahini-Obiatey H, Galyuon I, Osae-Awuku F, Muller E. Geographical distribution of *cacao swollen shoot virus* molecular variability in Ghana. Plant Dis. 2016;100:2011–7. https://doi.org/10.1094/PDIS-01-16-0081-RE.
2. Aguiar ERGR, Olmo RP, Marques JT. Virus-derived small RNAs: molecular footprints of host-pathogen interactions: Virus-derived small RNAs. WIREs RNA. 2016;7:824–37. https://doi.org/10.1002/wrna.1361.
3. Aguiar ERGR, Olmo RP, Paro S, Ferreira FV, de Faria IJDS, Todjro YMH, Lobo FP, Kroon EG, Meignin C, Gatherer D, Imler JL, Marques JT. Sequence-independent characterization of viruses based on the pattern of viral small RNAs produced by the host. Nucleic Acids Res. 2015;43:6191–6206. https://doi.org/10.1093/nar/gkv587.
4. Alexandre RS, Chagas K, Marques HIP, Costa PR, Cardoso J. Caracterização de frutos de clones de cacaueiros na região litorânea de São Mateus, ES. Rev Bras Eng Agric Ambient. 2015;19:785–90. https://doi.org/10.1590/1807-1929/agriambi.v19n8p785-790.
5. Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. Basic local alignment search tool. J Mol Biol. 1990;215:403–10. https://doi.org/10.1016/S0022-2836(05)80360-2.
6. Alverson WS, Whitlock BA, Nyffeler R, Bayer C, Baum DA. Phylogeny of the core Malvales: evidence from *ndhF* sequence data. Am J Bot. 1999;86:1474–86.
7. Ameyaw GA, Domfeh O, Gyamera E. Epidemiology and diagnostics of cacao swollen shoot disease in Ghana: past research achievements and knowledge gaps to guide future research. Viruses. 2023;16:43. https://doi.org/10.3390/v16010043.
8. Ameyaw GA, Dzahini-Obiatey HK, Domfeh O. Perspectives on cocoa swollen shoot virus disease (CSSVD) management in Ghana. Crop Prot. 2014;65:64–70. https://doi.org/10.1016/j.cropro.2014.07.001.
9. Ashfaq MA, Dinesh Kumar V, Soma Sekhar Reddy P, Anil Kumar C, Sai Kumar K, Narasimha Rao N, et al. Post-transcriptional gene silencing: basic concepts and applications. J Biosci. 2020;45:128. https://doi.org/10.1007/s12038-020-00098-3.
10. Bankevich A, Nurk S, Antipov D, Gurevich AA, Dvorkin M, Kulikov AS, et al. SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing. J Comput Biol. 2012;19:455–77. https://doi.org/10.1089/cmb.2012.0021.
11. Bigarré L, Chazly M, Salah M, Ibrahim M, Padidam M, Nicole M, et al. Characterization of a new begomovirus from Egypt infecting Hollyhock (*Althea rosea*). Eur J Plant Pathol. 2001;107:701–11. https://doi.org/10.1023/A:1011967232319.
12. Bömer M, Rathnayake AI, Visendi P, Silva G, Seal SE. Complete genome sequence of a new member of the genus *Badnavirus*, *Dioscorea* bacilliform RT virus 3, reveals the first evidence of recombination in yam badnaviruses. Arch Virol. 2018;163:533–8. https://doi.org/10.1007/s00705-017-3605-9.
13. CEPEC. Recomendação Clonal de Cacaueiro 2018 - Cepec-2176 e cepec-2204. 2022. https://www.gov.br/agricultura/pt-br/assuntos/ceplac/publicacoes/cultivares/bannerclones-def2.pdf/view.
14. Chingandu N, Kouakou K, Aka R, Ameyaw G, Gutierrez OA, Herrmann H-W, et al. The proposed new species, cacao red vein virus, and three previously recognized badnavirus species are associated with cacao swollen shoot disease. Virol J. 2017;14:199. https://doi.org/10.1186/s12985-017-0866-6.
15. CONAB – Companhia Nacional de Abastecimento. 2018. Cacau amêndoa – análise mensal. . https://www.conab.gov.br/info-agro/analises-do-mercado-agropecuario-e-extrativista/analises-do-mercado/historico-mensal-de-cacau.
16. Coslovsky. Oportunidades para aprimoramento da na Amazônia Brasileira. 2023.
17. Dallot S, Acuña P, Rivera C, Ramírez P, Côte F, Lockhart BE, et al. Evidence that the proliferation stage of micropropagation procedure is determinant in the expression of banana streak virus integrated into the genome of the FHIA 21 hybrid (Musa AAAB). Arch Virol. 2001;146:2179–90. https://doi.org/10.1007/s007050170028.
18. Darriba D, Posada D, Kozlov AM, Stamatakis A, Morel B, Flouri T. ModelTest-NG: a new and scalable tool for the selection of DNA and protein evolutionary models ed Keith Crandall. Mol Biol Evol. 2020;37:291–294. https://doi.org/10.1093/molbev/msz189.
19. De Souza PA, Moreira LF, Sarmento DHA, Da Costa FB. Cacao— Theobroma cacao. In Exotic Fruits, Elsevier. 2018 pp. 69–76. https://doi.org/10.1016/B978-0-12-803138-4.00010-1.
20. Doyle J. DNA protocols for plants. In Molecular Techniques in Taxonomy, eds. Godfrey M. Hewitt, Andrew W. B. Johnston, and J. Peter W. Young. Berlin, Heidelberg: Springer Berlin Heidelberg. 1991 pp. 283–293. https://doi.org/10.1007/978-3-642-83962-7_18.
21. Espinal RBA, De Santana SF, Santos VC, Lizardo GNR, Silva RJS, Corrêa RX, et al. Uncovering a complex virome associated with the cacao pathogens *Ceratocystis cacaofunesta* and *Ceratocystis fimbriata*. Pathogens. 2023;12(2):287. https://doi.org/10.3390/pathogens12020287.
22. Feder N, O'Brien TP. Plant microtechnique: some principles and new methods. Am J Bot. 1968;55:123–42. https://doi.org/10.1002/j.1537-2197.1968.tb06952.x.
23. Fiallo-Olivé E, Lett JM, Martin DP, Roumagnac P, Varsani A, Zerbini FM, Navas-Castillo J. ICTV virus taxonomy profile: Geminiviridae 2021: This article is part of the ICTV Virus Taxonomy Profiles collection. J Gen Virol. 2021;102. https://doi.org/10.1099/jgv.0.001696.
24. Fonseca PLC, Badotti F, De Oliveira TFP, Fonseca A, Vaz ABM, Tomé LMR, et al. Virome analyses of *Hevea brasiliensis* using small RNA deep sequencing and PCR techniques reveal the presence of a potential new virus. Virol J. 2018;15:184. https://doi.org/10.1186/s12985-018-1095-3.
25. Fontenele RS, Abreu RA, Lamas NS, Alves-Freitas DMT, Vidal AH, Poppiel RR, et al. Passion fruit chlorotic mottle virus: molecular characterization of a new divergent geminivirus in Brazil. Viruses. 2018;10:169. https://doi.org/10.3390/v10040169.
26. García-Andrés S, Monci F, Navas-Castillo J, Moriones E. Begomovirus genetic diversity in the native plant reservoir *Solanum nigrum*: evidence for the presence of a new virus species of recombinant nature. Virology. 2006;350:433–42. https://doi.org/10.1016/j.virol.2006.02.028.
27. Graham AP, Martin DP, Roye ME. Molecular characterization and phylogeny of two begomoviruses infecting *Malvastrum americanum* in Jamaica: evidence of the contribution of inter-species recombination to the evolution of malvaceous weed-associated begomoviruses from the Northern Caribbean. Virus Genes. 2010;40:256–66. https://doi.org/10.1007/s11262-009-0430-6.
28. Huang X, Madan A. CAP3: a DNA sequence assembly program. Genome Res. 1999;9:868–77. https://doi.org/10.1101/gr.9.9.868.
29. I Sardi SH, Carvalho RC, Pacheco LG, PD Almeida JP, MDA Belitardo EM, S Pinheiro C, S Campos G, RGR Aguiar E. 2020. High-quality resolution of the outbreak-related Zika virus genome and discovery of new viruses using ion torrent-based metatranscriptomics. Viruses 12:782. https://doi.org/10.3390/v12070782.
30. IBGE. 2023. Produção Agrícola - Lavoura Permanente.
31. Ivanova AA, Wegner C, Kim Y, Liesack W, Dedysh SN. Identification of microbial populations driving biopolymer degradation in acidic peatlands by metatranscriptomic analysis. Mol Ecol. 2016;25:4818–35. https://doi.org/10.1111/mec.13806.
32. Jacquot E, Hagen LS, Michler P, Rohfritsch O, Stussi-Garaud C, Keller M, et al. In situ localization of cacao swollen shoot virus in agroinfected *Theobroma cacao*. Arch Virol. 1999;144:259–71. https://doi.org/10.1007/s007050050502.
33. Johnson WE. Origins and evolutionary consequences of ancient endogenous retroviruses. Nat Rev Microbiol. 2019;17:355–70. https://doi.org/10.1038/s41579-019-0189-2.
34. Katoh K. MAFFT: a novel method for rapid multiple sequence alignment based on fast Fourier transform. Nucleic Acids Res. 2002;30:3059–66. https://doi.org/10.1093/nar/gkf436.
35. Koeppe S, Kawchuk L, Kalischuk M. RNA interference past and future applications in plants. IJMS. 2023;24:9755. https://doi.org/10.3390/ijms24119755.

36. Langmead B. Aligning short sequencing reads with bowtie. Curr Protoc Bioinformatics. 2010;32. https://doi.org/10.1002/0471250953.bi1107s32.
37. Leonetti P, Consiglio A, Arendt D, Golbik RP, Rubino L, Gursinsky T, et al. Exogenous and endogenous dsRNAs perceived by plant Dicer-like 4 protein in the RNAi-depleted cellular context. Cell Mol Biol Lett. 2023;28:64. https://doi.org/10.1186/s11658-023-00469-2.
38. Letunic I, Bork P. Interactive Tree Of Life (iTOL) v4: recent updates and new developments. Nucleic Acids Res. 2019;47:W256–9. https://doi.org/10.1093/nar/gkz239.
39. Lheureux F, Carreel F, Jenny C, Lockhart B, Iskra-Caruana M. Identification of genetic markers linked to banana streak disease expression in inter-specific *Musa* hybrids. Theor Appl Genet. 2003;106:594–8. https://doi.org/10.1007/s00122-002-1077-z.
40. Luna AP, Romero-Rodríguez B, Rosas-Díaz T, Cerero L, Rodríguez-Negrete EA, Castillo AG, et al. Characterization of Curtovirus V2 protein, a functional homolog of Begomovirus V2. Front Plant Sci. 2020. https://doi.org/10.3389/fpls.2020.00835.
41. MAPA. Cacau do Brasil. 2022.
42. MAPA. Recentes achados de virose em cacaueiros levam a CEPLAC a adotar medidas preventivas. 2023. https://www.gov.br/agricultura/pt-br/assuntos/ceplac/medidaspreventivasvirosecacaueiros. Accessed 5 Nov 2024.
43. Marques JPR, Nuevo LG. 2022. Double-Staining Method to Detect Pectin in Plant-Fungus Interaction. JoVE. 2022;63432. https://doi.org/10.3791/63432.
44. Martin M. Cutadapt removes adapter sequences from high-throughput sequencing reads. EMBnet j. 2011;17:10. https://doi.org/10.14806/ej.17.1.200.
45. Matthewman CA, Kawashima CG, Húska D, Csorba T, Dalmay T, Kopriva S. Mir395 is a general component of the sulfate assimilation regulatory network in Arabidopsis. FEBS Lett. 2012;586:3242–8. https://doi.org/10.1016/j.febslet.2012.06.044.
46. Miller MA, Pfeiffer W, Schwartz T. The CIPRES science gateway: a community resource for phylogenetic analyses. In Proceedings of the 2011 TeraGrid Conference: Extreme Digital Discovery, Salt Lake City Utah: ACM. 2011 pp. 1–8. https://doi.org/10.1145/2016741.2016785.
47. Moore MJ, Schwartzfarb EM, Silver PA, Yu MC. Differential recruitment of the splicing machinery during transcription predicts genome-wide patterns of mRNA splicing. Mol Cell. 2006;24:903–15. https://doi.org/10.1016/j.molcel.2006.12.006.
48. Motamayor JC, Risterucci AM, Lopez PA, Ortiz CF, Moreno A, Lanaud C. Cacao domestication I: the origin of the cacao cultivated by the Mayas. Heredity. 2002;89:380–6. https://doi.org/10.1038/sj.hdy.6800156.
49. Moura. Coconut crown anomaly. Centro de Pesquisas do Cacau (CEPEC). 2020;32:157–162. https://doi.org/10.21757/0103-3816.2020v32n2p157-162.
50. Muller E, Ullah I, Dunwell JM, Daymond AJ, Richardson M, Allainguillaume J, et al. Identification and distribution of novel badnaviral sequences integrated in the genome of cacao (*Theobroma cacao*). Sci Rep. 2021;11:8270. https://doi.org/10.1038/s41598-021-87690-1.
51. Paim MCA, Luz EDMN, De Souza JT, Cerqueira AO, Lopes JRM. Pathogenicity of *Phytophthora* species to *Anthurium andraeanum* in Brazil. Australas Plant Pathol. 2006;35:275. https://doi.org/10.1071/AP06005.
52. Pirovani CP, Carvalho HAS, Machado RCR, Gomes DS, Alvim FC, Pomella AWV, et al. Protein extraction for proteome analysis from cacao leaves and meristems, organs infected by *Moniliophthora perniciosa*, the causal agent of the witches' broom disease. Electrophoresis. 2008;29:2391–401. https://doi.org/10.1002/elps.200700743.
53. Pooggin MM. Small RNA-omics for plant virus identification, virome reconstruction, and antiviral defense characterization. Front Microbiol. 2018;9:2779. https://doi.org/10.3389/fmicb.2018.02779.
54. Ramos-Sobrinho R, Chingandu N, A Gutierrez O, Marelli J-P, K Brown J. A complex of badnavirus species infecting cacao reveals mixed infections, extensive genomic variability, and interspecific recombination. Viruses. 2020;12:443. https://doi.org/10.3390/v12040443.
55. Ramos-Sobrinho R, Ferro MMM, Lima GSA, Nagata T. First report of Cotton leafroll dwarf virus infecting cacao (*Theobroma cacao*) trees in Brazil. Plant Dis. 2023;107:1251. https://doi.org/10.1094/PDIS-07-22-1570-PDN.
56. Ramos-Sobrinho R, Ferro MMM, Nagata T, Puig AS, Keith CV, Britto DS, et al. Complete genome sequences of three newly discovered cacao mild mosaic virus isolates from *Theobroma cacao* L. in Brazil and Puerto Rico and evidence for recombination. Arch Virol. 2021;166:2027–31. https://doi.org/10.1007/s00705-021-05063-5.
57. Rao GP, Sharma SK, Singh D, Arya M, Singh P, Baranwal VK. Genetically diverse variants of *Sugarcane bacilliform virus* infecting sugarcane in India and evidence of a novel recombinant *badnavirus* variant. J Phytopathol. 2014;162:779–87. https://doi.org/10.1111/jph.12270.
58. Rasera JB, Amorim L, Marques JPR, Soares MKM, Appezzato-da-Glória B. Histopathological evidences of early grapevine leaf senescence caused by *Phakopsora euvitis* colonisation. Physiol Mol Plant Pathol. 2019;108:101434. https://doi.org/10.1016/j.pmpp.2019.101434.
59. Reyes BMD, Fonseca PLC, Heming NM, Conceição LBDA, Nascimento KTDS, Gramacho KP, et al. Characterization of the microbiota dynamics associated with *Moniliophthora roreri*, causal agent of cocoa frosty pod rot disease, reveals new viral species. Front Microbiol. 2023;13:1053562. https://doi.org/10.3389/fmicb.2022.1053562.
60. Ristaino JB, Anderson PK, Bebber DP, Brauman KA, Cunniffe NJ, Fedoroff NV, et al. The persistent threat of emerging plant disease pandemics to global food security. Proc Natl Acad Sci U S A. 2021;118:e2022239118. https://doi.org/10.1073/pnas.2022239118.
61. Ritsch M, Brait N, Harvey E, Marz M, Lequime S. Endogenous viral elements: insights into data availability and accessibility. Virus Evol. 2024;10:veae099. https://doi.org/10.1093/ve/veae099.
62. Roumagnac P, Lett J-M, Fiallo-Olivé E, Navas-Castillo J, Zerbini FM, Martin DP, et al. Establishment of five new genera in the family Geminiviridae: Citlodavirus, Maldovirus, Mulcrilevirus, Opunvirus, and Topilevirus. Arch Virol. 2022;167:695–710. https://doi.org/10.1007/s00705-021-05309-2.
63. Rubio L, Galipienso L, Ferriol I. Detection of plant viruses and disease management: relevance of genetic diversity and evolution. Front Plant Sci. 2020;11:1092. https://doi.org/10.3389/fpls.2020.01092.
64. Rusconi M, Conti A. *Theobroma cacao* L., the food of the gods: a scientific approach beyond myths and claims. Pharmacol Res. 2010;61:5–13. https://doi.org/10.1016/j.phrs.2009.08.008.
65. Sakai WS. Simple method for differential staining of paraffin embedded plant material using toluidine blue o. Stain Technol. 1973;48:247–9. https://doi.org/10.3109/10520297309116632.
66. Sánchez-Tovar MR, Rivera-Bustamante RF, Saavedra-Trejo DL, Guevara-González RG, Torres-Pacheco I. Mixed plant viral infections: complementation, interference and their effects, a review. Agronomy. 2025;15:620. https://doi.org/10.3390/agronomy15030620.
67. Shakir S, Mubin M, Nahid N, Serfraz S, Qureshi MA, Lee T-K, et al. Repercussions: how geminiviruses recruit host factors for replication. Front Microbiol. 2023;14:1224221. https://doi.org/10.3389/fmicb.2023.1224221.
68. Souza PFN, Garcia-Ruiz H, Carvalho FEL. What proteomics can reveal about plant–virus interactions? Photosynthesis-related proteins on the spotlight. Theor Exp Plant Physiol. 2019;31:227–48. https://doi.org/10.1007/s40626-019-00142-0.
69. Srivastava A, Agrawal L, Raj R, Jaidi M, Raj SK, Gupta S, et al. *Ageratum enation virus* infection induces programmed cell death and alters metabolite biosynthesis in *Papaver somniferum*. Front Plant Sci. 2017;8:1172. https://doi.org/10.3389/fpls.2017.01172.
70. Tatineni S, Alexander J, Qu F. Differential synergistic interactions among four different wheat-infecting viruses. Front Microbiol. 2022. https://doi.org/10.3389/fmicb.2021.800318.
71. Teixeira PJPL, Thomazella DPDT, Reis O, Do Prado PFV, Do Rio MCS, Fiorin GL, et al. High-resolution transcript profiling of the atypical biotrophic interaction between *Theobroma cacao* and the fungal pathogen *Moniliophthora perniciosa*. Plant Cell. 2014;26:4245–69. https://doi.org/10.1105/tpc.114.130807.
72. Teycheney P-Y, Geering ADW, Dasgupta I, Hull R, Kreuze JF, Lockhart B, et al. ICTV virus taxonomy profile: Caulimoviridae. J Gen Virol. 2020;101:1025–6. https://doi.org/10.1099/jgv.0.002114.
73. Ullah I, Dunwell JM. Bioinformatic, genetic and molecular analysis of several badnavirus sequences integrated in the genomes of diverse cocoa (*Theobroma cacao* L.) germplasm. Saudi J Biol Sci. 2023;30:103648. https://doi.org/10.1016/j.sjbs.2023.103648.
74. Untergasser A, Cutcutache I, Koressaar T, Ye J, Faircloth BC, Remm M, et al. Primer3–new capabilities and interfaces. Nucleic Acids Res. 2012;40:e115. https://doi.org/10.1093/nar/gks596.
75. Varsani A, Roumagnac P, Fuchs M, Navas-Castillo J, Moriones E, Idris A, et al. *Capulavirus* and *Grablovirus*: two new genera in the family *Geminiviridae*. Arch Virol. 2017;162:1819–31. https://doi.org/10.1007/s00705-017-3268-6.
76. Vaucheret H, Voinnet O. The plant siRNA landscape. Plant Cell. 2024;36:246–75. https://doi.org/10.1093/plcell/koad253.
77. Voinnet O. Fly antiviral RNA silencing and miRNA biogenesis claim ARS2. Cell Host Microbe. 2009;6:99–101. https://doi.org/10.1016/j.chom.2009.08.002.

78. Wang A, Zhou X eds. Current research topics in plant virology. 1st ed. 2016. Cham: Springer International Publishing : Imprint: Springer. 2016. https://doi.org/10.1007/978-3-319-32919-2.
79. Wang J, Chitsaz F, Derbyshire MK, Gonzales NR, Gwadz M, Lu S, et al. The conserved domain database in 2023. Nucleic Acids Res. 2023;51(D1):D384–8. https://doi.org/10.1093/nar/gkac1096.
80. Wang L, Fan P, Jimenez-Gongora T, Zhang D, Ding X, Medina-Puche L, et al. The V2 protein from the Geminivirus Tomato Yellow Leaf Curl Virus largely associates to the endoplasmic reticulum and promotes the accumulation of the viral C4 protein in a silencing suppression-independent manner. Viruses. 2022;14:2804. https://doi.org/10.3390/v14122804.
81. Ye X, Ding D, Chen Y, Liu C, Li Z, Lou B, et al. Identification of RNA silencing suppressor encoded by citrus chlorotic dwarf-associated virus. Front Microbiol. 2024;15:1328289. https://doi.org/10.3389/fmicb.2024.1328289.
82. Zhang J, Ma M, Liu Y, Ismayil A. Plant defense and viral counter-defense during plant-geminivirus interactions. Viruses. 2023;15:510. https://doi.org/10.3390/v15020510.

Publisher's Note