

Article

Theobroma cacao Virome: Exploring Public RNA-Seq Data for Viral Discovery and Surveillance

Gabriel Victor Pina Rodrigues ¹, João Pedro Nunes Santos ¹, Lucas Yago Melo Ferreira ¹,
Lucas Barbosa de Amorim Conceição ¹, Joel Augusto Moura Porto ¹ and Eric Roberto Guimarães Rocha Aguiar ^{2,*}

¹ Department of Biological Science, Center of Biotechnology and Genetics, Universidade Estadual de Santa Cruz, Ilhéus 45662-900, BA, Brazil; gvprodrigues.ppggbm@uesc.br (G.V.P.R.); jpnsantos.bio@uesc.br (J.P.N.S.); lucasmelobiomed@gmail.com (L.Y.M.F.); lucasbarbosa1714@gmail.com (L.B.d.A.C.); joelaugustomp@gmail.com (J.A.M.P.)

² Department of Engineering and Computing, Universidade Estadual de Santa Cruz, Ilhéus 45662-900, BA, Brazil

* Correspondence: ericgdp@gmail.com

Abstract: Cocoa (*Theobroma cacao* L.) is a major agricultural commodity, essential for the global chocolate industry and the livelihoods of millions of farmers. However, viral diseases pose a significant threat to cocoa production, with Badnavirus species causing severe losses in Africa. Despite its economic importance, the overall virome of *T. cacao* remains poorly characterized, limiting our understanding of viral diversity and potential disease interactions. This study aims to assess the cocoa-associated virome by analyzing 109 publicly available RNA-seq libraries from nine BioProjects, covering diverse conditions and geographic regions. We implemented a comprehensive bioinformatics pipeline integrating multiple viral sequence enrichment steps, a hybrid assembly strategy using different assemblers, and sequence similarity searches against NCBI non-redundant databases. Our approach identified ten putative novel viruses associated with the cocoa microbiome and a novel Badnavirus species. These findings provide new insights into the viral landscape of *T. cacao*, characterizing the diversity of cacao-associated viruses and their potential ecological roles. Expanding the catalog of viruses associated with cocoa plants not only enhances our understanding of plant–virus–microbiome interactions but also contributes to the development of more effective disease surveillance and management strategies, ultimately supporting sustainable cocoa production.

Keywords: cocoa; Badnavirus; metatranscriptome; metavirome; mycoviruses



Academic Editor: Svetlana Y. Folimonova

Received: 16 March 2025

Revised: 7 April 2025

Accepted: 11 April 2025

Published: 26 April 2025

Citation: Rodrigues, G.V.P.; Santos, J.P.N.; Ferreira, L.Y.M.; Conceição, L.B.d.A.; Porto, J.A.M.; Aguiar, E.R.G.R. *Theobroma cacao* Virome: Exploring Public RNA-Seq Data for Viral Discovery and Surveillance. *Viruses* **2025**, *17*, 624. <https://doi.org/10.3390/v17050624>

Copyright: © 2025 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

Cocoa (*Theobroma cacao* L.) is a tropical plant classified under the Malvaceae family and the *Theobroma* genus [1]. *Theobroma* comprises 22 species, but *Theobroma cacao* L. stands out as the most significant and widely cultivated, primarily for its highly valued seeds [2]. *T. cacao* is not just the backbone of the chocolate industry but also a vital agricultural commodity that sustains millions of livelihoods worldwide [1,3]. Each year, approximately 4.5 million tons of cocoa are produced, generating over USD 10 billion in income for farming families [4]. For many smallholder farmers, particularly in West Africa, cocoa is more than a cash crop; it is the foundation of their economic survival. Countries like Côte d'Ivoire and Ghana lead global production [5,6], making cocoa a critical driver of employment, foreign exchange earnings, and economic stability in the region [7,8].

Traditionally known for its role in chocolate production, cocoa is being explored for various other industrial applications. By-products like cocoa shells and pod husks are

increasingly used in animal feed, cosmetics, and even biofuels, adding both economic and environmental value to the crop [9,10]. As global demand for cocoa-based products continues to grow, particularly for chocolate, its economic significance remains as strong as ever [11].

Nevertheless, the industry faces mounting challenges. Climate change, market fluctuations, and sustainability concerns threaten cocoa production and pricing, putting pressure on the millions who depend on it [8,12]. In this scenario, one of the most destructive pathogens affecting *T. cacao* is *Moniliophthora perniciosa*, a hemibiotrophic fungus responsible for witches' broom disease, which significantly disrupts cocoa production [13–16]. Another major threat is *Moniliophthora roreri*, which specifically infects developing cocoa pods, leading to frosty pod rot and further compromising yields [17].

Apart from these fungal infections, virus-related diseases represent an additional major threat to *T. cacao* and global agriculture, with far-reaching consequences for crop yield, quality, and plant health [18,19]. These viruses hijack plant cellular processes to replicate and spread, often resulting in symptoms such as stunted growth, necrosis, leaf and pod deformation, and discoloration [20]. For cocoa specifically, viral infections can directly reduce bean production and quality, disrupting supply chains and threatening economic stability [21–23]. This is particularly concerning in Africa, where the *Cacao swollen shoot virus* (CSSV) has devastated vulnerable cocoa varieties, particularly in West Africa [24–26]. Member of the Badnavirus genus within the *Caulimoviridae* family, this virus was first identified in Ghana in 1936 and has since spread to other cocoa-producing countries in West Africa, including Côte d'Ivoire, Nigeria, Togo, and Benin [27–29]. It is primarily transmitted by mealybugs, which facilitate its spread among cocoa trees [30]. Infected trees exhibit severe symptoms, including shoot swelling, yellowing leaves, stunted growth, and, in many cases, premature tree death [31,32].

Several viral infections have been documented in cocoa production, predominantly involving members of the Badnavirus genus, such as *Cacao mild mosaic virus* (CaMMV), *Cacao yellow vein banding virus* (CYVBV) [33], *Cacao red vein virus* (CRVV) [34], and many different isolates of CSSV [28,35]. However, other viral families have also been reported infecting *T. cacao*, including the *Solemoviridae* family, represented by *Cacao leafroll virus* from the Polerovirus genus, and the *Tymoviridae* family, which includes *Cacao yellow mosaic virus* (CYMV) [36]. These other species may cause several losses to cocoa production; CaMMV alone has been responsible for lower annual yield ranging from approximately 7% to 33% in infected plants [37].

Cocoa viral infections are widespread and exhibit substantial genetic diversity across different regions, posing significant challenges for disease control [35,38]. In Ghana, CSSV management has traditionally relied on practices such as removing infected trees and replanting to eliminate viral reservoirs [39]. However, no effective treatments are currently available for viral infections in cocoa trees, underscoring the urgency of improved disease management strategies. Addressing these threats is essential not only for agricultural sustainability but also for preserving economic stability and international trade networks [40].

This study aims to analyze metatranscriptomic data from publicly available *Theobroma cacao* RNA-seq libraries across various regions worldwide, providing information regarding the circulation of known pathogenic viruses and the detection of new emergent species.

2. Materials and Methods

2.1. Acquisition of RNA-Seq Libraries

The RNA library accessions were retrieved from the Sequence Read Archive (SRA) database available on the NCBI platform. The search was conducted using the keywords

“Theobroma” AND “cacao”. Only RNA-seq libraries of *T. cacao* sequenced using the ILLUMINA® platform were selected, encompassing both *single-end* and *paired-end* layouts. Libraries corresponding to technical replicates were excluded from the selection process. The BioProjects associated with the selected libraries were PRJEB35419, PRJNA604260, PRJNA613342, PRJNA742476, PRJNA785999, PRJNA558793, PRJNA314774, PRJNA421343, and PRJNA326055 (Supplementary Table S1).

2.2. RNA-Seq Processing and Assembly

Publicly available RNA-seq libraries were preprocessed with Fastp (version 0.23.2) [41] to eliminate adapter sequences and low-quality reads. Host-derived reads were filtered out by aligning the processed reads to the respective reference genome, Criollo cocoa B97-61/B2 (GCF_000208745.1) [42], using the STAR aligner (version 2.7.11b) [43] with default parameters. Unmapped reads were subsequently used for de novo assembly, employing two distinct tools: SPAdes (version 4.0.0) [44] using default settings and the transcripts derived by SPAdes were further processed using CAP3 (VersionDate: 2 October 2015) [45].

2.3. RNA-Seq-Based Assessment of the Cocoa Microbiome

The unmapped reads were analyzed for taxonomic classification by comparing them with the reference database PlusPFP (version: 7 June 2022) on Kraken2 (Galaxy Version 2.1.3+galaxy1) [46]. Kraken-report (Galaxy Version 1.3.1) was further applied to generate a report file using the Minikraken database from which we extracted and refined our results.

2.4. Metaviromic Analysis

Viral contigs were identified through sequence similarity searches using Diamond (version 2.1.8) in blastX mode [47] against a local nr database (release 08/2024) and BLAST (version 2.12.0+) [48] alignment with a local nt database (release 08/2024). The identified assembled contigs were filtered based on a minimum threshold of 500 nucleotides. The largest open reading frames (ORFs) within each contig were predicted using ORFiPy (version 0.0.4) [49], and the resulting sequences were further analyzed with HMMER (version 3.4) [50], using the *hmmsearch* application, against the Reference Viral Database (RVDB) HMM profile [51], version 29.0, and conserved domains were predicted using Protein Family Database (Pfam) [52] release 37.2.

2.5. Novel Virus Definition

Virus classification was based on the latest ICTV criteria, along with phylogenetic placement, genome organization, and pairwise protein identity with defined viruses. Specifically, viral species demarcation followed ICTV criteria, whereas viruses belonging to families or genera without established demarcation criteria were classified based on a nucleotide identity threshold of 90%. Similarly, genus and family classification adhered to the latest ICTV guidelines.

2.6. Integrative Genome Assembly

In order to improve possible viral genomes, an integrative assembly strategy was also performed by running the following different tools: rnaviralSPAdes (version 4.0.0) [44], Megahit (version 1.2.9) [53], and Trinity (version 2.15.2) [54], all assemblers with default parameters. The results were consolidated using CD-HIT (version 4.8.1) [55] to join sequences with more than 90% similarity. The assembled sequences were manually investigated and used along with raw read alignment to correct misassembled regions. Manually curated viral sequences identified in our work can be found in the Supplementary Table S2.

2.7. Molecular Phylogeny of Virus-Derived Sequences

Sequences exhibiting similarity to genes encoding RNA polymerase and polyproteins were used to construct phylogenetic trees. Additionally, protein sequences representing specific phyla, including Kitrinoviricota, Pisuviricota, Larnaviricota, and Duplornaviricota, as well as the *Caulimoviridae* viral family, were selected based on their proximity and references from the International Committee on Taxonomy of Viruses (ICTV). For each set of sequences, a global alignment was performed using MAFFT v7.505 [56] to ensure accurate alignment. The alignment was visualized and trimmed with AliView [57] version 1.29. Phylogenetic trees were then inferred using IQ-TREE [58] version 2.1.4 on a local server, employing the maximum likelihood method with 1,000 bootstrap replicates to evaluate the robustness of the tree topology. The trees were visualized in the online tool Interactive Tree of Life (ITOL v6) [59]. All novel viral sequences are available in Supplementary Table S2.

2.8. RNA Abundance and the Widespread Presence of Viral Segments

The transcriptional activity of virus-derived sequences was evaluated using Salmon (version 1.10.3) [60]. To compare viral abundance, the host NADH(1-4) and COX2 genes were selected as endogenous and standard reference genes, respectively. A summary of the transcriptional activity of the viral sequences included in this analysis is provided in Supplementary Table S3.

3. Results

3.1. *Theobroma cacao* Microbiome Assessment

A total of 5,192,778,743 reads were obtained from all 109 *Theobroma cacao* RNA-seq libraries. Removal of host-derived reads led to 319,168,768 reads that were further classified using the microbiome composition-focused tool. The microbial diversity across all samples comprised 4577 species from 1428 genera, representing 452 distinct families. Elements from the phyla Proteobacteria and Ascomycota were the most abundantly represented and frequently detected in the cocoa RNA-seq libraries, followed by Actinobacteria and Firmicutes (Figure 1).

The geographic origin of each RNA-seq library spans a diverse range of regions worldwide. Among the libraries, 18 lack associated geographical metadata (from BioProjects PRJEB35419 and PRJNA314774), while the remaining 91 are derived from 10 countries: China, India, Colombia, the United States, Ecuador, Trinidad and Tobago, French Guiana, Brazil, Samoa, and Peru (Supplementary Table S1).

A subset of eight libraries, identified as wild samples based on NCBI SRA metadata (SRR3217278, SRR3217279, SRR3217280, SRR3217292, SRR3217281, SRR3217282, SRR3217283, and SRR3217284), originates from Ecuador, Colombia, and French Guiana. Analysis of species relative abundance revealed no distinct microbiome distribution patterns between these wild samples and those derived from greenhouse conditions (Supplementary Figure S1).

To further investigate geographic variations in species relative abundance, we compared samples based on their continent of origin (the Americas, Asia, and Oceania). A *t*-distributed stochastic neighbor embedding (*t*-SNE) analysis was performed to visualize the high-dimensional data structure. Samples from North and South America showed no clear separation, whereas libraries from Asia formed a distinct cluster. Additionally, libraries with unknown geographic metadata are grouped into a separate cluster, primarily composed of samples from the PRJEB35419 BioProject (Supplementary Figure S1).

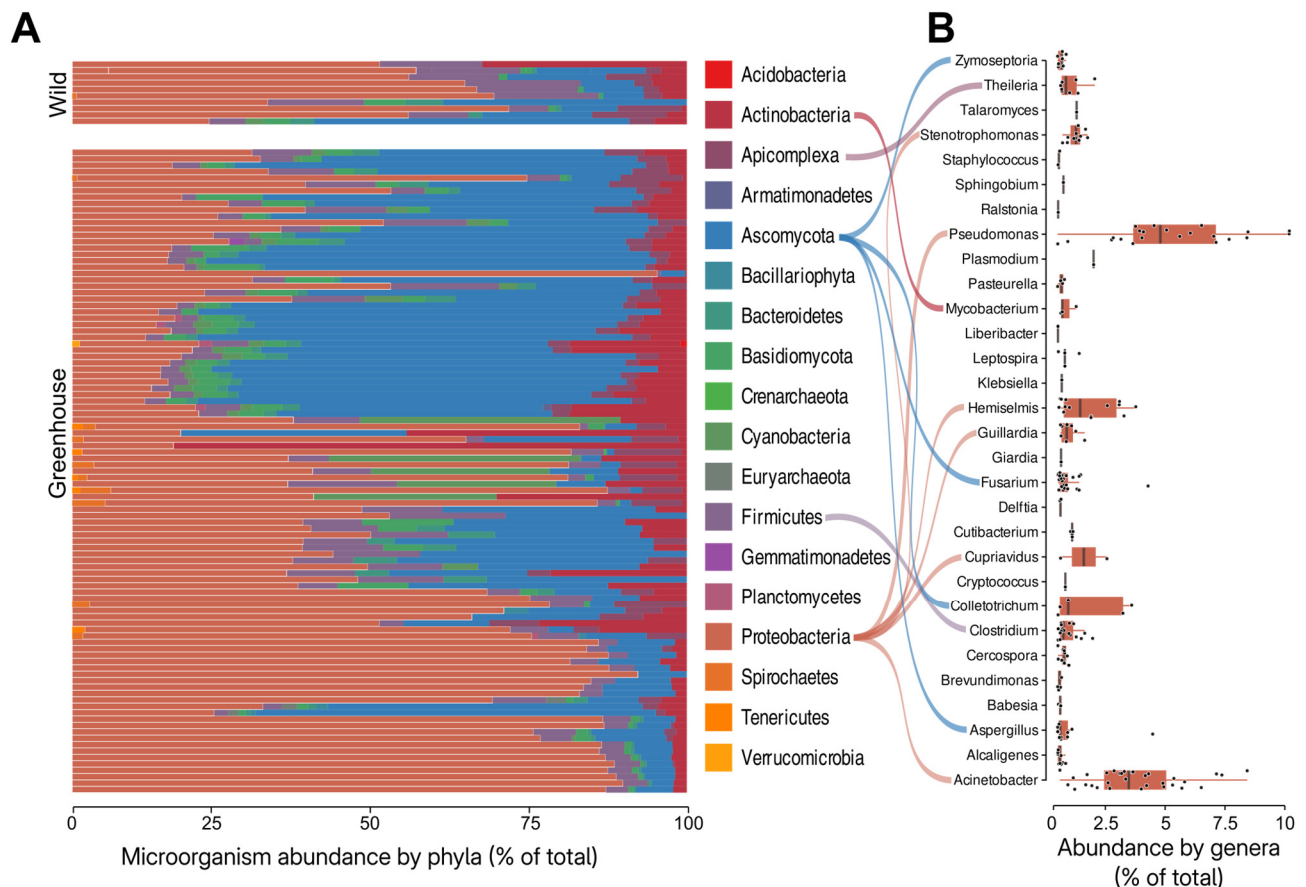


Figure 1. (A). Microbiome composition across the 109 *Theobroma cacao*-derived RNA-seq libraries, showing the relative abundance of the most frequent phyla. Phyla are represented only if they account for $\geq 0.05\%$ of the total reads in each library. The X-axis indicates the relative number of reads classified within each phylum. (B). Reads abundance (%) in total metatranscriptome of the most frequent genera across all samples, considering only genera that account for $\geq 0.5\%$ of the total reads in each library.

3.2. *Theobroma cacao*-Associated Virome

We identified a considerable number of viral sequences in the metatranscriptomic analysis, prompting a deeper investigation into the presence of viruses in *Theobroma cacao* samples. After manual curation of de novo assembled transcripts, 34 non-redundant putative viral sequences longer than 500 nt were selected for further characterization. Sequence similarity searches against the NCBI non-redundant databases revealed 14 transcripts closely related to previously described viruses, while the remaining 20 sequences likely represent new viral species.

Analysis of the virus-related sequences revealed a diverse range of species spanning different Baltimore classes, including positive-sense single-stranded RNA viruses (ssRNA+), double-stranded RNA viruses (dsRNA), and double-stranded DNA viruses (dsDNA) (Figure 2 and Supplementary Table S2). Among the putative novel viral sequences, eighteen exhibited conserved domains such as polymerase, helicase, and capsid. In contrast, the remaining sequences did not show conserved motifs, likely due to their shorter length or lack of a significant number of related sequences (Supplementary Figure S2).

Global alignment analysis

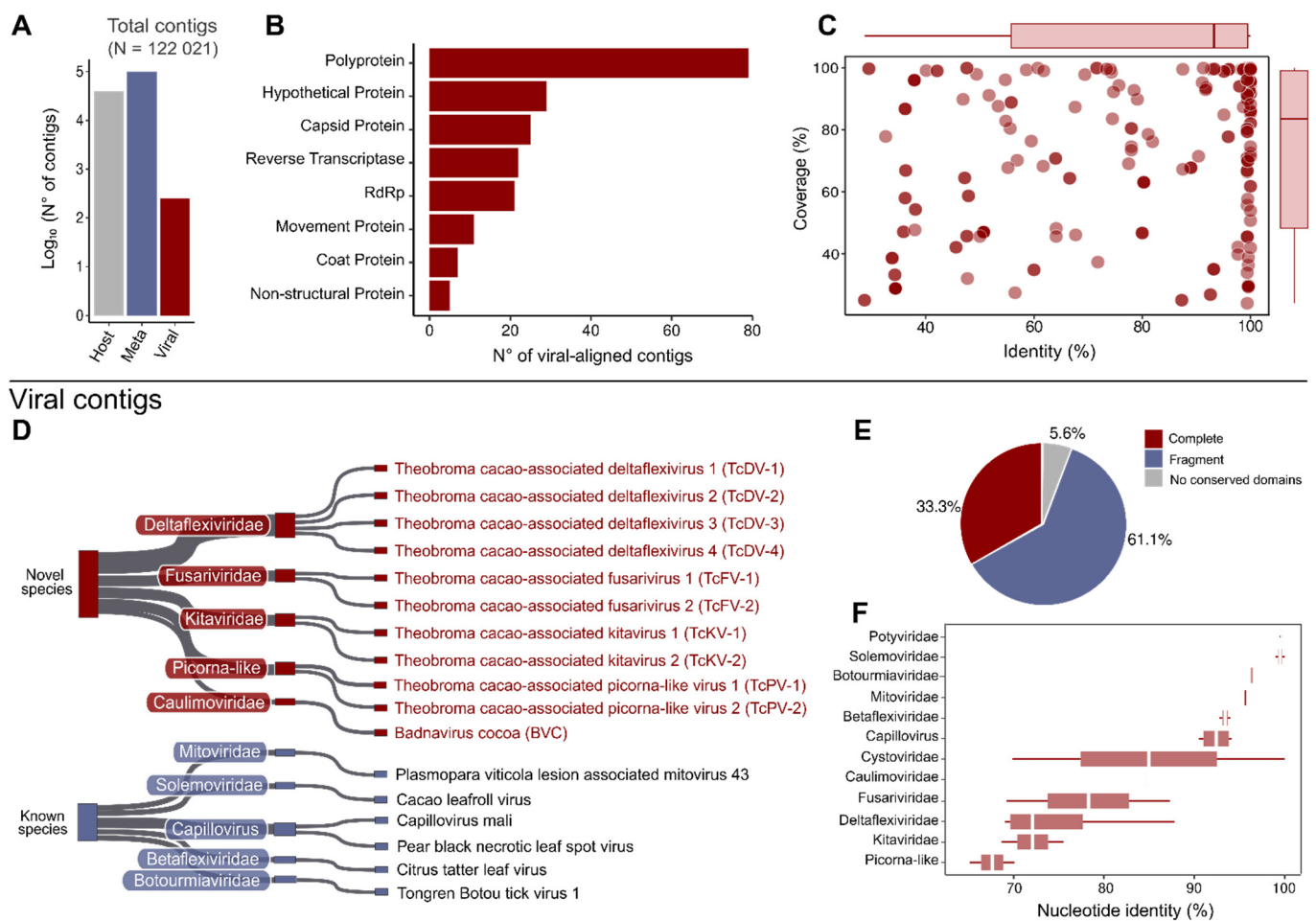


Figure 2. Metavirome overview. (A) Total contigs (N) from the de novo assembly of 109 cocoa RNA-seq libraries, classified as Meta, Host, or Viral based on protein alignment. Host material refers to residual reads that did not map to the reference genome. (B) Alignment targets of viral contigs based on protein alignment. (C) Identity and coverage of contigs aligned to viral protein sequences. (D) Classification of known and novel viruses by family and genetic material, as characterized in the contigs. (E) Distribution of complete, fragmented, and absent viral conserved domains in the characterized sequences. (F) Nucleotide similarity of contigs within each characterized viral family.

3.2.1. Characterization of Known Viral Species

Fourteen contigs exhibited high nucleotide sequence identity to the genomes of known viruses and encoded proteins with the same conserved domains as their closest hits. Among these, five contigs ranging from 530 to 1022 nt were related to three distinct strains of the Betaflexivirus *Capillovirus mali*, containing open reading frames (ORFs) for the polymerase, movement protein, and capsid. Notably, two complete ORFs of *Cacao leafroll virus*, a polerovirus, were assembled, encoding the RNA-dependent RNA polymerase (RdRP), Lutovirus P0 protein, and a peptidase. The remaining assembled contigs exhibited similarity to distinct viral sequences. One contig was related to the Red-mite-associated cystovirus, containing ORFs encoding an RdRP (632 nt) and a hydrolase (2353 nt). Additionally, two contigs were related to mitovirus sequences, with one matching the RdRP of *Plasmopara viticola* lesion-associated mitovirus 43 (1783 nt) and the other aligning to the RdRP of *Tongren Botou tick virus 1* (1254 nt) (Supplementary Table S2).

The last contig of 9498 nt showed high identity to the *Potato virus Y* (PVY). It contained a complete ORF of 8214 nt, encoding a polyprotein with conserved domains associated with

the polymerase, helicase, coat protein, and peptidase. However, all libraries containing PVY sequences originate from the BioProject PRJEB35419, which includes a total of 1818 high-similarity contigs matching *Solanum tuberosum* and an additional 1159 contigs associated with the *Solanum* genus. The substantial presence of *Solanum*-derived contigs suggests a potential contamination in these samples, and Supplementary Figure S3 exhibits more details of this pattern in all libraries that had contigs aligned with *S. tuberosum* sequences. The source of *S. tuberosum* remains unknown, such as the origin of the Potyvirus found, and further investigations are necessary.

3.2.2. Characterization of Novel Viral Species

Overall, 20 contigs showed sequence similarity to known viral species only at the amino acid level. Among these putative viral sequences, 11 contained conserved domains encoding for the polymerase/polyprotein, while the remaining contigs either displayed other domains or lacked recognizable domains and were therefore not subjected to phylogenetic analysis. Hence, six transcripts related to replicative sequences were associated with viruses within the Kitrinoviricota phylum and four with viruses from the Pisuviricota. Additionally, one contig exhibited similarity to dsDNA caulimoviruses, including *Cacao yellow vein banding virus*.

Kitrinoviricota

The *Deltaflexiviridae* family is a group of fungi- and plant-infecting viruses characterized by an ssRNA+ polyadenylated genome, typically ranging from 6 to 8.3 kb in size and containing four to five open reading frames. The genome encodes essential proteins such as the RNA-dependent RNA polymerase, viral helicase, and methyltransferase [61,62]. In our study, four putative new deltaflexivirus sequences were assembled, with their genomes ranging in size from 527 to 4432 nt.

The largest contig, measuring 4432 nt, although differing in genome size from previously established *Deltaflexiviridae* members, was assembled with a complete genome structure. It features a complete polyprotein ORF of 2199 nt, which encodes the polymerase and helicase domains, as well as three smaller ORFs that lack conserved domains, which is expected for this family. In contrast, the 3741-nt-long contig only contained the polyprotein ORF (3273 nt), which displayed the polymerase and helicase conserved domains. The two smaller contigs each contained an incomplete polyprotein ORF, ranging from 468 to 513 nt, yet both displayed the polymerase domain as their best hits (Supplementary Figure S2 and Supplementary Table S2).

Of note, the phylogenetic analysis clustered the two largest contigs with characterized viruses from the genus *Deltaflexivirus*, while the two smaller ones were grouped with unclassified members of the *Deltaflexiviridae* family, potentially due to their shorter genome lengths (Figure 3). The sequences were named *Theobroma cacao*-associated deltaflexivirus 1-4 (TcDV-1-4).

The *Kitaviridae* family comprises plant-infecting viruses with ssRNA+ genomes. Their genomes are multipartite, consisting of two to four segments depending on the viral genus, with at least one large open reading frame typically ranging from 5 to 8 kb. In addition to the replication-associated protein, their genomes encode movement proteins and structural proteins [63].

Two assembled transcripts exhibited amino-acid-level similarity to *Kitaviridae* members. The first, measuring 623 nt, contained an incomplete ORF of 588 nt, while the second, at 3666 nt, harbored a complete ORF of 3378 nt. Notably, both included conserved domains associated with the polymerase, whereas their top BLAST hits lacked identifiable conserved domains (Supplementary Figure S2 and Supplementary Table S2). Phylogenetic analysis using the RdRp ORFs, consistent with the similarity analysis, indicated that the assembled

contigs belong to the *Kitaviridae* family, as they clustered with unclassified members of this family. Consequently, these sequences were named *Theobroma cacao*-associated kitavirus 1 and 2 (TcKV 1-2) (Figure 3).

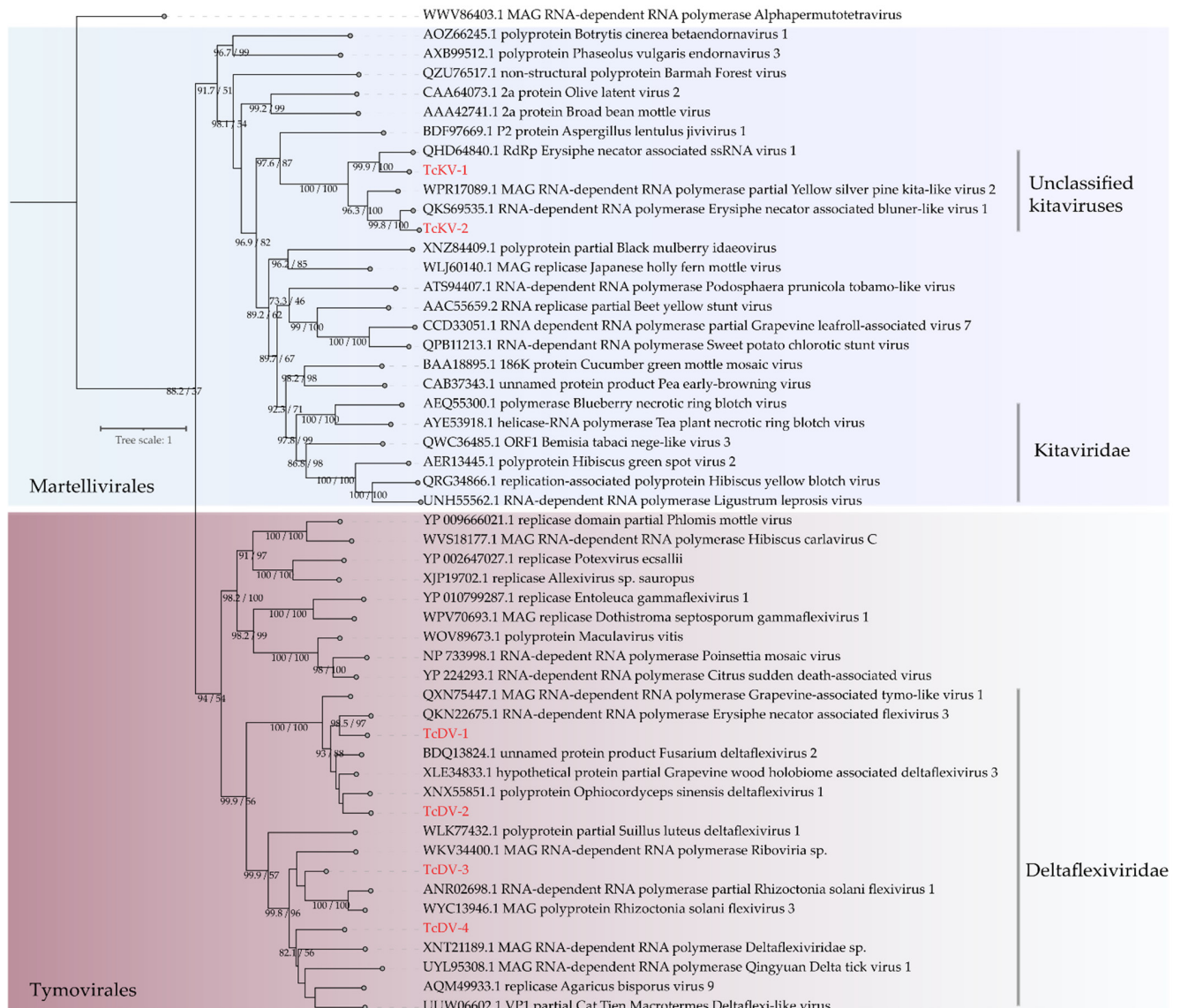


Figure 3. Phylogenetic analysis of Kitrinoviricota viruses. A maximum likelihood phylogenetic tree was constructed based on full-length amino acid sequences of the RNA-dependent RNA polymerase (RdRp). The tree was generated using IQ-TREE, employing the Blossum62+F+R5 evolutionary model with 1000 bootstrap pseudoreplicates for statistical support. The tree was rooted using the RdRp sequence from Alphapermutetetravirus (family *Alphapermutetetraviridae*) as the outgroup for comparison. The putative new viral species are highlighted in red.

Pisuviricota

The *Fusariviridae* family consists of positive-sense, single-stranded RNA viruses with genome sizes varying between 5.9 and 10.7 kb. Their genomes are typically bicistronic, with the larger ORF encoding an RNA-dependent RNA polymerase (RdRp) and an RNA helicase, while the function of the second ORF remains unclear. These viruses are primarily found in fungal hosts, though some evidence suggests that they may also infect oomycetes [64]. In our study, two sequences exhibited similarity to fusaviruses. The first, measuring 6556 nt, contained two complete ORFs of 4758 nt and 1296 nt. The larger ORF encoded the RdRp and a viral helicase, while the second ORF lacked conserved domains, which is characteristic of

Fusariviridae members (Supplementary Figure S2). The second contig of 1008 nt featured an incomplete ORF of 963 nt that retained the polymerase domain. Based on phylogenetic analysis, these sequences were designated as *Theobroma cacao*-associated fusarivirus 1-2 (TcFV1-2) (Figure 4).

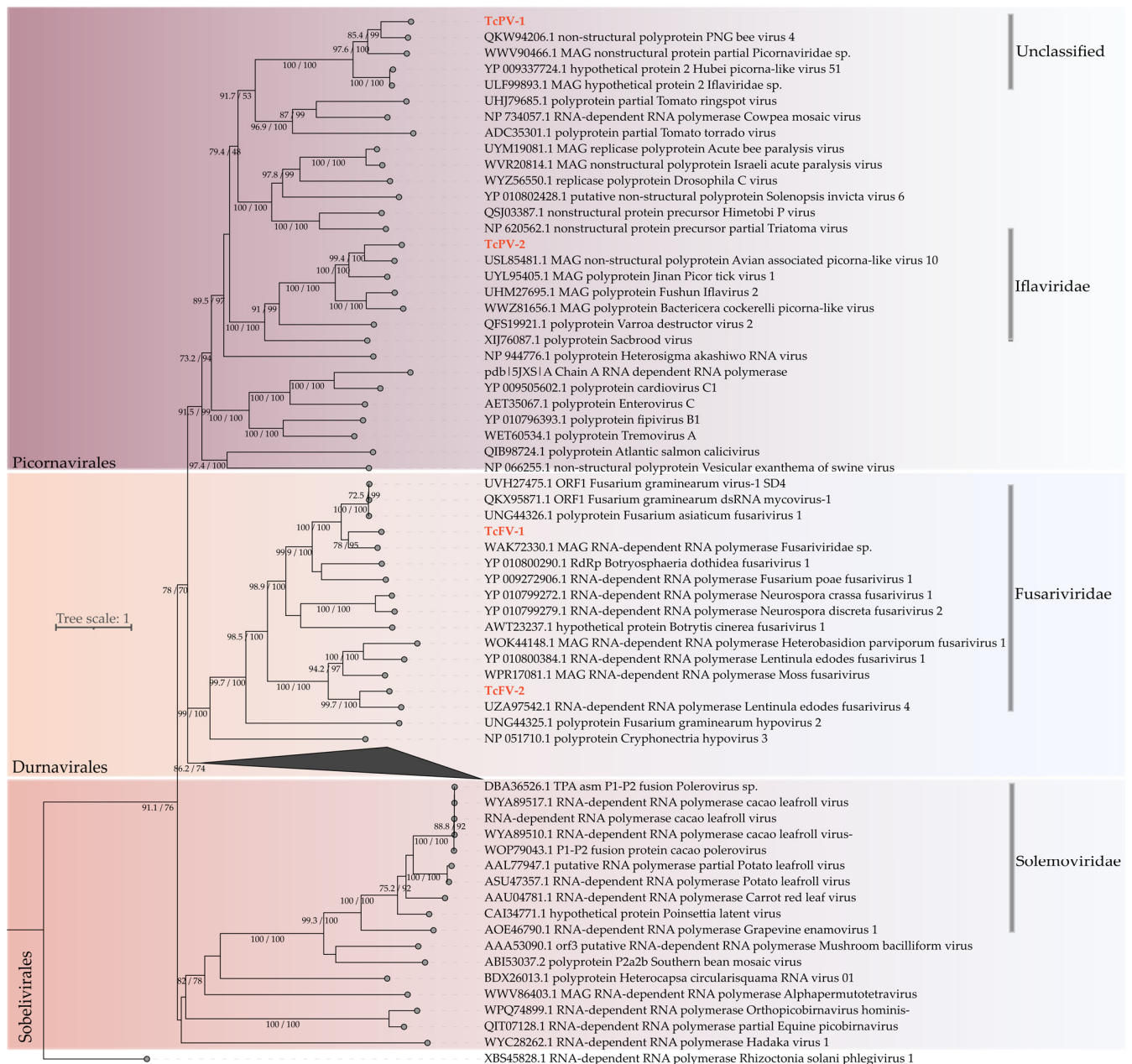


Figure 4. Phylogenetic analysis of Pisuviricota viruses. A maximum likelihood phylogenetic tree was constructed based on full-length amino acid sequences of the RNA-dependent RNA polymerase (RdRp). The tree was generated using IQ-TREE, employing the Blossum62+F+R5 evolutionary model with 1000 bootstrap pseudoreplicates for statistical support. The tree was rooted using the RdRp sequence from *Rhizoctonia solani* phlegivirus 1 (family *Phlegiviridae*) as the outgroup for comparison. The putative new viral species are highlighted in red.

The order Picornavirales comprises a diverse group of ssRNA+ viruses that infect a wide range of hosts, including plants, invertebrates, and vertebrates. Their genomes typically range from 6 to 12 kb and are characterized by a single large open reading frame encoding a polyprotein with RdRp, helicase, and protease-related domains. This order includes several well-known families, such as *Picornaviridae*, *Secoviridae*, and *Dicistroviri-*

dae [65]. In this study, two assembled transcripts were identified as related to viruses from the Picornavirales order, with lengths of 1331 nt and 1520 nt. Both contigs contained complete polymerase ORFs of 936 nt and 1314 nt, respectively, and notably, they presented a Picornavirales-specific polymerase domain as their best hit (Figure 4).

In the phylogenetic analysis, the larger sequence clustered in a clade of unclassified Picornavirales members, while the second contig grouped with members of the *Iflaviridae* family. Based on these findings, the sequences were designated Theobroma cacao-associated picorna-like virus 1 and 2 (TcPV-1 and TcPV-2).

Caulimoviridae

The *Caulimoviridae* family consists of dsDNA viruses that infect plants, often causing persistent infections. Their genomes, typically ranging from 7 to 9.8 kb, are circular and pararetroviral. The genome organization includes multiple open reading frames encoding proteins such as the reverse transcriptase (RT), coat protein, movement protein, and transcriptional regulators. Members of this family are non-enveloped and form icosahedral or bacilliform virions, depending on the genus [66]. A *Caulimoviridae*-related incomplete contig was assembled, measuring 2074 nt. It contained one ORF of 1377 nt, encoding the reverse transcriptase and ribonuclease H domains, respectively.

Phylogenetic analysis revealed that this putative viral sequence clustered with members of the Badnavirus genus, notably within the same clade as the well-known cacao pathogen *Cacao yellow vein banding virus* (Figure 5). Interestingly, it was distinct from other known cacao-infecting badnaviruses, such as *Cacao swollen shoot virus* and *Cacao mild mosaic virus*. Under these observations, this sequence was designated as Badnavirus cocoa (BCV).

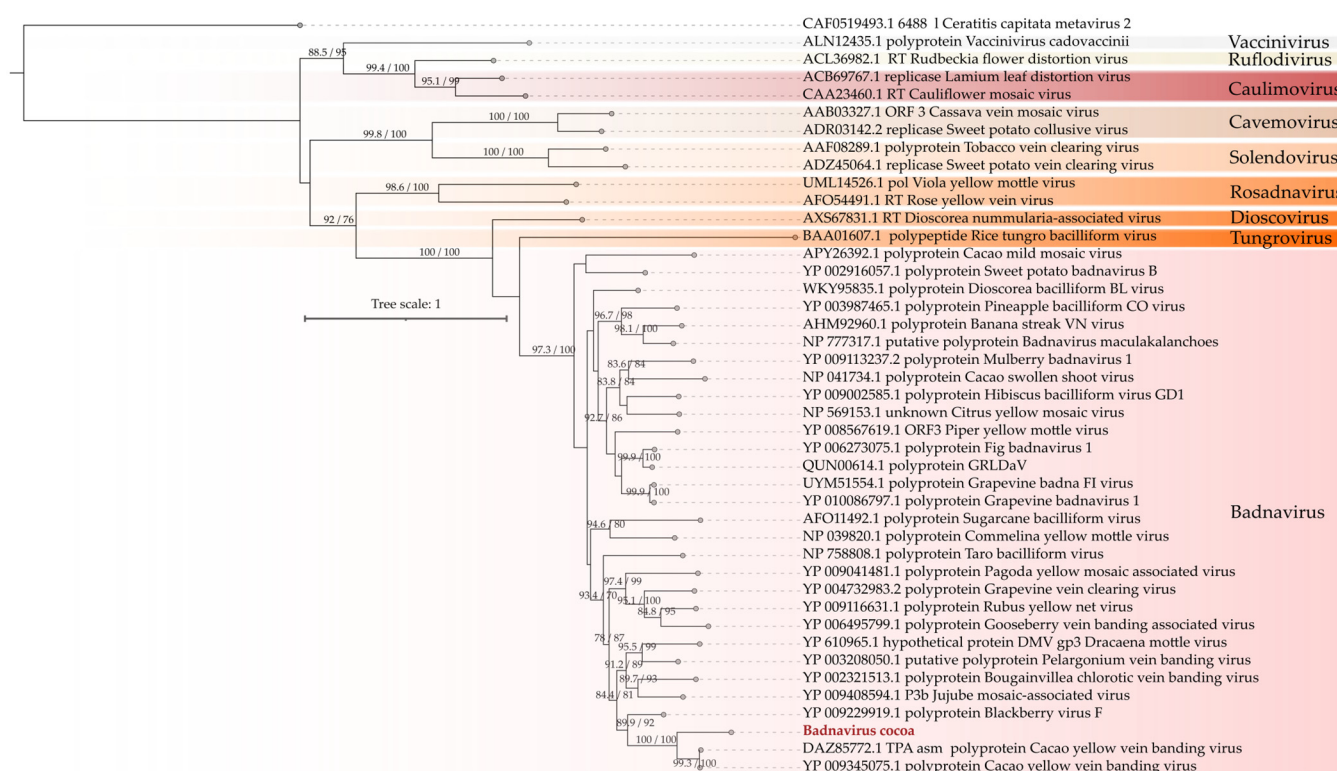


Figure 5. Phylogenetic analysis of the caulimovirus. A maximum likelihood phylogenetic tree was constructed based on full-length amino acid sequences of the reverse transcriptases (RTs). The tree was generated using IQ-TREE, employing the LG+F+R6 evolutionary model with 1000 bootstrap pseudoreplicates for statistical support. The tree was rooted using the RT from *Ceratitidis capitata metavirus 2* (family *Metaviridae*) as the outgroup for comparison. The putative new viral species is highlighted in red.

Of note, nucleotide similarity analysis showed relatively low identity between BCV and the endogenous *Theobroma cacao* viral sequences, specifically *Theobroma cacao* bacilliform virus 1, with 66.12% identity and 46% coverage.

3.3. Wide Spread of Known Pathogens and Novel *Theobroma cacao*-Associated Viruses

The *Caulimoviridae* family comprises dsDNA viruses that infect plants, often establishing persistent infections. In this context, we investigated the abundance of our viral sequences compared to known Badnavirus species. Endogenous viral elements (EVEs) from this genus have already been reported in the *Theobroma cacao* genome [67]. Notably, in our data, a sequence from *Cacao swollen shoot virus* (CSSV) was detected in 47 different libraries from distinct BioProjects through read quantification analysis (Supplementary Figure S4). However, the alignment of assembled contigs did not reveal any sequence with similarity to this viral species, likely due to its similarity to known endogenous viruses.

Other known viral sequences were also mapped to libraries where no sequences had previously been identified through contig assembly, including *Cacao leafroll virus*. A total of fifteen different libraries from distinct BioProjects were mapped to the complete genome of this virus; only four of them contained metatranscriptomic contigs aligning with the viral sequences.

The abundance of constitutive genes was higher than that of any known or novel viral sequence characterized in this study. Given the large number of processed libraries and their intrinsic specificities and the different characteristics of each viral sequence within the dataset, only a few clusters allow for the analysis of sequence quantification variability across samples (Figure 6). In general, the detection of the novel viral species was restricted to libraries associated with the same BioProject and geographic region of the library where the virus was first found. In contrast, the distribution of viral families exhibited a broader geographic distribution. Novel species of the *Fusariviridae* family were predominantly found in the Asian continent, particularly in India and China. *Deltaflexiviridae* species were detected in both the USA and Ecuador, while novel *Kitaviridae* sequences were identified in samples derived from the USA and India. However, it is important to point out that we were not able to determine a core virome for *Theobroma cacao*.

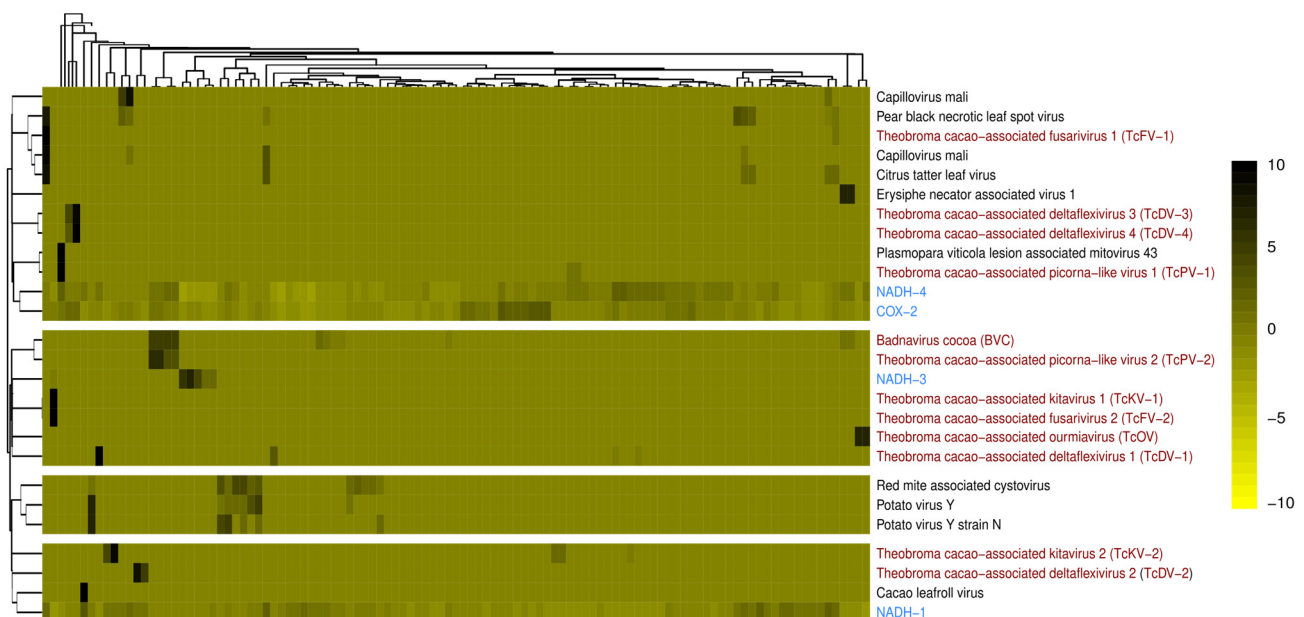


Figure 6. Transcriptional activity of *Theobroma cacao* RNA-seq libraries, highlighting both known and newly identified viral sequences. Novel viral sequences detected in *T. cacao* libraries are indicated in red. Expression profiles of reference cocoa genes across all 109 analyzed libraries are shown in blue. Known and new viral species are represented in black and red, respectively.

4. Discussion

The metatranscriptomic analysis of all libraries reveals common patterns characteristic of the relatively unexplored *Theobroma cacao* leaf microbiome [68,69]. Most studies on this subject have primarily focused on the cocoa bean microbiome during the fermentation process [70–75], and a few reported the diverse associated virome related to the cocoa microbiome [76,77]. Viral diversity and surveillance in *Theobroma cacao* remain relatively unexplored, as most studies have primarily focused on members of the Badnavirus and Polerovirus genera. This study aimed to comprehensively explore the total virome and its diversity using a public dataset of 109 RNA-seq libraries collected under different conditions, across various regions, and from multiple research projects. We were able to identify multiple viral sequences, and although in some cases they do not represent complete viral genomes, they present many characteristics of viral sequences, such as large open reading frames and the presence of conserved domains associated with viral species. However, these putative viral sequences might originate from symbiotic/pathogenic fungi or microbial communities in different plant tissues, requiring further experiments to determine the possible host.

Despite the unequal number of greenhouse-derived samples compared with wild accessions (only 7.33% of wild accessions libraries), there was no significant variation in the metatranscriptomic profile between these groups. However, due to the large difference in sample size, there is still a bias in the comparative analysis. Notably, only 11% of the samples originate from cocoa pod sequencing, specifically from BioProjects PRJNA613342 (China) and PRJNA604260 (India), derived from lab samples. These libraries exhibited novel cocoa-associated viruses and a high prevalence of fungal contigs. The first BioProject, which analyzed samples from China [78], appears to provide the first report of *Capillovirus mali* sequences in a *T. cacao* RNA-seq library. Commonly known as *Apple stem grooving virus* (ASGV) and *Citrus tatter leaf virus* (CTLV), this virus has been reported to have a global distribution. As for ASGV identification, *Capillovirus mali* has been documented infecting a broad range of hosts, including apples, pears, citrus, kiwifruit, and various other horticultural crops [79–81]. Meanwhile, CLTV has been widely recognized as a significant pathogen affecting citrus production worldwide [82]. Currently, there is no evidence based on the literature of *Capillovirus mali* infecting cocoa trees. Similarly, the other known viruses found in our analysis could be associated with a diverse host range inside the cocoa's microbial community, from symbiotic/pathogenic fungi or aphids and other insects that could have landed on cacao plants after interacting with virus-infected hosts. However, knowledge of these viruses is important to monitor possible spillover events that could occur from these interactions.

A newly characterized putative virus-like sequence from the *Fusariviridae* family, *Theobroma cacao*-associated fusarivirus 1 (TcFV-1), was identified in the same BioProject, specifically in the RNA-seq library under the SRA code SRR11389078. Although TcFV-1 is a fungal virus, the microbiome analysis of this library reveals a low abundance of fungal reads, with only 0.56% of the total metatranscriptome aligning with fungal targets. This contrasts with the host sample of TcFV-2, identified in SRR11060111, a library from India, which exhibits approximately 11.8% of total metatranscriptomic reads associated with fungal elements. The key commonality between the samples hosting TcFV-1 and TcFV-2 is the plant tissue used for library preparation—the cocoa pod.

The two infected libraries also share the same most abundant fungal genus, with *Fusarium* being the dominant fungal taxon in read alignment and quantification, followed by *Colletotrichum*. Members of *Fusarium* are well-known plant pathogens and have been previously reported to harbor a diverse range of dsRNA mycoviruses [83], particularly *Fusariviridae* infections in *Fusarium graminearum* by *Fusarium graminearum* virus 1 strain

DK21 (FgV1-DK21) [84,85]. Notably, read alignment and quantification in both libraries hosting TcFV-1 and TcFV-2 indicate the presence of *F. graminearum*, further supporting its association with these viral sequences.

The comprehensive microbiome analysis across all samples revealed a vast and diverse community of fungal and oomycete species within the endophytic environment of *Theobroma cacao*. One likely factor contributing to this high microbial diversity is the cultivation of cocoa in tropical and humid agricultural regions, which provide optimal conditions for the proliferation of fungal phytopathogens [86,87]. Despite the successful spread of these fungal pathogens in *T. cacao*, their uncontrolled proliferation does not go unnoticed. Mycoviruses, or fungal viruses, are widely distributed across diverse fungal taxa, including yeasts, mushrooms, and pathogenic fungi that infect plants, insects, and humans. These viruses are classified into multiple families based on their genomic composition, notably, the majority of mycoviruses possess RNA genomes, with those carrying +ssRNA being assigned to 11 distinct families, including *Alphaflexiviridae*, *Barnaviridae*, *Botourmiaviridae*, *Deltaflexiviridae*, *Endornaviridae*, *Gammapflexiviridae*, *Hadakaviridae*, *Hypoviridae*, *Mitoviridae*, *Narnaviridae*, and *Yadokariviridae* [88]. Within the family *Deltaflexiviridae*, the sole genus, *Deltaflexivirus*, includes viruses with both segmented and unsegmented genomes [89]. A conserved feature among all *Deltaflexivirus* members is the presence of a gene encoding an RNA replicase, which plays a crucial role in viral replication and maintenance [90].

Our analysis successfully identified four novel *Deltaflexiviridae* members in *Theobroma cacao* samples from Ecuador and the USA projects. This viral family was recently proposed by Kunfei et al. [91], who investigated *Sclerotinia sclerotiorum*, an important plant pathogenic fungus known to harbor a diverse range of mycoviruses. The newly identified viral sequences, *Theobroma cacao*-associated deltaflexivirus 1–4 (TcDV-1 to TcDV-4), were detected in three distinct BioProjects: PRJNA558793, PRJNA326055, and PRJNA785999. While the first two projects focused on the comparative analysis of genetic variants and domesticated trees, BioProject PRJNA785999 aimed to evaluate resistant and susceptible cocoa genotypes following *Phytophthora palmivora* inoculation. Notably, there are no known reports of *Deltaflexiviridae* members infecting *P. palmivora*. However, other *Phytophthora* species have been documented as hosts for viruses from the *Narnaviridae*, *Totiviridae*, *Endornaviridae*, *Megabirnaviridae*, and *Tombusviridae* families, as well as members of the order *Bunyavirales* [92–94].

The viral contigs TcDV-3 and TcDV-4 were identified in the SRA library SRR17587555, a *Phytophthora palmivora*-inoculated sample from the CCN51 genotype. Metatranscriptomic analysis of the assembled contigs in this library revealed that 56.9% of the contigs aligned with high-similarity protein sequences from the *Phytophthora* genus, while approximately 9.2% aligned with high-similarity targets from the *Moniliophthora* genus. The substantial presence of *Phytophthora* contigs, combined with the documented submitted inoculation conditions, suggests a potential pathogen–host interaction between TcDV-3/TcDV-4 and *P. palmivora*.

The sample infected with TcDV-1 exhibited 13% assembled contigs with high identity to sequences derived from species within the *Moniliophthora* genus in the total metatranscriptomic analysis. In contrast, TcDV-2 did not show significant fungal diversity, with only two contigs aligning to *Sphaceloma batatas*, a known phytopathogenic fungus. The sample containing TcDV-2 belongs to BioProject PRJNA558793, where another novel viral sequence, *Theobroma cacao*-associated kitavirus 2 (TcKV-2), was also identified.

Members of the *Kitaviridae* family are known to infect a wide range of host plants globally, causing economically significant diseases in crops such as citrus, tomato, and blueberry. These viruses are transmitted by mites of the *Brevipalpus* genus [63,95]. Recently, *Brevipalpus* mites have been reported infecting *Theobroma cacao* in Peru [96], and historical

data indicate their global widespread distribution [97,98]. The libraries infected with TcKV-1 and TcKV-2 originate from India and the USA, respectively. Despite the geographical distance between these samples, it is plausible to speculate that *Brevipalpus* mites could infect *T. cacao* in other regions and potentially serve as vectors for these novel kitaviruses.

The recently identified *Cacao leafroll virus* (CaLRV), a member of the Polerovirus genus within the *Solemoviridae* family, was also detected in our analysis. CaLRV was first described by Adegbola et al. (2023) [99] as “cacao polerovirus” and was subsequently fully characterized by the same research group in 2024 [67]. The samples containing CaLRV originate from BioProject PRJNA558793, which focuses on investigating disease resistance in *Theobroma cacao* through whole-genome and transcriptome sequencing of diverse genotypes.

The Badnavirus cocoa (BVC) was identified as an assembled contig in samples SRR9203102, SRR9203103, and SRR3217279 from French Guiana. These samples belong to BioProjects PRJNA326055 and PRJNA314774, both of which aimed to perform comparative analyses on the effects of domestication on genome evolution across different crop species and wild genotypes. Quantification analysis detected this viral sequence in a total of 11 RNA-seq libraries, encompassing additional geographic regions such as the USA and Colombia, but at lower TPM expression levels compared to the contigs from French Guiana.

Badnaviruses are known to integrate multiple sequences into the *Theobroma cacao* genome, with these integrations varying in type and prevalence among different genetic groups. This suggests that viral integration occurred before or during species diversification [100]. Recent studies have surveyed diverse cacao germplasm, identifying multiple insertions of varying lengths and orientations. Notably, these endogenous viral elements (EVEs) can interfere with host gene expression, potentially impacting plant performance [33]. The high diversity of integrated sequences, coupled with the limited characterization of different genotypes, highlights the largely unexplored nature of EVEs in the *T. cacao* genome. This may explain the observed quantification of BVC sequences in certain samples without direct assembly or alignment to known viral or EVE reference sequences.

While the exploration of public data is crucial, this approach is inherently accompanied by several limitations, such as incomplete metadata, potential biases in the sample origins (e.g., geographical location or tissue type), and the inability to perform experimental validation. Despite encountering some of these challenges in our study, we sought to present robust evidence through comprehensive bioinformatics analyses, which we believe substantiate the presence of viral sequences linked to cocoa-derived samples available in the NCBI SRA public database.

5. Conclusions

This study advances our understanding of the *Theobroma cacao* virome by uncovering a diverse array of previously uncharacterized viral sequences. The identification of eleven novel viruses, including a new Badnavirus, highlights the hidden complexity of cocoa-associated viruses, providing new insights into their diversity and ecological roles and interactions within the plant microbiome. The detection of *Capillovirus mali* in cocoa suggests a potential new host for this species, emphasizing the need for further surveillance and prospecting for emerging viral threats. Additionally, the discovery of novel mycoviruses reveals the intricate dynamics of cocoa-associated microbial communities, underscoring the complexity of plant–virus–microbiome interactions. These findings reinforce the importance of virome studies in agricultural systems, particularly for economically significant crops like cocoa, where viral infections pose a major threat to productivity. By expanding the catalog of cocoa-associated viruses, this research contributes to enhanced

disease monitoring and management strategies, ultimately supporting sustainable cocoa production and the resilience of global supply chains.

Supplementary Materials: The following supporting information can be downloaded at: <https://www.mdpi.com/article/10.3390/v17050624/s1>, Figure S1: t-SNE based on the relative abundance of viral species; Figure S2: Structural annotation of the novel viruses; Figure S3: Frequency of *S. tuberosum* contigs in *T. cacao*-derived samples; Figure S4: Quantification of previously known cocoa viruses; Table S1: Samples metadata; Table S2: Viral sequences characterization; Table S3: Overview of transcripts quantification.

Author Contributions: Conceptualization, E.R.G.R.A.; methodology, G.V.P.R., J.P.N.S., L.Y.M.F., J.A.M.P. and L.B.d.A.C.; software, G.V.P.R., J.P.N.S., L.Y.M.F. and L.B.d.A.C.; validation, G.V.P.R., J.P.N.S., L.Y.M.F. and E.R.G.R.A.; formal analysis, G.V.P.R., J.P.N.S., L.Y.M.F. and L.B.d.A.C.; resources, E.R.G.R.A.; data curation, G.V.P.R., J.P.N.S. and L.Y.M.F.; writing—original draft preparation, G.V.P.R., J.P.N.S., L.Y.M.F., J.A.M.P. and E.R.G.R.A.; writing—review and editing, G.V.P.R., J.P.N.S., L.Y.M.F., J.A.M.P., L.B.d.A.C. and E.R.G.R.A.; visualization, G.V.P.R., J.P.N.S. and L.Y.M.F.; supervision, E.R.G.R.A.; project administration, E.R.G.R.A.; funding acquisition, E.R.G.R.A. All authors have read and agreed to the published version of the manuscript.

Funding: This study was financed in part by the Coordenação de Aperfeiçoamento de Pessoal de Nível Superior—Brazil (CAPES)—Finance Code 001. E.R.G.R.A. is a researcher of the Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq).

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The data presented in this study are contained within the article and Supplementary Materials. Additional data may be available from the corresponding author upon reasonable request.

Acknowledgments: We thank all the members of the Virus Bioinformatics Laboratory—UESC for the fruitful discussions and the Laboratory of Molecular Biodiversity and Omics—LabiÔmicas for the local infrastructure.

Conflicts of Interest: The authors declare no conflicts of interest.

References

1. Nair, K.P.P. Cocoa (*Theobroma cacao* L.). In *The Agronomy and Economy of Important Tree Crops of the Developing World*; Elsevier: Amsterdam, The Netherlands, 2010; pp. 131–180. [\[CrossRef\]](#)
2. Kongor, J.E.; Hinneh, M.; de Walle, D.V.; Afoakwa, E.O.; Boeckx, P.; Dewettinck, K. Factors influencing quality variation in cocoa (*Theobroma cacao*) bean flavour profile—A review. *Food Res. Int.* **2016**, *82*, 44–52. [\[CrossRef\]](#)
3. Kongor, J.E.; Owusu, M.; Oduro-Yeboah, C. Cocoa production in the 2020s: Challenges and solutions. *CABI Agric. Biosci.* **2024**, *5*, 102. [\[CrossRef\]](#)
4. Paisic-Ramirez, R.; Hernández-Amasifuen, A.D.; Sánchez-Aguilar, W.D.; Corazon-Guivin, M.A.; Bobadilla, L.G.; Mansilla-Córdova, P.J.; Caetano, A.C.; Zuta, M.Z.S.; Guerrero-Abad, J.C. Effect of osmoregulatory on the secondary somatic embryogenesis of cocoa (*Theobroma cacao* L.). *J. Appl. Biol. Biotechnol.* **2024**, *12*, 177–183. [\[CrossRef\]](#)
5. Alain, B.K. Economic impact of cocoa culture in Ivory Coast and in Ghana from 1980 to 2015. *J. Hist. Archaeol. Anthr. Sci.* **2024**, *9*, 62–67. [\[CrossRef\]](#)
6. Obodai, J.; Asamoah, P.K.B.; Edusei, J. Cocoa Purchasing and the Issue of Insecurity in the Akontombra District in the Western Region of Ghana. *Pelita Perkeb.* **2018**, *34*, 128–136. [\[CrossRef\]](#)
7. Widhiyoga, G.; Wijayati, H. Challenges Faced by Cocoa-based Industries from Indonesia in Global Value Chains. *Husnayain Bus. Rev.* **2022**, *2*, 1–10. [\[CrossRef\]](#)
8. Edet, E.O.; Udoe, P.O.; Abang, S.O. Economic impact of climate change on cocoa production among South-Western states, Nigeria: Results from ricardian analysis. *Glob. J. Pure Appl. Sci.* **2018**, *24*, 171–180. [\[CrossRef\]](#)
9. Förste, F.; Bauer, L.; Streeck, C.; Radtke, M.; Reinholz, U.; Kadow, D.; Keil, C.; Mantouvalou, I. Quantitative Analysis and 2D/3D Elemental Imaging of Cocoa Beans Using X-ray Fluorescence Techniques. *Anal. Chem.* **2023**, *95*, 5627–5634. [\[CrossRef\]](#)

10. Rojo-Poveda, O.; Barbosa-Pereira, L.; Zeppa, G.; Stévigny, C. Cocoa Bean Shell—A By-Product with Nutritional Properties and Biofunctional Potential. *Nutrients* **2020**, *12*, 1123. [\[CrossRef\]](#)
11. Forbes, S.J.; Northfield, T.D. Increased pollinator habitat enhances cacao fruit set and predator conservation. *Ecol. Appl.* **2017**, *27*, 887–899. [\[CrossRef\]](#)
12. Kadio, A.K.C. Challenges and Perspectives of Local Cocoa Transformation in Côte D’Ivoire: A Case Study on the Cocoa Industry in Different Production Areas. *Open J. Bus. Manag.* **2023**, *11*, 2849–2867. [\[CrossRef\]](#)
13. Lopes, M.; Junior, B.H.; Dias, C.; Santos, G.; Gramacho, K.; Cascardo, J.; Gesteira, A.; Micheli, F. Expression analysis of transcription factors from the interaction between cacao and *Moniliophthora perniciosa* (Tricholomataceae). *Genet. Mol. Res.* **2010**, *9*, 1279–1297. [\[CrossRef\]](#) [\[PubMed\]](#)
14. Sena, K.; Alemanno, L.; Gramacho, K.P. The infection process of *Moniliophthora perniciosa* in cacao. *Plant Pathol.* **2014**, *63*, 1272–1281. [\[CrossRef\]](#)
15. Litholdo, C.G.; Leal, G.A.; Albuquerque, P.S.B.; Figueira, A. Differential expression of jasmonate biosynthesis genes in cacao genotypes contrasting for resistance against *Moniliophthora perniciosa*. *Plant Cell Rep.* **2015**, *34*, 1747–1759. [\[CrossRef\]](#)
16. Rubini, M.R.; Silva-Ribeiro, R.T.; Pomella, A.W.V.; Maki, C.S.; Araújo, W.L.; dos Santos, D.R.; Azevedo, J.L. Diversity of endophytic fungal community of cacao (*Theobroma cacao* L.) and biological control of *Crinipellis perniciosa*, causal agent of Witches’ Broom Disease. *Int. J. Biol. Sci.* **2005**, *1*, 24–33. [\[CrossRef\]](#)
17. Phillips-Mora, W.; Wilkinson, M.J. Frosty Pod of Cacao: A Disease with a Limited Geographic Range but Unlimited Potential for Damage. *Phytopathology* **2007**, *97*, 1644–1647. [\[CrossRef\]](#)
18. Ploetz, R.C. Cacao Diseases: Important Threats to Chocolate Production Worldwide. *Phytopathology* **2007**, *97*, 1634–1639. [\[CrossRef\]](#)
19. Hebbar, P.K. Cacao Diseases: A Global Perspective from an Industry Point of View. *Phytopathology* **2007**, *97*, 1658–1663. [\[CrossRef\]](#)
20. Adomako, D.; Hutcheon, W.V. Carbohydrate Metabolism and Translocation in Healthy and Cocoa Swollen Shoot Virus-Infected Cocoa Plants. *Physiol. Plant.* **1974**, *30*, 90–96. [\[CrossRef\]](#)
21. Ploetz, R. The Impact of Diseases on Cacao Production: A Global Overview. In *Cacao Diseases*; Bailey, B.A., Meinhardt, L.W., Eds.; Springer International Publishing: Cham, Switzerland, 2016; pp. 33–59. [\[CrossRef\]](#)
22. Marelli, J.-P.; Guest, D.I.; Bailey, B.A.; Evans, H.C.; Brown, J.K.; Junaid, M.; Barreto, R.W.; Lisboa, D.O.; Puig, A.S. Chocolate Under Threat from Old and New Cacao Diseases. *Phytopathology* **2019**, *109*, 1331–1343. [\[CrossRef\]](#)
23. Dzahini-Obiatey, H.; Fox, R.T.V. Early signs of infection in Cacao swollen shoot virus (CSSV) inoculated cocoa seeds and the discovery of the cotyledons of the resultant plants as rich sources of CSSV. *Afr. J. Biotechnol.* **2010**, *9*, 593–603. [\[CrossRef\]](#)
24. Ramos-Sobrinho, R.; Chingandu, N.; Gutierrez, O.A.; Marelli, J.-P.; Brown, J.K. A Complex of Badnavirus Species Infecting Cacao Reveals Mixed Infections, Extensive Genomic Variability, and Interspecific Recombination. *Viruses* **2020**, *12*, 443. [\[CrossRef\]](#) [\[PubMed\]](#)
25. Obok, E.; Aikpokpodion, P.; Ani, O.; Allainguillaume, J.; Wetten, A. Cacao swollen shoot virus detection and DNA barcoding of its vectors and putative vectors in *Theobroma cacao* L. by using polymerase chain reaction. *BioTechnologia* **2021**, *102*, 229–244. [\[CrossRef\]](#) [\[PubMed\]](#)
26. Andres, C.; Hoerler, R.; Home, R.; Joerin, J.; Dzahini-Obiatey, H.K.; Ameyaw, G.A.; Domfeh, O.; Blaser, W.J.; Gatteringer, A.; Offei, S.K.; et al. Social network to inform and prevent the spread of cocoa swollen shoot virus disease in Ghana. *Agron. Sustain. Dev.* **2018**, *38*, 53. [\[CrossRef\]](#)
27. Muller, E. Cacao Swollen Shoot Virus (CSSV): History, Biology, and Genome. In *Cacao Diseases*; Bailey, B.A., Meinhardt, L.W., Eds.; Springer International Publishing: Cham, Switzerland, 2016; pp. 337–358. [\[CrossRef\]](#)
28. Abrokwa, 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*, 2011–2017. [\[CrossRef\]](#)
29. Ameyaw, G.A.; Wetten, A.; Dzahini-Obiatey, H.; Domfeh, O.; Allainguillaume, J. Investigation on Cacao swollen shoot virus (CSSV) pollen transmission through cross-pollination. *Plant Pathol.* **2013**, *62*, 421–427. [\[CrossRef\]](#)
30. Trebissou, C.I.; N’guessan, R.K.; Tahi, M.G.; Aïdara, S.; Akaffou, S.D.; N’guetta, S.-P.A. Response of Forty (40) High-Producing Cacao Genotypes to Cacao Swollen Shoot Disease. *J. Exp. Agric. Int.* **2020**, *42*, 1–10. [\[CrossRef\]](#)
31. Ameyaw, G.; Dzahini-Obiatey, H.; Domfeh, O. Perspectives on cocoa swollen shoot virus disease (CSSVD) management in Ghana. *Crop. Prot.* **2014**, *65*, 64–70. [\[CrossRef\]](#)
32. Friscina, A.; Chiappetta, L.; Jacquemond, M.; Tepfer, M. Infection of non-host model plant species with the narrow-host-range Cacao swollen shoot virus. *Mol. Plant Pathol.* **2017**, *18*, 293–297. [\[CrossRef\]](#)
33. Chingandu, N.; Zia-Ur-Rehman, M.; Sreenivasan, T.N.; Surujdeo-Maharaj, S.; Umaharan, P.; Gutierrez, O.A.; Brown, J.K. Molecular Characterization of Previously Elusive Badnaviruses Associated with Symptomatic Cacao in the New World. *Arch. Virol.* **2017**, *162*, 1363–1371. [\[CrossRef\]](#)

34. Chingandu, N.; Kouakou, K.; Aka, R.; Ameyaw, G.; Gutierrez, O.A.; Herrmann, H.-W.; Brown, J.K. 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. [CrossRef] [PubMed]
35. Kouakou, K.; Kébé, B.I.; Kouassi, N.; Aké, S.; Cilas, C.; Muller, E. Geographical Distribution of Cacao swollen shoot virus Molecular Variability in Côte d'Ivoire. *Plant Dis.* **2012**, *96*, 1445–1450. [CrossRef] [PubMed]
36. Ding, S.-W.; Mackenzie, A.; Torronen, M.; Gibbs, A. Nucleotide Sequence of the Virion Protein Gene of Cacao Yellow Mosaic Tymovirus. *Nucleic Acids Res.* **1990**, *18*, 5886. [CrossRef] [PubMed]
37. Puig, A.S. Detection of Cacao Mild Mosaic Virus (CaMMV) Using Nested PCR and Evidence of Uneven Distribution in Leaf Tissue. *Agronomy* **2021**, *11*, 1842. [CrossRef]
38. Guiraud, B.S.H.B.; Tahi, G.M.; Trebissou, I.C.; Coulibaly, K.; N'Guessan, W.P.; N'Guessan, K.F.; Muller, E.; Zoro, B.I. Genetic Diversity and Identification of Cocoa Swollen Shoot Virus Detected in Tolerant Cocoa Trees (*Theobroma cacao* L.) in West and Central West Department of Côte d'Ivoire 2023. Available online: <https://www.researchsquare.com/article/rs-3307863/v1> (accessed on 10 April 2025).
39. Ameyaw, G.A.; Dzahini-Obiatey, H.K.; Domfeh, O.; Oppong, F.K.; Abaka-Ewusie, K. History and Data Analyses of 'Cutting Out' Method for Cocoa Swollen Shoot Virus Disease (CSSVD) Control in Ghana. *J. Plant Dis. Prot.* **2015**, *122*, 200–206. [CrossRef]
40. Ofori, A.; Padi, F.K.; Ameyaw, G.A.; Dadzie, A.M.; Opoku-Agyeman, M.; Domfeh, O.; Ansah, F.O. Field evaluation of the impact of cocoa swollen shoot virus disease infection on yield traits of different cocoa (*Theobroma cacao* L.) clones in Ghana. *PLoS ONE* **2022**, *17*, e0262461. [CrossRef]
41. Chen, S.; Zhou, Y.; Chen, Y.; Gu, J. fastp: An ultra-fast all-in-one FASTQ preprocessor. *Bioinformatics* **2018**, *34*, i884–i890. [CrossRef]
42. Argout, X.; Salse, J.; Aury, J.-M.; Guiltinan, M.J.; Droc, G.; Gouzy, J.; Allegre, M.; Chaparro, C.; Legavre, T.; Maximova, S.N.; et al. The Genome of *Theobroma cacao*. *Nat. Genet.* **2011**, *43*, 101–108. [CrossRef]
43. Dobin, A.; Davis, C.A.; Schlesinger, F.; Drenkow, J.; Zaleski, C.; Jha, S.; Batut, P.; Chaisson, M.; Gingeras, T.R. STAR: Ultrafast universal RNA-seq aligner. *Bioinformatics* **2013**, *29*, 15–21. [CrossRef]
44. Bankevich, A.; Nurk, S.; Antipov, D.; Gurevich, A.A.; Dvorkin, M.; Kulikov, A.S.; Lesin, V.M.; Nikolenko, S.I.; Pham, S.; Pribelski, A.D.; et al. SPAdes: A New Genome Assembly Algorithm and Its Applications to Single-Cell Sequencing. *J. Comput. Biol.* **2012**, *19*, 455–477. [CrossRef]
45. Huang, X.; Madan, A. CAP3: A DNA Sequence Assembly Program. *Genome Res.* **1999**, *9*, 868–877. [CrossRef] [PubMed]
46. Wood, D.E.; Lu, J.; Langmead, B. Improved Metagenomic Analysis with Kraken 2. *Genome Biol.* **2019**, *20*, 257. [CrossRef] [PubMed]
47. Buchfink, B.; Xie, C.; Huson, D.H. Fast and Sensitive Protein Alignment Using DIAMOND. *Nat. Methods* **2015**, *12*, 59–60. [CrossRef] [PubMed]
48. Altschul, S.F.; Gish, W.; Miller, W.; Myers, E.W.; Lipman, D.J. Basic Local Alignment Search Tool. *J. Mol. Biol.* **1990**, *215*, 403–410. [CrossRef]
49. Singh, U.; Wurtele, E.S. Orfipy: A Fast and Flexible Tool for Extracting ORFs. *Bioinformatics* **2021**, *37*, 3019–3020. [CrossRef]
50. Wistrand, M.; Sonnhammer, E.L.L. Improved Profile HMM Performance by Assessment of Critical Algorithmic Features in SAM and HMMER. *BMC Bioinform.* **2005**, *6*, 99. [CrossRef]
51. Bigot, T.; Temmam, S.; Pérot, P.; Eloit, M. RVDB-Prot, A Reference Viral Protein Database and Its HMM Profiles. *F1000Research* **2019**, *8*, 530. [CrossRef]
52. Finn, R.D.; Bateman, A.; Clements, J.; Coghill, P.; Eberhardt, R.Y.; Eddy, S.R.; Heger, A.; Hetherington, K.; Holm, L.; Mistry, J.; et al. Pfam: The Protein Families Database. *Nucleic Acids Res.* **2014**, *42*, D222–D230. [CrossRef]
53. Li, D.; Liu, C.-M.; Luo, R.; Sadakane, K.; Lam, T.-W. MEGAHIT: An Ultra-Fast Single-Node Solution for Large and Complex Metagenomics Assembly via Succinct de Bruijn Graph. *Bioinformatics* **2015**, *31*, 1674–1676. [CrossRef]
54. Grabherr, M.G.; Haas, B.J.; Yassour, M.; Levin, J.Z.; Thompson, D.A.; Amit, I.; Adiconis, X.; Fan, L.; Raychowdhury, R.; Zeng, Q.D.; et al. Full-Length Transcriptome Assembly from RNA-Seq Data Without a Reference Genome. *Nat. Biotechnol.* **2011**, *29*, 644–652. [CrossRef]
55. Li, W.; Godzik, A. Cd-hit: A Fast Program for Clustering and Comparing Large Sets of Protein or Nucleotide Sequences. *Bioinformatics* **2006**, *22*, 1658–1659. [CrossRef] [PubMed]
56. Katoh, K.; Standley, D.M. MAFFT Multiple Sequence Alignment Software Version 7: Improvements in Performance and Usability. *Mol. Biol. Evol.* **2013**, *30*, 772–780. [CrossRef] [PubMed]
57. Larsson, A. AliView: A Fast and Lightweight Alignment Viewer and Editor for Large Datasets. *Bioinformatics* **2014**, *30*, 3276–3278. [CrossRef] [PubMed]
58. Nguyen, L.-T.; Schmidt, H.A.; Von Haeseler, A.; Minh, B.Q. IQ-TREE: A Fast and Effective Stochastic Algorithm for Estimating Maximum-Likelihood Phylogenies. *Mol. Biol. Evol.* **2015**, *32*, 268–274. [CrossRef]
59. Letunic, I.; Bork, P. Interactive Tree Of Life (iTOL) v5: An Online Tool for Phylogenetic Tree Display and Annotation. *Nucleic Acids Res.* **2021**, *49*, W293–W296. [CrossRef]

60. Patro, R.; Duggal, G.; Love, M.I.; Irizarry, R.A.; Kingsford, C. Salmon Provides Fast and Bias-Aware Quantification of Transcript Expression. *Nat. Methods* **2017**, *14*, 417–419. [\[CrossRef\]](#)
61. Xiao, J.; Wang, X.; Zheng, Z.; Wu, Y.; Wang, Z.; Li, H.; Li, P. Molecular Characterization of a Novel Deltaflexivirus Infecting the Edible Fungus *Pleurotus Ostreatus*. *Arch. Virol.* **2023**, *168*, 162. [\[CrossRef\]](#)
62. Bejerman, N.; Debat, H. Exploring the Tymovirales Landscape Through Metatranscriptomics Data. *Arch. Virol.* **2022**, *167*, 1785–1803. [\[CrossRef\]](#)
63. Ramos-González, P.L.; Arena, G.D.; Tassi, A.D.; Chabi-Jesus, C.; Kitajima, E.W.; Freitas-Astúa, J. Kitaviruses: A Window to Atypical Plant Viruses Causing Nonsystemic Diseases. *Annu. Rev. Phytopathol.* **2023**, *61*, 97–118. [\[CrossRef\]](#)
64. Chiba, S.; Suzuki, N.; Velasco, L.; Ayllón, M.A.; Lee-Marzano, S.-Y.; Sun, L.; Sabanadzovic, S.; Turina, M. ICTV Virus Taxonomy Profile: Fusariviridae 2024. *J. Gen. Virol.* **2024**, *105*, 001973. [\[CrossRef\]](#)
65. Le Gall, O.; Christian, P.; Fauquet, C.M.; King, A.M.Q.; Knowles, N.J.; Nakashima, N.; Stanway, G.; Gorbalenya, A.E. Picornavirales, a Proposed Order of Positive-Sense Single-Stranded RNA Viruses with a Pseudo-T = 3 Virion Architecture. *Arch. Virol.* **2008**, *153*, 715–727. [\[CrossRef\]](#) [\[PubMed\]](#)
66. Teycheney, P.-Y.; Geering, A.D.W.; Dasgupta, I.; Hull, R.; Kreuze, J.F.; Lockhart, B.; Muller, E.; Olszewski, N.; Pappu, H.; Pooggin, M.M.; et al. ICTV Virus Taxonomy Profile: Caulimoviridae. *J. Gen. Virol.* **2020**, *101*, 1025–1026. [\[CrossRef\]](#) [\[PubMed\]](#)
67. Muller, E.; Ullah, I.; Dunwell, J.M.; Daymond, A.J.; Richardson, M.; Allainguillaume, J.; Wetten, A. Identification and Distribution of Novel Badnaviral Sequences Integrated in the Genome of Cacao (*Theobroma cacao*). *Sci. Rep.* **2021**, *11*, 8270. [\[CrossRef\]](#) [\[PubMed\]](#)
68. Wemheuer, F.; Berkemann, D.; Wemheuer, B.; Daniel, R.; Vidal, S.; Bisseleua Daghele, H.B. Agroforestry Management Systems Drive the Composition, Diversity, and Function of Fungal and Bacterial Endophyte Communities in *Theobroma cacao* Leaves. *Microorganisms* **2020**, *8*, 405. [\[CrossRef\]](#)
69. Santana, J.O.; Gramacho, K.P.; de Souza Eduvirgens Ferreira, K.T.; Rezende, R.P.; Mangabeira, P.A.O.; Dias, R.P.M.; Couto, F.M.; Pirovani, C.P. Witches' Broom Resistant Genotype CCN51 Shows Greater Diversity of Symbiont Bacteria in Its Phylloplane than Susceptible Genotype Catongo. *BMC Microbiol.* **2018**, *18*, 194. [\[CrossRef\]](#)
70. Illegheems, K.; De Vuyst, L.; Papalexandratou, Z.; Weckx, S. Phylogenetic Analysis of a Spontaneous Cocoa Bean Fermentation Metagenome Reveals New Insights into Its Bacterial and Fungal Community Diversity. *PLoS ONE* **2012**, *7*, e38040. [\[CrossRef\]](#)
71. Agyirifo, D.S.; Wamalwa, M.; Otwe, E.P.; Galyuon, I.; Runo, S.; Takrama, J.; Ngeranwa, J. Metagenomics Analysis of Cocoa Bean Fermentation Microbiome Identifying Species Diversity and Putative Functional Capabilities. *Heliyon* **2019**, *5*, e02170. [\[CrossRef\]](#)
72. Almeida, O.G.G.; De Martinis, E.C.P. Metagenome-Assembled Genomes Contribute to Unraveling of the Microbiome of Cocoa Fermentation. *Appl. Environ. Microbiol.* **2021**, *87*, e0058421. [\[CrossRef\]](#)
73. Fernández-Niño, M.; Rodríguez-Cubillos, M.J.; Herrera-Rocha, F.; Anzola, J.M.; Cepeda-Hernández, M.L.; Mejía, J.L.A.; Chica, M.J.; Olarte, H.H.; Rodríguez-López, C.; Calderón, D.; et al. Dissecting Industrial Fermentations of Fine Flavour Cocoa Through Metagenomic Analysis. *Sci. Rep.* **2021**, *11*, 8638. [\[CrossRef\]](#)
74. Mota-Gutierrez, J.; Ferrocino, I.; Giordano, M.; Suarez-Quiroz, M.L.; Gonzalez-Ríos, O.; Cocolin, L. Influence of Taxonomic and Functional Content of Microbial Communities on the Quality of Fermented Cocoa Pulp-Bean Mass. *Appl. Environ. Microbiol.* **2021**, *87*, e0042521. [\[CrossRef\]](#)
75. Santos, J.P.N.; Rodrigues, G.V.P.; Ferreira, L.Y.M.; Monteiro, G.P.; Fonseca, P.L.C.; Lopes, Í.S.; Florêncio, B.S.; da Silva Junior, A.B.; Ambrósio, P.E.; Pirovani, C.P.; et al. The Virome of Cocoa Fermentation-Associated Microorganisms. *Viruses* **2024**, *16*, 1226. [\[CrossRef\]](#) [\[PubMed\]](#)
76. Espinal, R.B.A.; de Santana, S.F.; Santos, V.C.; Lizardo, G.N.R.; Silva, R.J.S.; Corrêa, R.X.; Loguercio, L.L.; Góes-Neto, A.; Pirovani, C.P.; Fonseca, P.L.C.; et al. Uncovering a Complex Virome Associated with the Cacao Pathogens *Ceratocystis Cacaofunesta* and *Ceratocystis Fimbriata*. *Pathogens* **2023**, *12*, 287. [\[CrossRef\]](#) [\[PubMed\]](#)
77. Reyes, B.M.D.; Fonseca, P.L.C.; Heming, N.M.; Conceição, L.B.d.A.; Nascimento, K.T.d.S.; Gramacho, K.P.; Arevalo-Gardini, E.; Pirovani, C.P.; Aguiar, E.R.G.R. Characterization of the Microbiota Dynamics Associated with *Moniliophthora roreri*, Causal Agent of Cocoa Frosty Pod Rot Disease, Reveals New Viral Species. *Front. Microbiol.* **2024**, *13*, 1053562. [\[CrossRef\]](#)
78. Li, F.; Wu, B.; Yan, L.; Qin, X.; Lai, J. Metabolome and Transcriptome Profiling of *Theobroma cacao* Provides Insights into the Molecular Basis of Pod Color Variation. *J. Plant Res.* **2021**, *134*, 1323–1334. [\[CrossRef\]](#) [\[PubMed\]](#)
79. Shujai, K.; Gibbs, A.J.; Hajizadeh, M. On the Evolution and Biogeography of Apple Stem Grooving Capillovirus. *AgriXiv* **2020**, 20203318783. [\[CrossRef\]](#)
80. Wang, Y.; Zhuang, H.; Yang, Z.; Wen, L.; Wang, G.; Hong, N. Molecular Characterization of an Apple Stem Grooving Virus Isolate from Kiwifruit (*Actinidia chinensis*) in China. *Can. J. Plant Pathol.* **2018**, *40*, 76–83. [\[CrossRef\]](#)
81. Chung, B.N.; Kwon, S.-J.; Yoon, J.-Y.; Cho, I.-S. First Report of *Cnidium officinale* as a Natural Host Plant of Apple Stem Grooving Virus in South Korea. *Plant Dis.* **2022**, *106*, 338. [\[CrossRef\]](#)
82. EFSA Panel on Plant Health (PLH); Jeger, M.; Bragard, C.; Caffier, D.; Dehnen-Schmutz, K.; Gilioli, G.; Gregoire, J.; Miret, J.A.J.; MacLeod, A.; Navarro, M.N.; et al. Pest Categorisation of Tatter Leaf Virus. *EFSA J.* **2017**, *15*, e05033. [\[CrossRef\]](#)

83. Chu, Y.-M.; Lim, W.-S.; Yea, S.-J.; Cho, J.-D.; Lee, Y.-W.; Kim, K.-H. Complexity of dsRNA Mycovirus Isolated from *Fusarium Graminearum*. *Virus Genes* **2004**, *28*, 135–143. [\[CrossRef\]](#)
84. Li, P.; Bhattacharjee, P.; Wang, S.; Zhang, L.; Ahmed, I.; Guo, L. Mycoviruses in *Fusarium* Species: An Update. *Front. Cell. Infect. Microbiol.* **2019**, *9*, 257. [\[CrossRef\]](#)
85. Marzano, S.-Y.L.; Nelson, B.D.; Ajayi-Oyetunde, O.; Bradley, C.A.; Hughes, T.J.; Hartman, G.L.; Eastburn, D.M.; Domier, L.L. Identification of Diverse Mycoviruses through Metatranscriptomics Characterization of the Viromes of Five Major Fungal Plant Pathogens. *J. Virol.* **2016**, *90*, 6846–6863. [\[CrossRef\]](#) [\[PubMed\]](#)
86. Leandro-Muñoz, M.E.; Tixier, P.; Germon, A.; Rakotobe, V.; Phillips-Mora, W.; Maximova, S.; Avelino, J. Effects of Microclimatic Variables on the Symptoms and Signs Onset of *Moniliophthora roreri*, Causal Agent of Moniliophthora Pod Rot in Cacao. *PLoS ONE* **2017**, *12*, e0184638. [\[CrossRef\]](#) [\[PubMed\]](#)
87. Delgado-Ospina, J.; Molina-Hernández, J.B.; Chaves-López, C.; Romanazzi, G.; Paparella, A. The Role of Fungi in the Cocoa Production Chain and the Challenge of Climate Change. *J. Fungi* **2021**, *7*, 202. [\[CrossRef\]](#) [\[PubMed\]](#)
88. Myers, J.M.; James, T.Y. Mycoviruses. *Curr. Biol.* **2022**, *32*, R150–R155. [\[CrossRef\]](#)
89. Mu, F.; Li, B.; Cheng, S.; Jia, J.; Jiang, D.; Fu, Y.; Cheng, J.; Lin, Y.; Chen, T.; Xie, J. Nine Viruses from Eight Lineages Exhibiting New Evolutionary Modes That Co-Infect a Hypovirulent Phytopathogenic Fungus. *PLOS Pathog.* **2021**, *17*, e1009823. [\[CrossRef\]](#)
90. Li, K.; Zheng, D.; Cheng, J.; Chen, T.; Fu, Y.; Jiang, D.; Xie, J. Characterization of a Novel Sclerotinia Sclerotiorum RNA Virus as the Prototype of a New Proposed Family Within the Order Tymovirales. *Virus Res.* **2016**, *219*, 92–99. [\[CrossRef\]](#)
91. Poimala, A.; Vainio, E.J. Complete Genome Sequence of a Novel Toti-like Virus from the Plant-Pathogenic Oomycete Phytophthora Cactorum. *Arch. Virol.* **2020**, *165*, 1679–1682. [\[CrossRef\]](#)
92. Botella, L.; Jung, M.H.; Rost, M.; Jung, T. Natural Populations from the *Phytophthora palustris* Complex Show a High Diversity and Abundance of ssRNA and dsRNA Viruses. *J. Fungi* **2022**, *8*, 1118. [\[CrossRef\]](#)
93. Raco, M.; Vainio, E.J.; Sutela, S.; Eichmeier, A.; Hakalová, E.; Jung, T.; Botella, L. High Diversity of Novel Viruses in the Tree Pathogen Phytophthora Castaneae Revealed by High-Throughput Sequencing of Total and Small RNA. *Front. Microbiol.* **2022**, *13*, 911474. [\[CrossRef\]](#)
94. Quito-Avila, D.F.; Freitas-Astúa, J.; Melzer, M.J. Bluner-, cile-, and higreviruses (Kitaviridae). In *Encyclopedia of Virology*; Elsevier: Amsterdam, The Netherlands, 2021; pp. 247–251, ISBN 9780128145166.
95. Escobar-Garcia, H.A.; de Andrade, D.J.; Carrillo, D.; Ochoa, R. *Theobroma cacao*, a New Host for Brevipalpus Yothersi (Acari: Tenuipalpidae) in Peru. *Acarologia* **2021**, *61*, 211–216. [\[CrossRef\]](#)
96. Knorr, L.C.; Denmark, H.A. Injury to Citrus by the Mite Brevipalpus Phoenixis12. *J. Econ. Entomol.* **1970**, *63*, 1996–1998. [\[CrossRef\]](#)
97. Rodrigues, J.C.V.; Gallo-Meagher, M.; Ochoa, R.; Childers, C.C.; Adams, B.J. Mitochondrial DNA and RAPD Polymorphisms in the Haploid Mite Brevipalpus Phoenixis (Acari: Tenuipalpidae). *Exp. Appl. Acarol.* **2004**, *34*, 275–290. [\[CrossRef\]](#) [\[PubMed\]](#)
98. Adegbola, R.O.; Keith, C.V.; Gutierrez, O.A.; Goenaga, R.; Brown, J.K. A Previously Undescribed Polerovirus (Solemoviridae) Infecting *Theobroma cacao* Germplasm. *Plant Dis.* **2022**, *107*, 975. [\[CrossRef\]](#) [\[PubMed\]](#)
99. Adegbola, R.O.; Keith, C.V.; Gutierrez, O.; Goenaga, R.; Brown, J.K. Complete Genome Characterization of Cacao Leafroll Virus, a Newly Described Cacao-Infecting Polerovirus. *Arch. Virol.* **2024**, *169*, 83. [\[CrossRef\]](#)
100. Ullah, I.; Dunwell, J.M. 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. [\[CrossRef\]](#)

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.