

A new genus of alphasatellites associated with banana bunchy top virus in Southeast Asia

Valentin Guyot,¹ Tien-Dung Trieu,^{2,†} Oudomphone Insiengmay,^{3,†} Ting Zhang,^{4,†} Marie-Line Iskra-Caruana,^{5,†} BforBB Consortium,[†] and Mikhail M. Pooggin^{1,*†}

¹PHIM Plant Health Institute, University of Montpellier, INRAE, CIRAD, IRD, Institute Agro, Montpellier 34398, France, ²Northern Mountainous Agriculture and Forestry Science Institute, Phu Tho City 290000, Vietnam, ³Life Science Research Centre, Science and Innovation Research Institute, Ministry of Education and Sports, Dontiew Road, Xaythany District, Vientiane Capital 99241, Laos, ⁴Kunming Institute of Botany, Chinese Academy of Sciences, Kunming 650201, China and ⁵DGD-RS, CIRAD, Montpellier 34398, France

[†]BforBB Consortium. Other members of the Consortium: Ngoc-Sam Ly (Institute of Tropical Biology, Vietnam), Thomas Haeveermans (Muséum National d'Histoire Naturelle, France), Matthieu Chabannes (CIRAD, France), and Gabriel Sachter-Smith (Hawaii Banana Source, USA).

[†]<https://orcid.org/0000-0003-2308-393X>

*Corresponding author: E-mail: mikhail.pooggin@inrae.fr

Abstract

Autonomously replicating alphasatellites (family Alphasatellitidae) are frequently associated with plant single-stranded (ss)DNA viruses of the families Geminiviridae, Metaxyviridae, and Nanoviridae. Alphasatellites encode a single replication-initiator protein (Rep) similar to Rep proteins of helper viruses and depend on helper viruses for encapsidation, movement, and transmission. Costs versus benefits of alphasatellite-helper virus association are poorly understood. Our surveys in Southeast Asia (SEA) for wild and cultivated banana plants infected with banana bunchy top virus (BBTV, Nanoviridae) and Illumina sequencing reconstruction of their viromes revealed, in addition to a six-component BBTV genome, one to three distinct alphasatellites present in sixteen of twenty-four BBTV-infected plants. Comparative nucleotide and Rep protein sequence analyses classified these alphasatellites into four distinct species: two known species falling into the genus *Muscarsatellite* (subfamily Petromoalphasatellitinae) previously identified in SEA and two novel species falling into the tentative genus *Banaphisatellite* (subfamily Nanoalphasatellitinae) so far containing a single species recently identified in Africa. The banaphisatellites were found to be most related to members of the genus *Fabenesatellite* of subfamily Nanoalphasatellitinae and the genus *Gosmusatellite* of subfamily Geminialphasatellitinae, both infecting dicots. This suggests a dicot origin of banaphisatellites that got independently associated with distinct strains of monocot-infecting BBTV in Africa and SEA. Analysis of conserved sequence motifs in the common regions driving replication and gene expression of alphasatellites and BBTV strains revealed both differences and similarities, pointing at their ongoing co-evolution. An impact of alphasatellites on BBTV infection and evasion of RNA interference-based antiviral defences was evaluated by measuring relative abundance of BBTV genome components and alphasatellites and by profiling BBTV- and alphasatellite-derived small interfering RNAs. Taken together, our findings shed new light on the provenance of alphasatellites, their co-evolution with helper viruses, and potential mutual benefits of their association.

Keywords: alphasatellite; banana bunchy top virus; multipartite; ssDNA virus; evolution; RNAi

Introduction

Banana bunchy top virus (BBTV, genus *Babuvirus*, family Nanoviridae) is a six-component circular single-stranded (ss)DNA virus that infects monocots mainly from the families Musaceae and Zingiberaceae, causing one of the most devastating diseases of cultivated bananas and plantains (*Musa* sp.) (Dale 1987; Qazi 2016). BBTV is mainly transmitted by the aphid vector *Pentalonia nigronervosa* in a circulative non-propagative manner (Magee 1940; Watanabe et al. 2013). Banana bunchy top disease is present in all continents except the Americas (Kumar et al. 2015). Sequenced isolates of BBTV are classified into two phylogenetic groups with distinct geography: the Southeast Asia (SEA) group (China, Indonesia, Japan, Philippines, Taiwan, Thailand, and Vietnam) and

the Pacific and Indian Oceans (PIO) group (Australia, Burundi, Cameroon, Democratic Republic of the Congo, Egypt, Fiji, Gabon, Hawaii, India, Malawi, Myanmar, Pakistan, Republic of Congo, Rwanda, Samoa, Sri Lanka, and Tonga) (Stainton et al. 2015).

The BBTV genome is composed of DNA-C encoding a cell-cycle link protein (Clink) (Lageix et al. 2007), DNA-M encoding a movement protein (MP) (Wanitchakorn et al. 2000), DNA-N encoding a nuclear shuttle protein (NSP) (Ji et al. 2019) implicated in aphid transmission (Guyot et al. 2022), DNA-R encoding a master replication protein (M-Rep) mediating rolling-circle replication of all six BBTV components (Horser, Harding, and Dale 2001), DNA-S encoding a coat protein (CP) encapsidating each viral ssDNA component individually (Wanitchakorn, Harding, and Dale 1997),

and DNA-U3 encoding a small protein of unknown function (Beetham, Harding, and Dale 1999). BBTV genome components have a size of ca. 1.0–1.1 kb and share two common regions with high sequence identities: the common-region stem-loop (CR-SL) and the common-region major (CR-M) (Burns, Harding, and Dale 1995). The CR-SL is an origin of rolling-circle replication that contains an inverted repeat forming a stem-loop secondary structure with the invariant nonanucleotide sequence TATTATTAC in the loop and serves as a binding site of M-Rep (Hafner et al. 1997). For specific recognition by M-Rep, the CR-SL possesses three iterated 5-nt long repeats (iterons): two tandem repeats in forward orientation (iterons F1 and F2) and the third repeat in reverse orientation (iteron R). The iterons are essential for efficient rolling-circle replication (Herrera-Valencia et al. 2006). The CR-M is a primer binding site for the synthesis of the viral DNA complementary strand on the virion strand template, which produces circular double-stranded DNA, the template for both rolling-circle replication and Pol II transcription of viral mRNAs (Hafner, Harding, and Dale 1997; Guyot et al. 2022).

Like other nanovirids and other plant circular ssDNA viruses of the families Geminiviridae and Metaxyviridae, BBTV is frequently associated with circular ssDNA alphasatellites (family Alphasatellitidae) which are non-essential for viral replication and infection cycles (Briddon and Stanley 2006; Briddon et al. 2018). Alphasatellites are classified into the subfamilies Geminialphasatellitinae associated with helper viruses of the family Geminiviridae, Nanoalphasatellitinae associated with helper viruses of the genus *Nanovirus* of family Nanoviridae, and Petromalphasatellitinae associated with helper viruses of the genus *Babuvirus* of family Nanoviridae and the genus *Cofodevirus* of family Metaxyviridae (Varsani et al. 2021). All alphasatellites encode a single Rep protein, related to M-Rep encoded by nanovirid DNA-R, which mediates replication of alphasatellite DNA but cannot mediate trans-replication of helper virus genome components (Horser, Harding, and Dale 2001). Alphasatellites depend on the helper viruses for encapsidation, movement, and transmission by insect vectors (Briddon and Stanley 2006).

Impacts of alphasatellites on nanovirid replication, systemic infection, and transmission by aphids have recently been studied for faba bean necrotic yellows virus (FBNYV, genus *Nanovirus*) and BBTV. Subterranean clover stunt alphasatellite 1 (genus *Subclousatellite*, Nanoalphasatellitinae) associated with FBNYV was shown to modify the relative abundance of helper virus DNA components (genome formula) in plants and aphids and to increase the helper virus transmission rate, despite a substantial reduction of helper virus DNA accumulation in aphids (Mansourpour et al. 2022). In the case of BBTV, a newly emerging alphasatellite associated with BBTV in Democratic Republic of the Congo (DRC) was shown to accumulate at high levels in plants and aphids, thereby reducing helper virus loads, altering its genome formula, and interfering with virus transmission by aphids (Guyot et al. 2022). Illumina sequencing analysis of viral transcripts and small interfering (si)RNAs revealed that the DRC alphasatellite decreases transcription efficiency of DNA-N encoding a putative aphid transmission factor and increases siRNA production rates from M-Rep- and MP-encoding components. Notably, the plant RNA interference (RNAi) machinery was found to silence DRC alphasatellite Rep-gene expression at both transcriptional and posttranscriptional levels, generating highly abundant 21, 22, and 24 nucleotide (nt) siRNAs, suggesting that the alphasatellite may serve as a decoy protecting its helper virus from antiviral RNAi (Guyot et al. 2022). DRC alphasatellite is the first satel-

lite reported to be associated with BBTV isolates from the PIO phylogenetic group and represents a new species classified in a tentative genus *Banaphisatellite*. It is more related to alphasatellites of the subfamily Nanoalphasatellitinae infecting dicots rather than to other BBTV alphasatellites of the subfamily Petromalphasatellitinae infecting monocots and previously identified only in SEA (Guyot et al. 2022). More specifically, the latter alphasatellites were discovered first in Taiwan in 1994 and later in China and Vietnam (2001–2013) and were classified into two distinct genera of *Petromalphasatellitinae*—*Babusatellite* and *Muscarsatellite* (see Supplementary Table S1 and references therein).

Here, we describe the identification and molecular characterization of multiple alphasatellites associated with BBTV isolates from wild and cultivated bananas sampled during recent surveys in Vietnam, Laos, and China. These alphasatellites include several new genetic variants of two previously reported species from the genus *Muscarsatellite* as well as several alphasatellites classified into two novel species falling within the tentative genus *Banaphisatellite*. Besides comparative analysis of complete nucleotide sequences of the newly discovered and previously reported alphasatellites, we performed clustering analysis of alphasatellite Rep proteins, which revealed evolutionary links of banaphisatellites to alphasatellites classified into the genus *Fabenesatellite* of Nanoalphasatellitinae and the genus *Gosmusatellite* of subfamily Geminialphasatellitinae, both infecting dicots. This supports the hypothesis on a dicot origin of banaphisatellites that became independently associated with monocot-infecting BBTV in PIO (DRC) and SEA (Vietnam, Laos, and China). Furthermore, we identified conserved sequence motifs in the common regions driving replication and gene expression of alphasatellites and BBTV from PIO and SEA phylogenetic groups. Finally, we compared relative abundances of alphasatellites and BBTV genome components as well as profiles of alphasatellite- and BBTV-derived siRNAs, which revealed the impact of alphasatellites on helper virus replication and evasion of RNAi-based antiviral defences.

Results and discussion

Diversity of BBTV alphasatellites in SEA

Leaf samples of wild and cultivated *Musa* plants exhibiting bunchy top disease symptoms were collected in Northern Vietnam ($n=11$), Laos ($n=10$), and the south of Yunnan province of China near the borders with Vietnam and Laos ($n=5$) during surveys conducted in 2018 and 2019 (Table 1; Supplementary Fig. S1). Total DNA was extracted from each sample and circular viral DNA was enriched by rolling-circle amplification (RCA), followed by Illumina sequencing of the RCA products and *de novo* assembly of the sequencing reads as described in Materials and Methods. BLAST analysis of the resulting consensus contigs revealed that they represent terminally redundant sequences of six circular components of BBTV genome in all samples, except one sample from Laos (ALYU-46) and one from China (ALYU-52). In the latter two samples, only negligible numbers of viral reads were detected by mapping to the reference BBTV genome: these reads likely represent cross-contamination from the other samples multiplexed and sequenced in one flow cell of Illumina HiSeq2500. These two samples served as negative controls to establish the cross-contamination threshold. In addition, BLASTn analysis revealed that sixteen of the twenty-four BBTV-infected samples contain a total of twenty-eight complete circular genomes of alphasatellites (1–3 per sample) (Table 1; Supplementary Dataset S1A).

Table 1. Virome components identified in wild and cultivated banana (*Musa*) sampled in SEA.

Sample/iso-late	Country	Plant species/genome/cultivar	BBTV genome Illumina	No. of alphasatellites	BBTA2	BBTA3	BBTA5	BBTA6	Defective (d) molecules	Other viruses
ALYU-25	Vietnam	<i>Musa itinerans</i>	full	2	1		1			Badnavirus BSMIV
ALYU-26	Vietnam	<i>Musa</i> sp.	full	3	1	1		1		
ALYU-29	Vietnam	<i>Musa</i> AA Pisang mas	full	3	1	1	1		dA5, dA2, dR	
ALYU-32	Vietnam	<i>Musa</i> sp. sweet banana	full	2	1		1		dA2	
ALYU-33	Vietnam	<i>Musa</i> AAB Chuoi Ngop	full	2	1		1			Badnavirus BSVNV
ALYU-34	Vietnam	<i>Musa</i> sp.	full	2	1		1			
ALYU-35	Vietnam	<i>Musa</i> sp.	full	1				1		
ALYU-36	Vietnam	<i>Musa</i> sp.	full	1	1					
ALYU-37	Vietnam	<i>Musa</i> sp.	full	1	1					
ALYU-39	Vietnam	<i>Musa</i> sp.	full	2	1			1		
ALYU-40	Vietnam	<i>Musa</i> AAA red banana	full	1			1			
ALYU-42	Laos	<i>Musa</i> AAA Cavendish	full	1			1			
ALYU-43	Laos	<i>Musa</i> AAA Cavendish	full	1			1			
ALYU-44	Laos	<i>Musa ornata</i>	full							
ALYU-45	Laos	<i>Musa</i> sp.	full							
ALYU-46	Laos	<i>Musa</i> ABB Klue Tiparot	no ^a							
ALYU-47	Laos	<i>Musa yunnanensis</i>	full							
ALYU-48	Laos	<i>Musa</i> sp.	full							
ALYU-49	Laos	<i>Musa</i> sp.	full							
ALYU-50	Laos	<i>Musa</i> AA Kouay niew mung	full							
ALYU-51	Laos	<i>Musa</i> ABB Pisang Awak?	full							
ALYU-52	China	<i>Musa acuminata</i> wild	no ^a							
ALYU-53	China	<i>Musa yunnanensis</i>	full							
ALYU-54	China	<i>Musa</i> AAA Cavendish	full	2	1			1	dA6	
ALYU-55	China	<i>Musa</i> AAA Cavendish	full	3	1	1		1	A5-U3 chimera	
ALYU-56	China	<i>Musa</i> AAA Cavendish	full	1	1					

^aBelow or at the cross-contamination threshold level.

Alphasatellites were present in all eleven samples from Vietnam, two of nine samples from Laos and three of four samples from China, indicating their high prevalence in SEA. Moreover, among the samples co-infected with BBTV and alphasatellites, two samples from Vietnam (ALYU-33 and ALYU-25) contained complete circular genomes of two distinct badnaviruses (genus *Badnavirus*, family Caulimoviridae) representing the known species of banana streak VN virus (BSNVV; binominal *Badnavirus iotavirgamusae*) and a novel virus species named after its host plant banana streak *Musa itinerans* virus (BSMIV; proposed binominal *Badnavirus kapavirgamusae*) and three samples contained circular molecules representing defective viral DNA components lacking protein-coding capacity (Table 1; Supplementary Dataset S1A). As a notable example, the sample ALYU-29 from Vietnam possessed not only three distinct alphasatellites but also defective molecules derived from two of the three alphasatellites and from BBTV DNA-R. In the follow up analysis of alphasatellites and the helper BBTV

from SEA, we also included Illumina data for the BBTV isolates from PIO described in our previous study (Guyot et al. 2022). They originated from DRC ($n=6$), Gabon ($n=3$), Benin ($n=1$), Malawi ($n=1$), and New Caledonia ($n=1$). Two samples from DRC (JGF-1 and ALYU-21) contained distinct variants of banana bunchy top alphasatellite 4 (BBTA4) (Supplementary Table S2; Supplementary Dataset S1A).

Pairwise comparison and phylogenetic analysis of nucleotide sequences of our new alphasatellites identified in banana samples from Vietnam, Laos, and China and of previously reported alphasatellites (Varsani et al. 2021) revealed that, based on the current species and genus demarcation criteria (~81 per cent and ~68 per cent sequence identity, respectively), the new satellites represent four different species, including two known species of BBTV alphasatellites—BBTA2 and BBTA3—both classified into the genus *Muscarsatellite* of subfamily Petromoalphasatellitinae, and two novel species that we classified together with BBTA4 into the

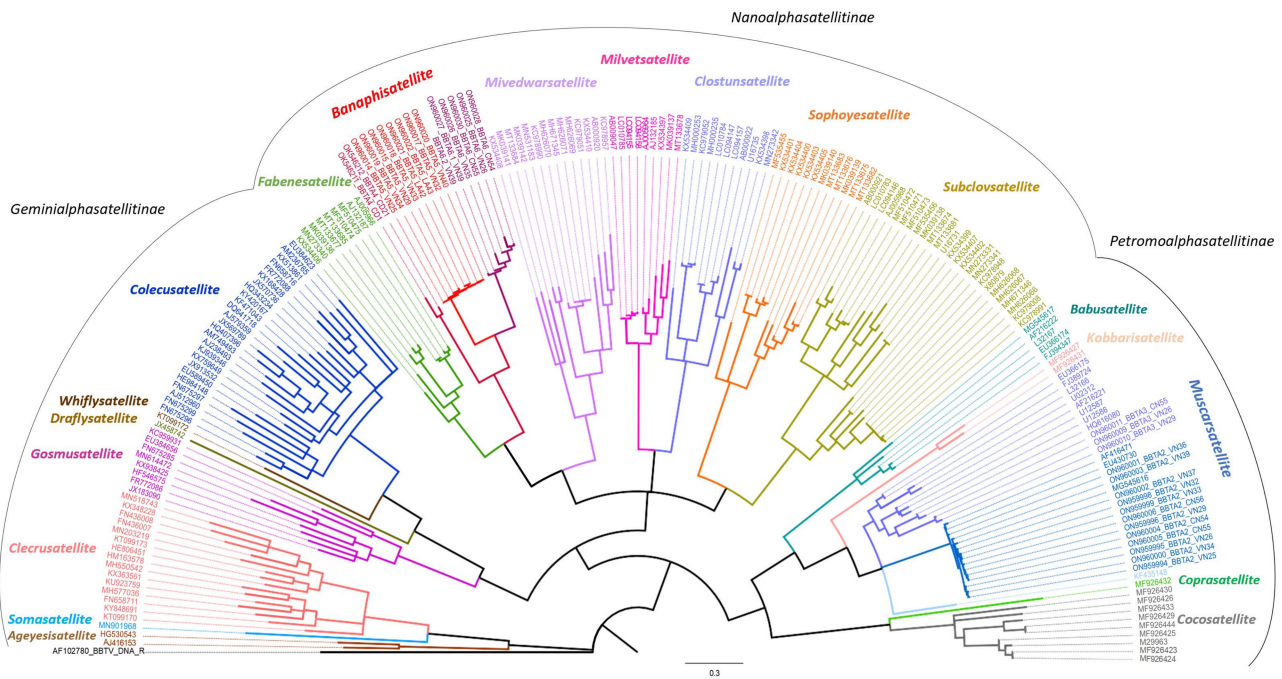


Figure 1. Phylogenetic analysis of complete nucleotide sequences of alphasatellites (family Alphasatellitidae). A maximum likelihood phylogenetic tree of complete nucleotide sequences of BBTV alphasatellites and alphasatellites associated with other helper viruses is rooted with BBTV DNA-R. Alphasatellites' genera are color-coded, and their subfamilies delineated. The NCBI Genbank accession number is given for each alphasatellite, and references to the original publications describing respective alphasatellites are provided in Guyot et al. (2022) (see the legend to Fig. 1 of that paper) and Varsani et al. (2021). In the case of the alphasatellites reported by Guyot et al. (2022) and here, the sequence IDs include in addition to the Genbank accession number, the alphasatellite species name (BETA2, BETA3, BETA4, BETA5, and BETA6), the two-letter country code (<https://www.iban.com/country-codes>), and the isolate number (1 for JGF-1, 21 for ALYU-21 and 25–56 for ALYU-25 to ALYU-56). In comparison with the phylogenetic tree shown in Guyot et al. (2022), bootstrap values <60 were collapsed.

genus *Banaphisatellite* of subfamily Nanoalphasatellitinae (Fig. 1; Supplementary Dataset S2). Keeping in line with previous nomenclature and numbering of BBTV alphasatellites, the newly discovered alphasatellites will hereafter be named banana bunchy top alphasatellite 5 (BETA5) and BETA6, respectively. To conform the newly adopted binominal nomenclature of viral species, we propose to name the three species classified in the genus *Banaphisatellite* as *Banaphisatellite alphamusae* comprising BETA4 isolates, *Banaphisatellite betamusae* comprising BETA5 isolates, and *Banaphisatellite gammamusae* comprising BETA6 isolates; note that all the species epithets include a latinized name of the host plant genus *Musa* (*musae*). BETA5 is represented with eight isolates and is prevalent in Vietnam (six of the eleven alphasatellite-containing samples) and Laos (each of the two alphasatellite-containing samples) but absent in samples from China. BETA6 is represented with five isolates and is present in Vietnam (3/11 samples) and China (two of the four alphasatellite-containing isolates) but absent in samples from Laos. The muscarsatellites BETA2 and BETA3 are prevalent in Vietnam (nine isolates of BETA2 and two isolates of BETA3) and China (two isolates of BETA2 and one isolate of BETA3), but absent in samples from Laos, and are frequently found in coinfections with banaphisatellites (eight samples in Vietnam and 2 samples in China) (Table 1). Among sixteen alphasatellite-containing samples, four samples contain only banaphisatellites (ALYU-35, ALYU-40, ALYU-42–43), four samples only muscarsatellites (ALYU-36–37, ALYU-56), and nine samples contain two or more alphasatellites with at least one member from both *Muscarsatellite* and *Banaphisatellite* genera (ALYU-25–26, ALYU-29, ALYU-32–34, ALYU-39, ALYU-54–55). In the samples from Vietnam and China, mixed infections of alphasatellites are

more frequent than infections with single alphasatellites (Table 1; Supplementary Dataset S1A).

Previously, babusatellite BETA1 and muscarsatellite BETA3 were reported only from China (Hainan province) and Taiwan, whereas muscarsatellite BETA2 was reported from China (Hainan province) and Vietnam (Supplementary Table S1 and references therein). In our survey, new isolates of BETA2 and BETA3 were detected in the north of Vietnam and the south of Yunnan province of China, geographically close to each other (Supplementary Fig. S1; Table 1). However, our sampling in Northern Laos and China near the border with Laos did not identify any BBTV alphasatellite and only samples from Central Laos near the border with Thailand contained BETA5 (Supplementary Fig. S1; Table 1). Thus, BBTV alphasatellites are less prevalent in the samples from Laos.

Using Sequence Demarcation Tool (SDT) (Muhire, Varsani, and Martin 2014), the intraspecies pairwise sequence identity of isolates of the banaphisatellites BETA5 and BETA6 was found to be 97.9–99.8 per cent and 93.5–98.7 per cent, respectively, and of the muscarsatellites BETA2 and BETA3 —96.3–99.4 per cent and 74.1–99.5 per cent, respectively (Supplementary Fig. S2). Thus, the most represented alphasatellites—BETA2 and BETA5—appear to have higher genetic stability than less represented ones, i.e. BETA3 and BETA6. Likewise, the babusatellite BETA1 (absent in our samples) has a low genetic stability among its previously identified isolates (84.7–97.7 per cent; Supplementary Fig. S2; Supplementary Dataset S2).

Considering sampling chronology, comparison of our new isolates of BETA2 with its isolates identified in 2002 and 2008 in Vietnam (AF416471 and EU430730) and in 2013 in China (MG545616)

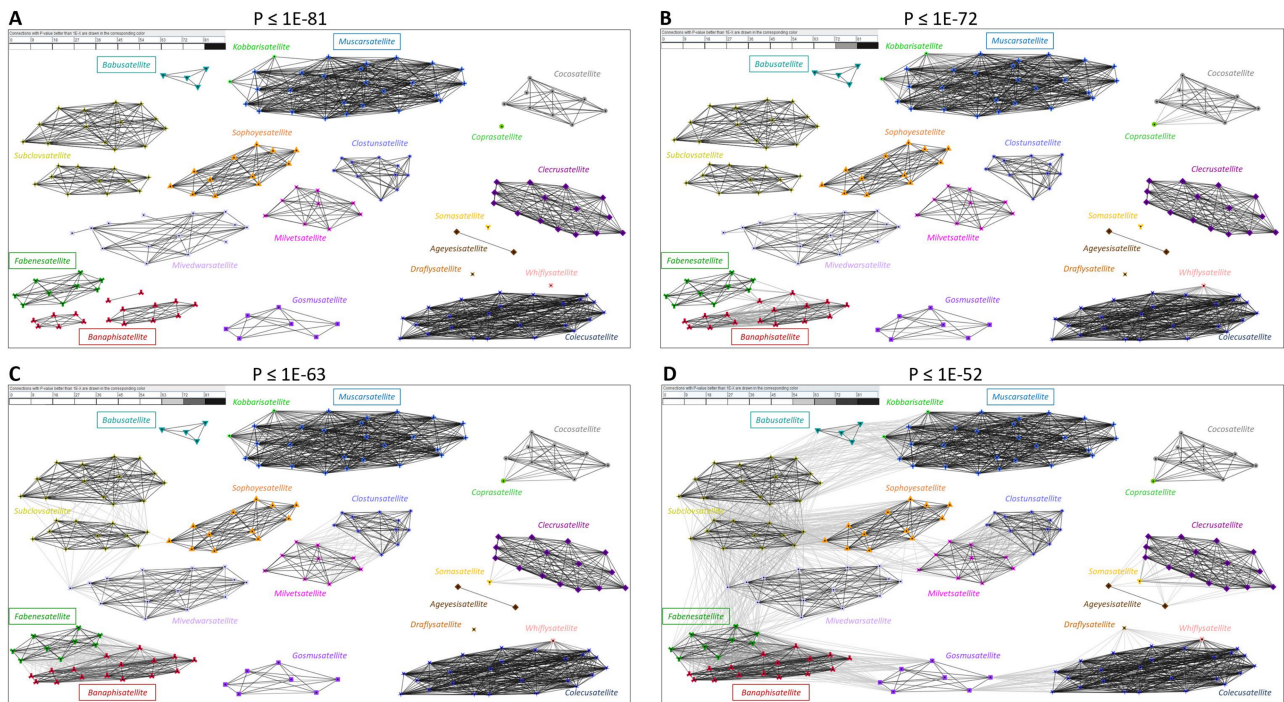


Figure 2. Sequence similarity-based clustering analysis of Rep proteins encoded by alphasatellites (family Alphasatellitidae). Rep proteins of the alphasatellites classified by [Bridson et al. \(2018\)](#) (see [Table 1](#) of that paper for the NCBI accession numbers) and [Varsani et al. \(2021\)](#) (see it for additional accession numbers) were compared ‘all-against-all’ and clustered using CLANS ([Frickey and Lupas 2004](#)). Members of the genus *Banaphisatellite* and all other genera of Alphasatellitidae are color-coded, and their evolutionary relatedness (link) to other alphasatellites is shown with solid grey lines whose colour intensity—from lightest to darkest (black)—indicates the strength of connections from worse (no direct link) to best at different threshold P -values (obtained using the Fruchterman–Reingold force-directed layout algorithm): $P \leq 1E-81$ (strongest link/black) (A), $P \leq 1E-72$ (weaker link/dark grey) (B), $P \leq 1E-63$ (weaker link/grey) (C), and $P \leq 1E-52$ (weakest link/light grey) (D).

shows the high genetic stability of this alphasatellite over 17 years. In contrast, our isolates of BBT43 from Vietnam (ALYU-26 and ALYU-29) share only 74.1–76.8 per cent pairwise identity with the isolates L32166 and U02312 (1994, Taiwan) and ca. 79.2–83.1 per cent identity with EU366175 (2007, Taiwan), FJ389724 (2009, Taiwan), and HQ616080 (2012, China), whereas its new isolate from China (ALYU-55) shares 82.4–97.4 per cent identity with the previous isolates from China and Taiwan ([Supplementary Fig. S2](#); [Supplementary Dataset S2](#)).

Clustering analysis of alphasatellite Rep proteins supports a dicot origin of banaphisatellites

To further understand the provenance of banaphisatellites, we performed sequence similarity-based clustering analysis of Rep proteins encoded by all isolates of banaphisatellites from DRC, Vietnam, Laos, and China in comparison with Rep proteins encoded by all alphasatellites from other genera of the subfamilies Nanoalphasatellitinae, Petromorphalphasatellitinae, and Geminalphasatellitinae. This analysis confirmed the nucleotide sequence-based classification of banaphisatellites in three species and their close evolutionary links with fabenesatellites (genus *Fabenesatellite*, subfamily Nanoalphasatellitinae) ([Fig. 2](#) vs [Fig. 1](#)). Indeed, at the most stringent threshold ($P \leq 1E-81$), isolates of each of the three banaphisatellites—BBTA4 ($n=2$), BBT45 ($n=8$), and BBT46 ($n=5$)—are linked to each other within each species but not between species ([Fig. 2A](#); [Supplementary Fig. S3A](#)), while at a less stringent threshold ($P \leq 1E-72$) all isolates of all the three banaphisatellites become interconnected within the genus *Banaphisatellite* ([Fig. 2B](#); [Supplementary Fig. S3B](#)). Notably at the latter threshold, both isolates of BBT44 (DRC alphasatellite)

become connected to some but not all fabenesatellites classified in a single species *Fabenesatellite* 2, the only species in the genus *Fabenesatellite*, whereas at a lower stringency ($P \leq 1E-63$) all alphasatellites of the *Fabenesatellite* and *Banaphisatellite* genera become interconnected, with none of them being linked to other genera ([Fig. 2C](#); [Supplementary Fig. S3C](#)). Finally, further decrease in stringency ($P \leq 1E-52$) resulted in the appearance of evolutionary links of both banaphisatellites and fabenesatellites to four of the seven alphasatellites from the genus *Gosmusatellite* of subfamily Geminalphasatellitinae ([Fig. 2D](#); [Supplementary Fig. S3D](#)). Notably, while the gosmusatellites are interconnected in one cluster at more stringent thresholds, at the less stringent threshold ($P \leq 1E-52$) they become connected to other clusters including the *Banaphisatellite-Fabenesatellite* cluster of Nanoalphasatellitinae and the cluster of *Colecusatellite*, *Draflysatellite*, and *Whiflysatellite* genera of Geminalphasatellitinae ([Fig. 2D](#)). The four gosmusatellites connected to the *Banaphisatellite-Fabenesatellite* cluster—Okra yellow crinkle Cameroon alphasatellite (FN675286), Hollyhock yellow vein virus associated symptomless alphasatellite (FR772086), Mesta yellow vein mosaic alphasatellite (JX183090), and Eclipta yellow vein alphasatellite (KX938425)—infect dicots of the families Malvaceae (FN675286, FR772086, JX183090) and Asteraceae (KX938425). The other three gosmusatellites (two of which are connected to the *Colecusatellite-Draflysatellite-Whiflysatellite* cluster)—Cotton leaf curl Gezira alphasatellite 3 (MN614472), *Gossypium mustelinum* symptomless alphasatellite (EU384656, LN880828, KF471041), and Vernonia yellow vein Fujian alphasatellite (KC959931, KC959932, JF733780)—have a wider dicot host range comprising the families Malvaceae (MN614472, EU384656), Asteraceae (JF733780),

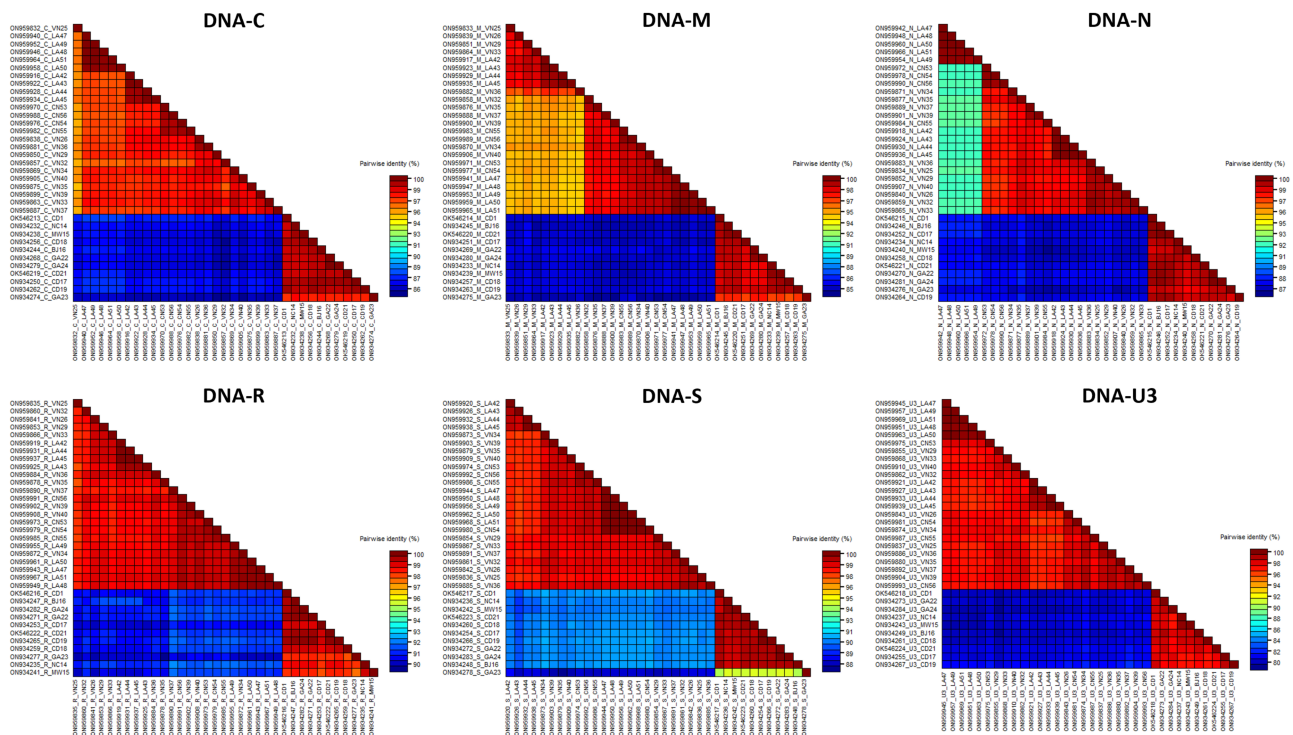


Figure 3. Pairwise sequence comparison of BBTV genome components from SEA and PIO. The nucleotide sequences of BBTV DNA-C, DNA-M, DNA-N, DNA-R, DNA-S, and DNA-U3 were compared using SDT v1.2 with Muscle (Muhire, Varsani, and Martin 2014) and their pairwise identities (in %) were plotted as heatmap diagrams. Sequence names include the NCBI Genbank accession number, the component name (C, M, N, R, S, U3), the two-letter country code (<https://www.iban.com/country-codes>; LA, Laos; VN, Vietnam; CN, China; CD, Democratic Republic of the Congo; BJ, Benin; GA, Gabon; MW, Malawi; NC, New Caledonia), and the isolate number (1 for JGF-1 and 14 to 56 for ALYU-14 to ALYU-56, respectively). Note that the U3 component was not detected in the BBTV-infected sample ALYU-23 from Gabon.

Urticaceae (LN880828), Caricaceae (KF471041), and Fabaceae (KC959931, KC959932). It should be noted that all fabenesatellites, which are most closely linked to banaphisatellites, infect dicots of the families Fabaceae (AJ132187, AJ005966, MF510474, MF510475, MN273340, KX534406) and Apiaceae (MK039136, MT133677, MT133685).

Interestingly, only at the least stringent threshold applied in our analysis ($P \leq 1E-52$), did fabenesatellites (but not banaphisatellites) become connected to other genera of the subfamily Nanoalphasatellitinae, namely *Midvesatellite* and *Subclovsatellite* (Fig. 2D). The genus *Subclovsatellite* is composed of two distinct clusters, with one cluster being linked to the genera *Fabenesatellite*, *Midvesatellite*, and *Sophoyesatellite* of Nanoalphasatellitinae and another being linked to the genera *Babusatellite* and *Muscarsatellite* of the subfamily Petromoalphasatellitinae (Fig. 2D) that contain all previously identified BBTV alphasatellites from SEA representing the babusatellite BBT1 and the muscarsatellites BBT2 and BBT3. Thus, Rep proteins of the babusatellite BBT1 and the muscarsatellites BBT2 and BBT3 are only distantly related to Rep proteins of the banaphisatellites BBT4, BBT5, and BBT6 that have much closer evolutionary links to Rep proteins of the dicot-infecting fabenesatellites and gosmusatellites. This supports the hypothesis on the provenance of banaphisatellites from a dicot-infecting alphasatellite ancestor associated with a helper virus from the genus *Nanovirus* (Nanoviridae) containing helper viruses of fabenesatellites, or from the genus *Begomovirus* (Geminiviridae) containing helper viruses of gosmusatellites.

Genome sequence comparison of BBTV isolates from SEA and PIO

Comparative nucleotide sequence analysis of our SEA and PIO isolates of BBTV showed their clear differences in each component of the BBTV genome, with the highest pairwise identity between the groups being in DNA-R (90.0–92.6 per cent), followed by DNA-S (87.1–90.7 per cent), DNA-N (85.8–87.7 per cent), DNA-C (85.0–87.6 per cent), DNA-M (83.7–85.6 per cent), and DNA-U3 (79.2–82.5 per cent) (Fig. 3; Supplementary Dataset S2). This confirms the previously established structuration of BBTV isolates into the two phylogenetic groups showing a distinct geographical delineation (Stainton et al. 2015). Interestingly, SEA isolates of DNA-N are split in two subgroups with 91.6–92.3 per cent between-subgroup identities and 97.2–100 per cent within-subgroup identities, respectively (Fig. 3; Supplementary Dataset S2): one subgroup is represented with five isolates from Northern Laos (ALYU-47–51) lacking alphasatellites, whereas another subgroup comprises nineteen BBTV isolates, sixteen of which contain alphasatellites. In the case of PIO isolates, a notable example is ALYU-23 from Gabon, which lacks DNA-U3 and whose DNA-S is a chimeric component with a 189 nt non-coding sequence derived from DNA-N (Supplementary Dataset S1A). Importantly, the two isolates of BBTV associated with banaphisatellite BBT4 in DRC (JGF-1 and ALYU-21) do not substantially differ from other isolates from DRC or other PIO countries (except ALYU-23 from Gabon) in any component of the BBTV genome (Fig. 3; Supplementary Dataset S2).

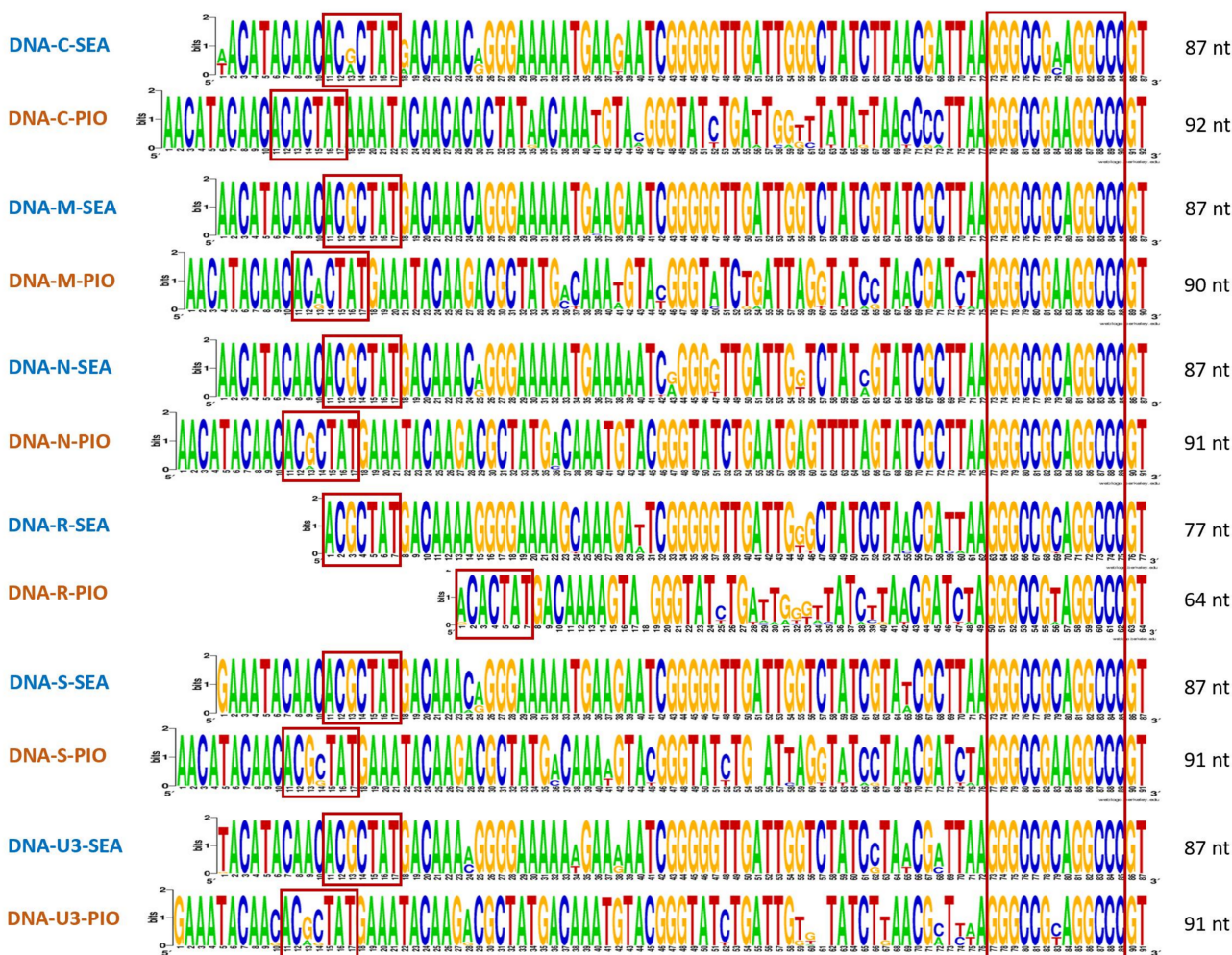


Figure 4. Logo of the CR-M sequences of the SEA and PIO isolates of BBTV. The CR-M sequences of BBTV DNA-C, DNA-M, DNA-N, DNA-R, DNA-S, and DNA-U3 of our PIO and SEA isolates were aligned separately using SeaView with Muscle. Their consensus sequence (logo) was generated using WebLogo v2.8.2. The conserved sequence motifs/cis-acting elements are boxes.

Our analysis of BBTV protein-coding sequences revealed that M-Rep (DNA-R), CP (DNA-S), Clink (DNA-C), MP (DNA-M), and NSP (DNA-N) proteins are well conserved in length and amino acid sequence between all PIO and SEA isolates, whereas no common ORF was found between PIO and SEA isolates of DNA-U3 (Supplementary Dataset S1). Our PIO isolates of DNA-U3 share one conserved ORF potentially encoding a seventy-seven amino acid protein of unknown function, which has previously been recognized in other PIO isolates (India, Pakistan, Sri Lanka, South Africa), although in one of our isolates (ALYU-17) this ORF is truncated due to a premature stop codon (Supplementary Dataset S1B). Likewise, in an Australian isolate (Burns, Harding, and Dale 1995) the conserved stop codon TGA is mutated to CG, thereby elongating an encoded protein to eighty-eight amino acids as we noted previously (Guyot et al. 2022). All our SEA isolates of DNA-U3 share a different ORF, potentially encoding a thirty-nine amino acid protein of unknown function, except for ALYU-40 where a premature stop codon makes an encoded protein 5 amino acid shorter (Supplementary Dataset S1A, B). We found this ORF to be conserved in some isolates previously reported from Thailand and to be elongated in some isolates from China. Given that no TATA-box can be identified upstream of this ORF, its functionality remains questionable. Interestingly, despite the drastic difference

in protein-coding capacities, the TATA-box and poly(A) signals previously identified to drive Pol II transcription and polyadenylation of U3-mRNA in the Australian (Beetham, Harding, and Dale 1999; Herrera-Valencia et al. 2007) and Congo-DRC (JGF-1) (Guyot et al. 2022) isolates are preserved in all our PIO and SEA isolates (Supplementary Dataset S1B). Taken together, SEA and PIO strains of DNA-U3 do not possess any conserved protein-coding capacity, suggesting either distinct strain-specific functions, or a non-protein-coding function of this component of BBTV. In a previous study, we have identified a plant microRNA of the conserved miR156 family that can potentially target U3-mRNA of the BBTV isolate from Congo-DRC for cleavage and degradation in *Musa acuminata* Cavendish (Guyot et al. 2022). Notably, the miR156 target site is well conserved in DNA-U3 of our PIO and SEA isolates with a single SNP position distinguishing the PIO and SEA strains (Supplementary Dataset S1B). Given the above-mentioned conservation of the TATA-box and poly(A) signal, the conserved Pol II transcription unit of DNA-U3 in both PIO and SEA strains of BBTV may serve a non-coding function by generating a decoy transcript diverting miR156 from its plant target genes, that is, *SQUAMOSA Promoter-Binding Protein-Like* (SPL) genes (Chen et al. 2010; Zhu et al. 2019). Since the SPL genes encode transcription factors that play important roles in plant growth and development (phase transition,

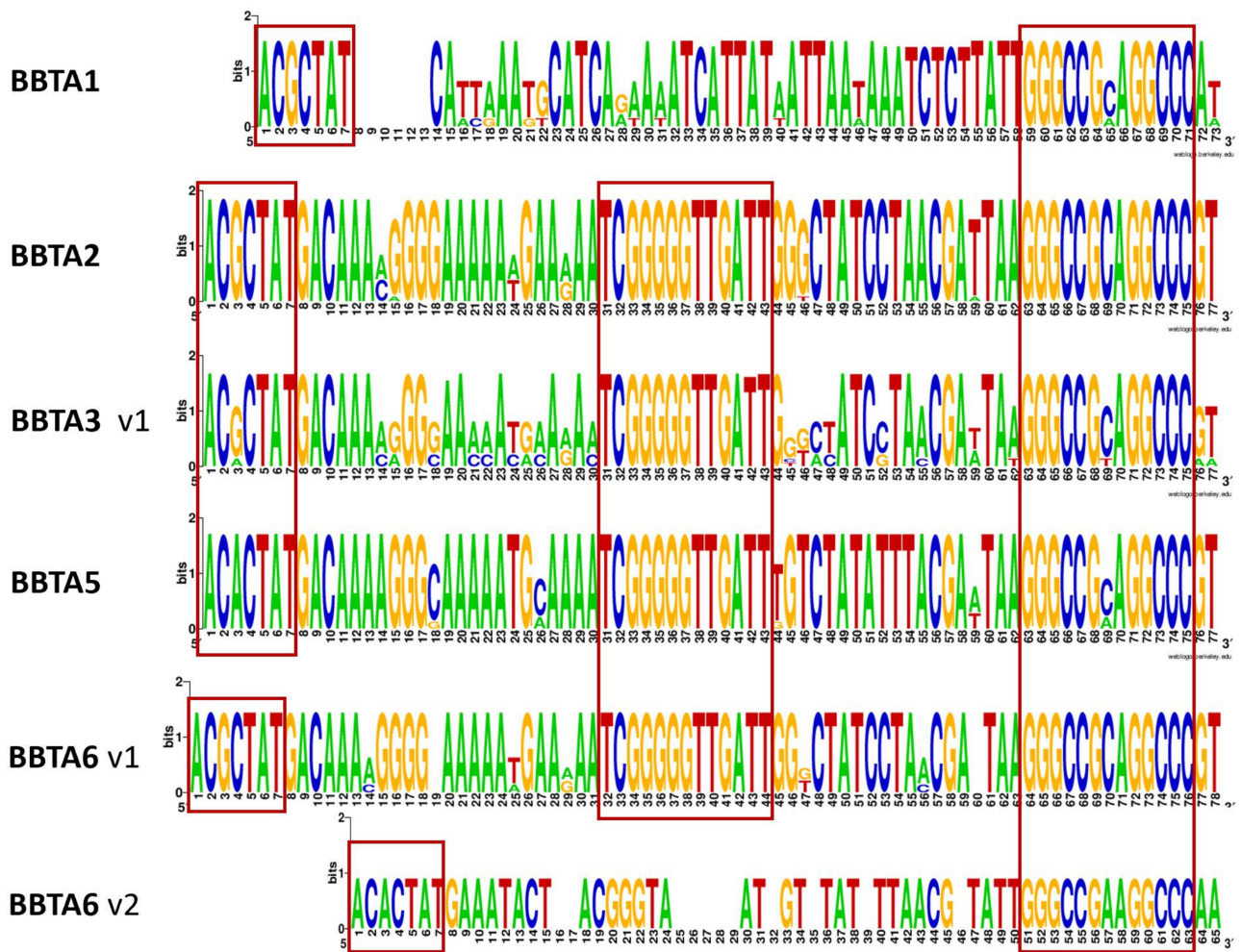


Figure 5. Logo of the CR-M sequences of SEA and PIO isolates of BBTV alphasatellites. The CR-M sequences of all isolates of the alphasatellite species BBTa1, BBTa2, BBTa5, and BBTa6 were aligned separately using SeaView with Muscle. Their consensus sequence (logo) was generated using WebLogo v2.8.2. The conserved sequence motifs/cis-acting elements are boxes.

flower and fruit development, architecture, hormone signalling, etc.; [Chen et al. 2010](#)), it is conceivable that DNA-U3-mediated deregulation of plant miR156-SPL pathway(s) may contribute to symptoms of the banana bunchy top disease such as stunting, smaller leaves with short petioles and no fruits.

Conserved sequence elements driving replication and gene expression of alphasatellites and helper viruses

To investigate molecular interactions of alphasatellites with helper viruses and potential similarities in their replication and gene expression mechanisms, we first compared conserved elements in the common regions shared by all BBTV genome components: the CR-SL, a Rep-binding origin of replication ([Herrera-Valencia et al. 2006](#)), and the CR-M containing the primer binding site for complementary DNA synthesis, located upstream of the CR-SL at various distances ([Burns, Harding, and Dale 1995](#); [Hafner, Harding, and Dale 1997](#)).

CR-M

Alignment of the CR-M sequences of BBTV components of our SEA and PIO isolates revealed that they range in size from 64 to 92

nts depending on the component and the phylogenetic group with the exception for ALYU-23 (Gabon) and ALYU-35 (Vietnam), whose DNA-N components have a longer CR-M with an identical 17 nt insert at the 5'-end (a direct repeat of the downstream sequence) ([Supplementary Fig. S4](#)). All CR-M sequences contain a highly conserved GC-rich element at the 3'-end that can potentially form a small stem-loop structure—AGGCCGHHAGGCCCGT (an inverted repeat is underlined; H = C, A, or T) ([Fig. 4](#); [Supplementary Fig. S4](#)). Sequences upstream of this conserved element fall into two distinct clades—SEA and PIO—and share high identities and lengths within the clades with the exception for DNA-R components that are shortened at the 5'-end by 10 and 26 nts, respectively ([Supplementary Fig. S4](#)). The 26 nt deletion in DNA-R CR-M has previously been reported for the Australian (PIO) isolate ([Burns, Harding, and Dale 1995](#); [Hafner, Harding, and Dale 1997](#)). Differences in CR-M sequences of PIO and SEA isolates have been mentioned, although not specified, by [Karan, Harding, and Dale \(1994\)](#).

We next inspected whether BBTV alphasatellites possess any region with homology to the CR-M of helper BBTV and located upstream of the CR-SL. The homologous sequences with a characteristic stem-loop region (GGGCCGHHAGGCC) at the 3'-end were identified in all BBTV alphasatellites except BBTa4 ([Fig. 5](#)). This shows that the banaphisatellite BBTa4 that recently emerged in

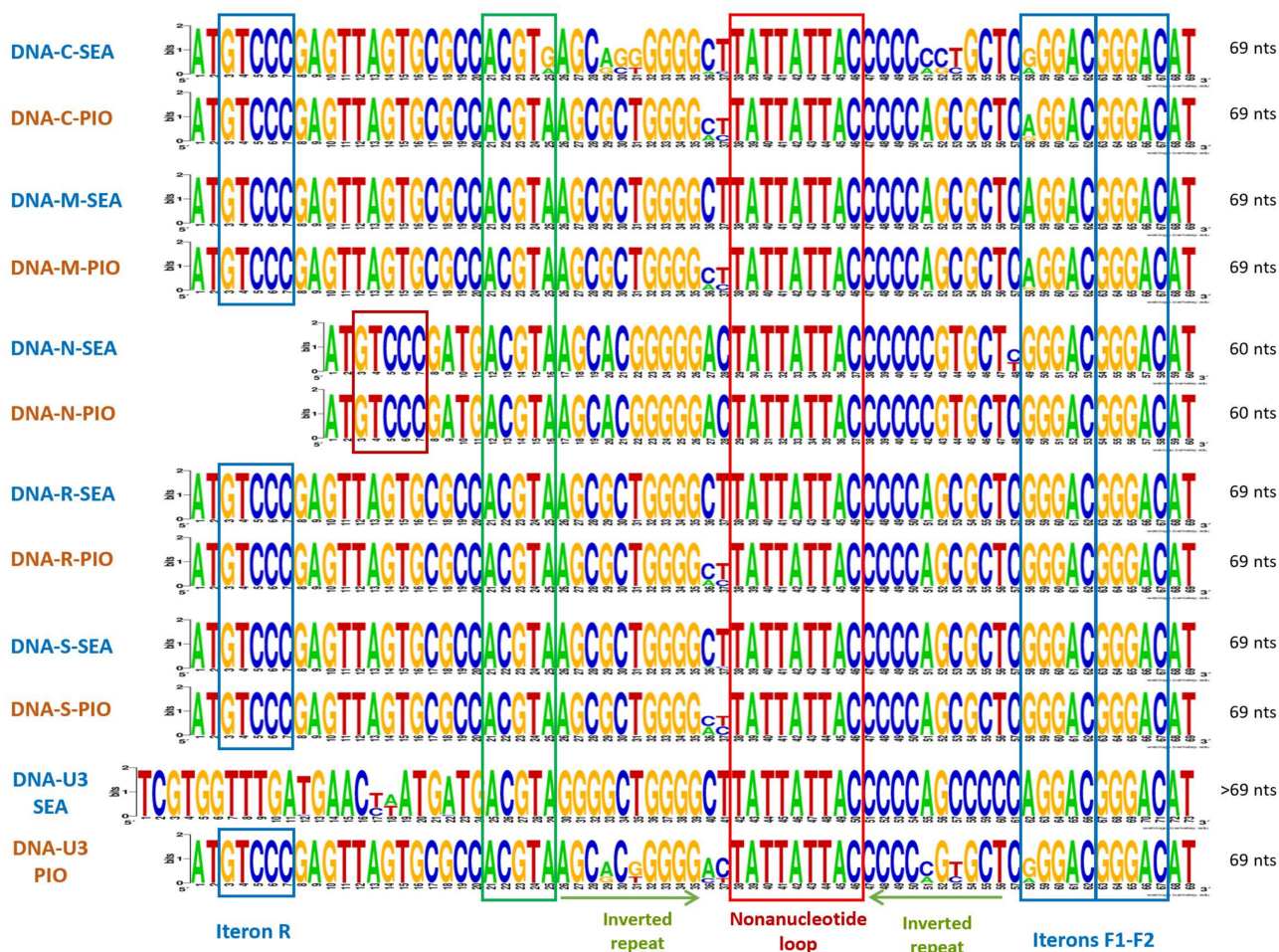


Figure 6. Logo of the CR-SL sequences of the SEA and PIO isolates of BBTv. The CR-SL sequences of BBTv DNA-C, DNA-M, DNA-N, DNA-R, DNA-S, and DNA-U3 of our PIO and SEA isolates were aligned separately using SeaView with Muscle. Their consensus sequence (logo) was generated using WebLogo v2.8.2. The conserved sequence motifs/cis-acting elements are boxes or underlined and named.

Africa (Guyot et al. 2022) differs from the SEA alphasatellites (i.e. banaphisatellites BBTA5 and BBTA6, muscarsatellites BBTA2 and BBTA3, and babusatellite BBTA1) in a priming mechanism of complementary DNA synthesis. Interestingly, the A and GT nucleotides flanking the stem-loop structure in all BBTv components (AGGCCCGHAGGCCCGT) are also conserved in some but not all the five alphasatellites. Moreover, the sequences upstream of the stem-loop region, which are well conserved in the BBTv components of SEA isolates, are highly similar to those of all isolates of BBTA2 and BBTA5 as well as seven of the eight isolates of BBTA3 and four of the six isolates of BBTA6 (Fig. 5; Supplementary Fig. S5, note that the BBTA3 isolates L32166, U02312 and U12586 with major alterations in CR-M and/or CR-SL were excluded from alignments). In contrast, one isolate of BBTA3 (ALYU-55), two isolates of BBTA6 (ALYU-26, ALYU-35), and all five isolates of BBTA1 (identified only in previous surveys) have deletions or alterations in the conserved sequence TCGGGGTTGATTG and further upstream sequences, except for the motif ACRCTAT (R=G or A) preserved at the 5'-end of CR-M of all alphasatellites (Fig. 5; Supplementary Fig. S5). The same motif is present at the 5'-end of CR-M of all the six BBTv components of both SEA and PIO isolates (Fig. 4; Supplementary Fig. S4), suggesting its involvement in priming the complementary DNA synthesis.

CR-SL

Alignment of the CR-SL sequences of BBTv components of our PIO and SEA isolates confirmed the presence of conserved cis-acting elements essential for rolling-circle replication (Herrera-Valencia et al. 2006): (i) the invariant nonanucleotide TATTATTAC flanked with 10 nt inverted repeats forming the stem-loop secondary structure and (ii) three 5 nt itérons—the two adjacent itérons F1 and F2 in the forward orientation (RGGACGGGAC, R=G or A) one nucleotide downstream of the stem-loop structure base and the iteron R in the reverse orientation (GTCCC) at a distance upstream of the stem-loop structure (Fig. 6; Supplementary Fig. S6A). The distance between the iteron R and the stem-loop structure base is conserved (26 nts) with an exception for DNA-N of both SEA and PIO isolates (9 nts) and DNA-U3 of SEA isolates (90 nts) (Supplementary Fig. S6B; note that DNA-U3 of PIO isolates has the second iteron GTCCC at the same distance of 90 nts). Thus, our PIO and SEA isolates of BBTv share all the previously described elements essential for rolling-circle replication. Further inspection of the CR-SL sequences revealed a previously unrecognized motif ACTGA located just upstream of the stem-loop structure and therefore may also be required for replication (Fig. 6). This motif is invariant in all the BBTv components of our PIO and SEA isolates with the exception of DNA-C of SEA isolates having a variation at the

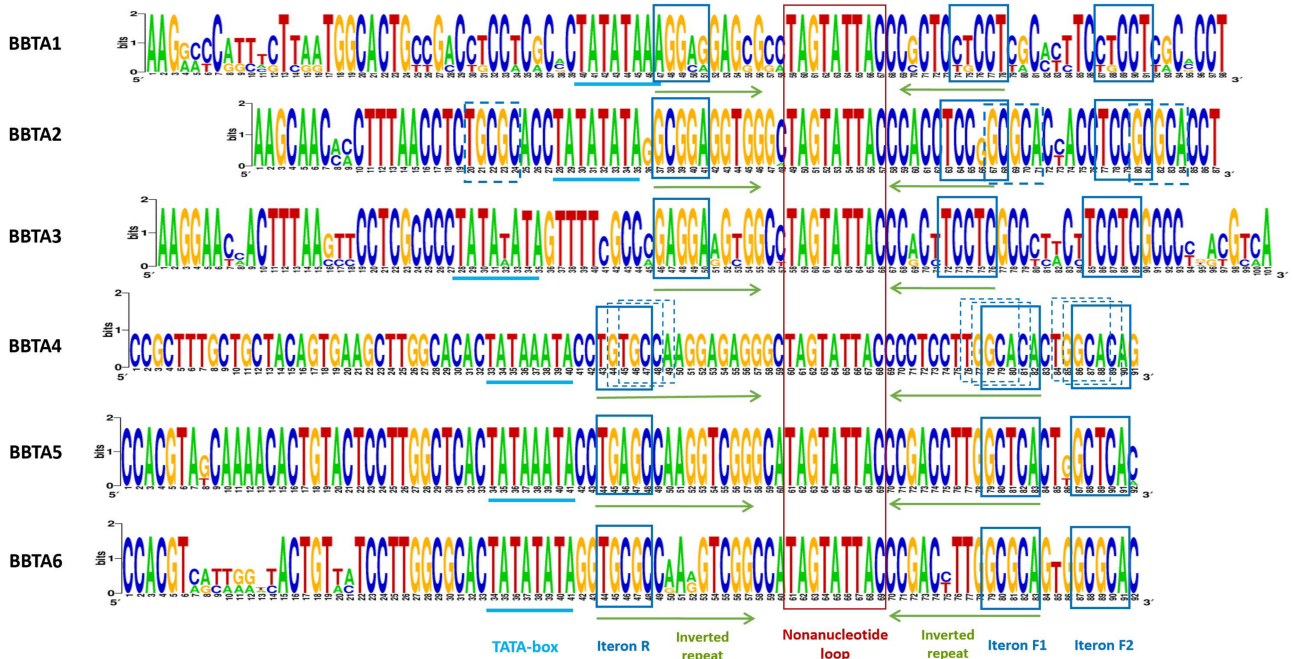


Figure 7. Logo of the CR-SL sequences of BBTV alphasatellites. The CR-SL sequences of the alphsatellites representing BBTa1, BBTa2, BBTa4, BBTa5, and BBTa6 species were aligned separately using SeaView with Muscle. Their consensus sequence (logo) was generated using WebLogo v2.8.2. The conserved sequence motifs/cis-acting elements are boxes or underlined and named.

3'-terminal nucleotide (ACTGG) in all isolates from Vietnam and China and four of the nine isolates from Laos (Supplementary Fig. S6A).

Inspection of the CR-SL sequences of BBTV alphsatellites revealed (i) the invariant nonnucleotide TAGTATTAC which is preserved in most alphsatellites (Stainton et al. 2017) and that differs at one position from the helper virus nonnucleotide (TATTATTAC), (ii) inverted repeats flanking the nonnucleotide sequence, (iii) putative 5-nt iterons R, F1, and F2, and (iv) the TATA-box of Pol II promoter driving transcription of Rep mRNA (Fig. 7). Interestingly, the inverted repeats in the banaphisatellites BBTa4, BBTa5, and BBTa6 are longer than those of the babusatellite BBTa1 and the muscarsatellites BBTa2 and BBTa3 (15 vs 10 nts, respectively). In all BBTV alphsatellites, the iterons R and F1 are parts of the inverted repeat sequences forming the stem-loop structure (Fig. 7). This is unlike the helper BBTV having all the three iterons outside of the stem-loop structure (Fig. 6). Note, however, that two of the three previously reported isolates (but not our twelve isolates) of BBTa2 have an insertion of G in the iteron F1 (TCGGGC) (Supplementary Fig. S7) and in these two isolates we identified alternative iterons F1' (GCGCA), F2' (GCGCA), and R' (TGCGC) (Fig. 7). Another notable exception to the within-species iteron identity is BBTa1 for which one of the five previously reported isolates has a different version of both the iteron R (AGGCA vs AGGAG) and iterons F1 and F2 (TGCCT vs CTCCT) (Supplementary Fig. S7). The alterations of two nucleotides in each of the three iterons (underlined) do not affect their nature as inverted (R) and direct (F1 and F2) repeats, thus supporting their functionality in alphsatellite replication. In the case of BBTa3, however, two of the eleven isolates (L32166 and U02312) have a deletion of iteron R and alterations in iterons F1 and F2 which do not appear to recreate three new iterons. Other alphsatellites have invariant iterons in all isolates within each species (Fig. 7; Supplementary Fig. S7). The iterons of three banaphisatellites

differ only at the 3rd nucleotide position and thus have the same consensus sequences of the iteron R (TGHGC, where H is T in BBTa4, A in BBTa5, or C in BBTa6) and iterons F1 and F2 (GCDCA, where D is A in BBTa4, T in BBTa5, or G in BBTa6). This indicates the conservation of Rep recognition and further supports their classification in one separate genus. Keeping in line with this notion, the muscarsatellites BBTa2 and BBTa3 also share the consensus sequences of iteron R (CMGGA, where M is C in BBTa2, or T in BBTa3) and iterons F1 and F2 (TCCKG, where K is G in BBTa2 or T in BBTa3), which differ from those of the babusatellite BBTa1 (Fig. 7).

Another notable difference between the CR-SL sequences of alphsatellites and helper BBTV is that the iterons F1 and F2 of alphsatellites are not adjacent to each other and are separated by 3 nts in all three banaphisatellite species or by 8 nts in the babusatellite BBTa1 and both muscarsatellite species (Fig. 7). This indicates differences in iteron recognition domains of alphsatellite Rep and helper virus M-Rep proteins. In further contrast to all six components of the helper BBTV genome, the TATA-box of Pol II promoter driving alphsatellite Rep transcription (as demonstrated for BBTa4 by Guyot et al. 2022) is located just upstream of the stem-loop structure in all BBTV alphsatellites, although the distance between these elements is not conserved, with the exception for all three banaphisatellite species (3 nts). Such location of the TATA-box implies that alphsatellite Rep binding to the origin of replication would interfere with Pol II transcription, thus providing a feedback loop to regulate Rep gene expression at the transcriptional level, as demonstrated for geminiviruses (Eagle, Orozco, and Hanley-Bowdoin 1994).

Our inspection of the CR-SL sequences of alphsatellites from *Fabenesatellite* and *Gosmusatellite* genera did not reveal any similarities to the iterons of banaphisatellites, but showed that the TATA-box with flanking CG-rich sequences comprising the core promoter are quite conserved between banaphisatellites and

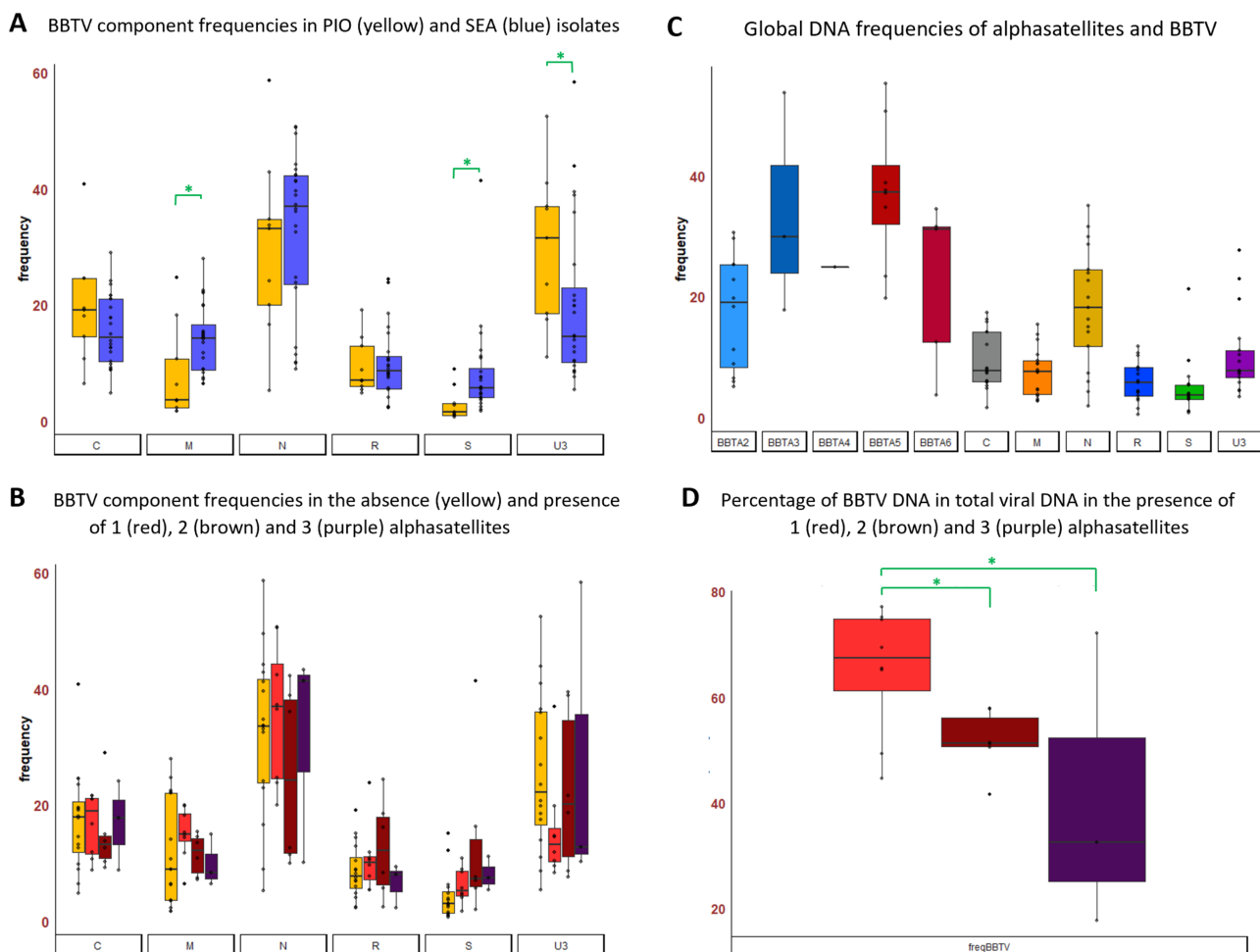


Figure 8. Relative abundance of BBTV genome components and alphasatellites. **(A)** BBTV genome component frequencies in PIO (yellow) and SEA (blue) isolates. **(B)** BBTV genome component frequencies in the absence (yellow) and presence of one (red), two (brown) and three (purple) alphasatellites. **(C)** Global DNA frequencies of alphasatellites and BBTV. **(D)** Percentage of BBTV DNA in total viral DNA in the presence of one (red), two (brown), and three (purple) alphasatellites. Statistical significance of the observed differences in median values was evaluated using a non-parametric Kruskal–Wallis test (Kruskal and Wallis 1952). * Kruskal–Wallis $P \leq 0.05$.

fabenesatellites (Supplementary Fig. S7B–E), further supporting their close evolutionary link.

Taking our findings together, BBTV M-Rep and alphasatellite Rep appear to have different modes of interactions with their respective iterons and the stem-loop structure within the origin of replication. By analogy with geminiviruses, co-variation in the iteron sequence and the iteron-binding domain of Rep protein would allow for sequence-specific recognition of the origin of replication by Rep. With the exception for a small proportion of isolates, members of each of the three genera of BBTV alphasatellites share the consensus iteron sequences, suggesting that alphasatellite Rep can potentially mediate trans-replication of other alphasatellites from the same genus. In contrast, none of the four BBTV alphasatellite genera shares iteron sequences with their helper BBTV, which explains why alphasatellite Rep cannot mediate trans-replication of helper virus components, as demonstrated for the muscarsatellite BBTA3 and BBTV from Taiwan (SEA) (Horser, Harding, and Dale 2001), and suggests that BBTV M-Rep may not be able to mediate trans-replication or directly interfere with autonomous replication (via CR-SL binding) of any BBTV alphasatellite and vice versa.

Competition between BBTV alphasatellites and their helper virus components for the host replication machinery

To further investigate the interactions between alphasatellites and helper BBTV, we compared relative abundance (frequency) of six BBTV components and alphasatellites by counting Illumina sequencing reads in our samples from PIO and SEA and then calculating the percentage of reads representing each viral component in a total number of viral (BBTV or BBTV + alphasatellite) reads. We have already calculated relative abundances for the BBTV components with and without banaphisatellite BBTA4 from DRC (PIO) using both quantitative (q)PCR analysis of total DNA extracted from infected banana leaves and by counting Illumina sequencing reads representing viral DNA amplified by RCA from the same total DNA (Guyot et al. 2022) and found that both methods give similar results. Another study comparing the relative abundances of genome components of Faba bean necrotic stunt virus (genus *Nanovirus*) by qPCR and Illumina sequencing showed some discrepancy between the two methods (Gallet et al. 2017).

Comparison of relative abundances of BBTV genome components in PIO ($n=9$) and SEA ($n=23$) isolates revealed differences

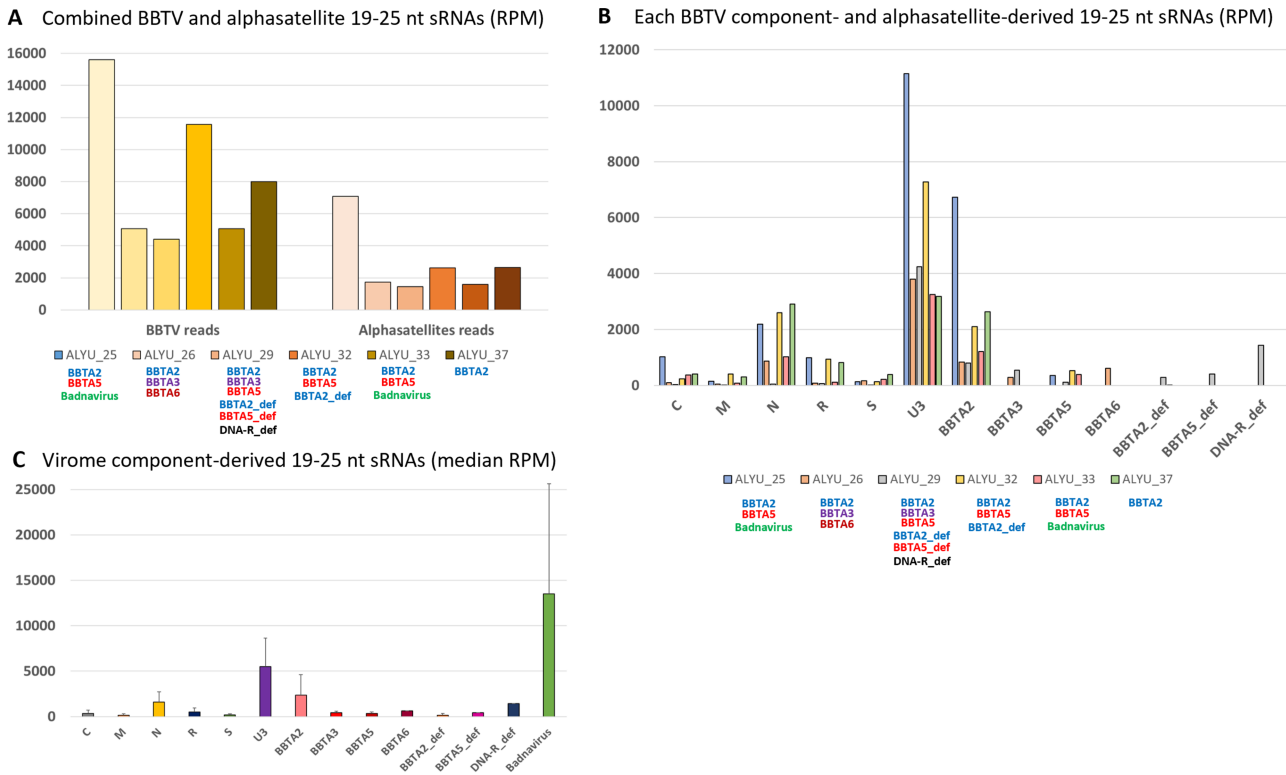


Figure 9. Counts of BBTV- and alphasatellite-derived 19–25 nt sRNAs in RPM. (A) Counts of combined BBTV and alphasatellite 19–25 nt sRNAs in reads per million of total 19–25 nt sRNAs (RPM). (B) Counts of each BBTV genome component- and each alphasatellite-derived 19–25 nt sRNAs in RPM. (C) Counts of virome component-derived 19–25 nt sRNAs in median RPM in all six samples. Names of the banana samples (ALYU) and their virome components—BBTV DNAs C, M, N, R, S and U3, BBTV alphasatellites, their defective (def) variants and badnaviruses—are indicated.

in accumulation of DNA-M and DNA-S, whose median levels were higher in SEA isolates (Kruskal–Wallis $P=0.02$ and $P=0.01$, respectively), and of DNA-U3 whose median level was lower in SEA isolates (Kruskal–Wallis $P=0.05$) (Fig. 8A). This is despite a big variation within each phylogenetic group (Supplementary Fig. S8). Further comparison of relative abundances of BBTV components in the isolates without alphasatellites and the isolates with one, two, or three alphasatellites did not reveal any statistically significant differences (Fig. 8B).

Notably, all five alphasatellites have a high median frequency that is comparable (BBTA2) or exceeding (BBTA3–6) the median frequency of DNA-N, one of the most abundant components of the helper virus in the presence of alphasatellites (Fig. 8C; Supplementary Fig. S9). The biggest variation was observed for the muscarsatellite BBTA2 whose relative abundance was high (and comparable to that of DNA-N) in the isolates where it was present alone ($n=3$), variable from high to low in the isolates with one additional alphasatellite ($n=6$), and intermediate in the isolates with two additional alphasatellites ($n=3$) (Supplementary Fig. S9B–D). Interestingly, the muscarsatellite BBTA3 that was always found together with two additional alphasatellites accumulated at comparable or higher levels than other alphasatellites in all three isolates (Supplementary Fig. S9D). All the banaphisatellites from PIO (BBTA4) and SEA (BBTA5 and BBTA6) accumulated at high frequency in all their single infections ($n=7$), either exceeding the frequency of any BBTV components (BBTA5 and BBTA6) or being comparable to the frequencies of DNA-N and DNA-U3 (BBTA4) (Supplementary Fig. S9B). Previously, the banaphisatellite BBTA4 associated with another BBTV isolate from DRC was also found to accumulate at high levels, comparable to those of

DNA-N and DNA-U3 (Guyot et al. 2022). In coinfections with one additional alphasatellite (always BBTA2) ($n=3$) the frequencies of banaphisatellites BBTA5 and BBTA6 were lower, compared to single infections, but still remained the highest among viral DNA components, with the exception for two isolates in which BBTA2 frequency was the highest (Supplementary Fig. S9C). Likewise, in coinfections with two alphasatellites (all three isolates with BBTA2 and BBTA3), the banaphisatellite BBTA5 (one isolate) had higher frequency than any BBTV component and BBTA2, but lower frequency than BBTA3, while BBTA6 had the highest frequency in one isolate and the lowest in another one (Supplementary Fig. S9C). A previous qPCR analysis of two SEA isolates of BBTV associated with one (BBTA2) or two (BBTA1 + BBTA2) alphasatellites has revealed that in a single infection the muscarsatellite BBTA2 accumulates at the highest level, followed by DNA-N, while in a double infection the babusatellite BBTA1 accumulates at the highest level, followed by BBTA2 and DNA-U3 (Yu et al. 2019).

Because in all our isolates one or more alphasatellites accumulated at high levels, the percentage of helper virus DNA in total viral DNA was found to range from 20 to 77 per cent: 45 to 77 per cent (median ca. 67 per cent) in the presence of one alphasatellite, 43 to 57 per cent (median ca. 50 per cent) in the presence of two alphasatellites and 20 to 70 per cent (median ca. 30 per cent) in the presence of three alphasatellites (Fig. 8D). This is consistent our previous study in which the percentage of helper virus DNA was ca. 65 per cent in the presence of one alphasatellite (BBTA4) (Guyot et al. 2022). In the latter study, we also found that the median loads of helper virus DNA (calculated by qPCR) was reduced by ca. 25 per cent in the presence of BBTA4. By extrapolation, the high abundance of alphasatellite DNA in both PIO and

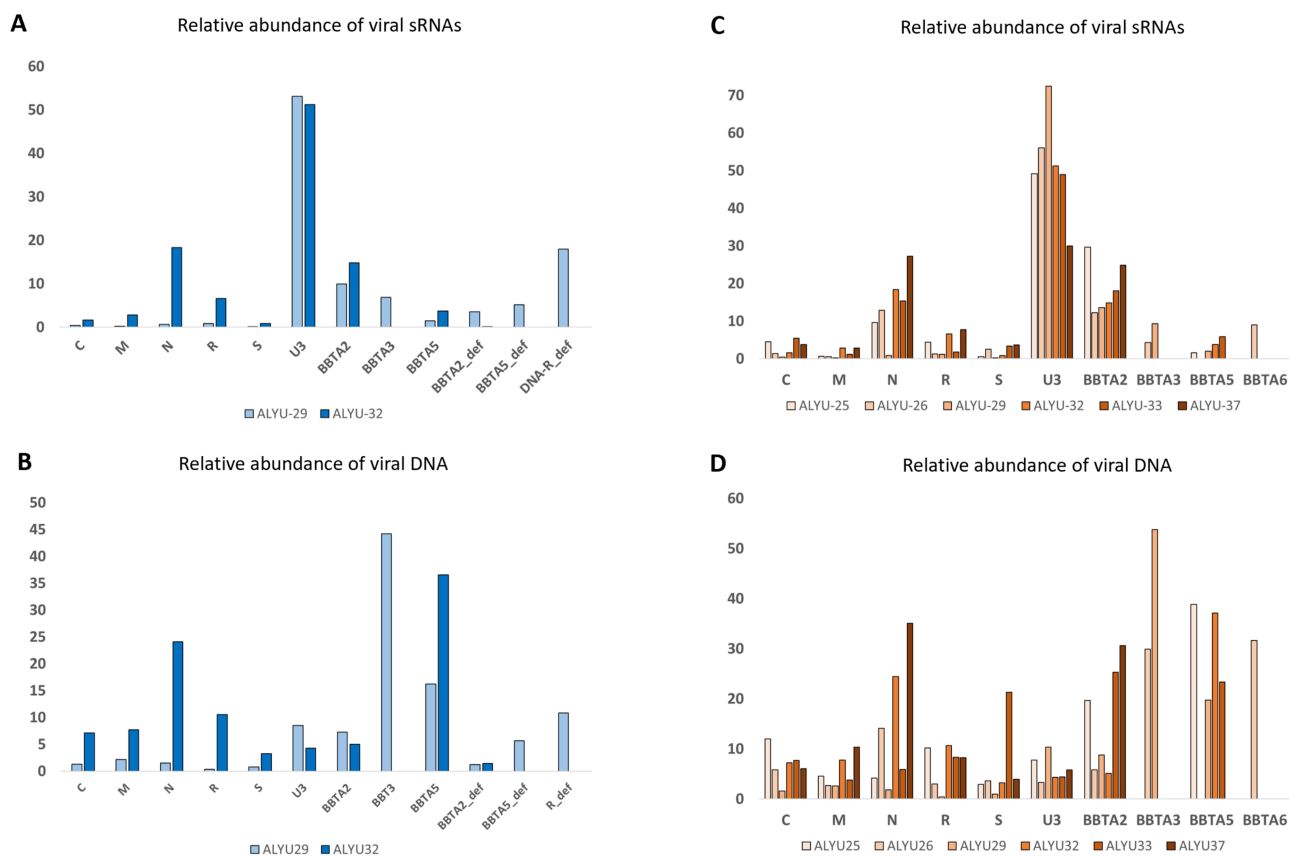


Figure 10. Comparison of viral sRNA and DNA relative abundances. Relative abundance of viral DNA and viral 19–25 nt sRNA reads representing each virome component—BBTV DNAs (C, M, N, R, S and U3), BBTV alphasatellites (BbTA2, BbTA3, BbTA5, BbTA6) and their defective (def) variants—present in each banana sample (ALYU) was calculated and plotted in percentage of total viral reads. Panels (A) and (C) show relative abundances of viral sRNAs, while panels (B) and (D) show relative abundances of viral DNA components.

SEA isolates would imply that all BBTV alphasatellites have the ability to compete with the helper virus for components of the host replication machinery, thereby reducing accumulation of helper BBTV in banana plants. Moreover, coinfections of BBTV with two or more alphasatellites would probably further reduce helper virus loads and likely interfere with its acquisition and transmission by banana aphids as we established for BbTA4 (Guyot et al. 2022).

Interactions of alphasatellites and helper BBTV with RNAi-based antiviral defences

To investigate the plant RNAi responses to co-infections with BBTV and its alphasatellites, we employed Illumina sequencing of small (s)RNA populations from six leaf samples collected in Vietnam and selected based on their complex viromes, each containing one (ALYU-37), two (ALYU-25, ALYU-32, ALYU-33), or three (ALYU-26, ALYU-29) alphasatellites. In addition, four of these samples contained either one badnavirus (ALYU-25, ALYU-33) or one and more defective viral molecules (ALYU-32 and ALYU-29, respectively). The resulting Illumina reads in a size-range from 19 to 25 nts were mapped to the reference sequences of the respective virome components. Reads representing each virome component or combinations thereof were then counted and normalized per million of total reads (RPM). As a result, the percentage of BBTV-derived sRNAs in the total (plant + viral) sRNA-ome was found to range from ca. 0.5 to 1.6 per cent (ca. 4500 to 15,500 RPM), while the percentage of combined alphasatellite-derived sRNAs ranged from ca. 0.2 to 0.7 per cent (ca. 1500 to 7000 RPM) (Fig. 9A). Notably, the ratio of BBTV- to alphasatellite-derived sRNAs was comparable

in all samples: alphasatellite sRNAs constituted 25–31 per cent of the total BBTV + alphasatellite sRNAs. This is consistent with our previous findings for banaphisatellite BbTA4 whose sRNAs constituted ca. 27 per cent of total BBTV + alphasatellite sRNAs (Guyot et al. 2022). The muscarsatellite BbTA2 present in all six samples was found to spawn the most abundant sRNAs, while sRNAs derived from the muscarsatellite BbTA3 present in two samples accumulated at lower levels. Likewise, sRNAs derived from the banaphisatellites BbTA5 (present in four samples) and BbTA6 (present in one sample) accumulated at lower levels, comparable to those of BbTA3 (Fig. 9B–C). Comparison of alphasatellite sRNA and DNA frequencies revealed an inverse correlation for BbTA3, BbTA5, and BbTA6 spawning less abundant sRNAs and accumulating their DNA at higher levels, with the exception for one of the four samples with BbTA5 (ALYU-33) where its DNA accumulated at a level comparable with that of BbTA2 (Fig. 10). Among BBTV components, DNA-U3 was found to spawn the most abundant sRNAs in all six samples, whose accumulation levels exceeded those of BbTA2-derived sRNAs. DNA-N spawned the second most abundant sRNAs in all samples except ALYU-29 where defective DNA-R was the second major source of viral sRNAs (Fig. 9B). Comparison of sRNA and DNA loads and frequencies revealed that despite being the best producer of viral sRNAs in all samples, DNA-U3 has the highest frequency only in the sample ALYU-29 where defective DNA-R becomes the second major source of viral sRNAs. DNA-N has the highest frequency in three of the five samples where it is the second best producer of viral sRNAs (Fig. 10; Supplementary Fig. S10).

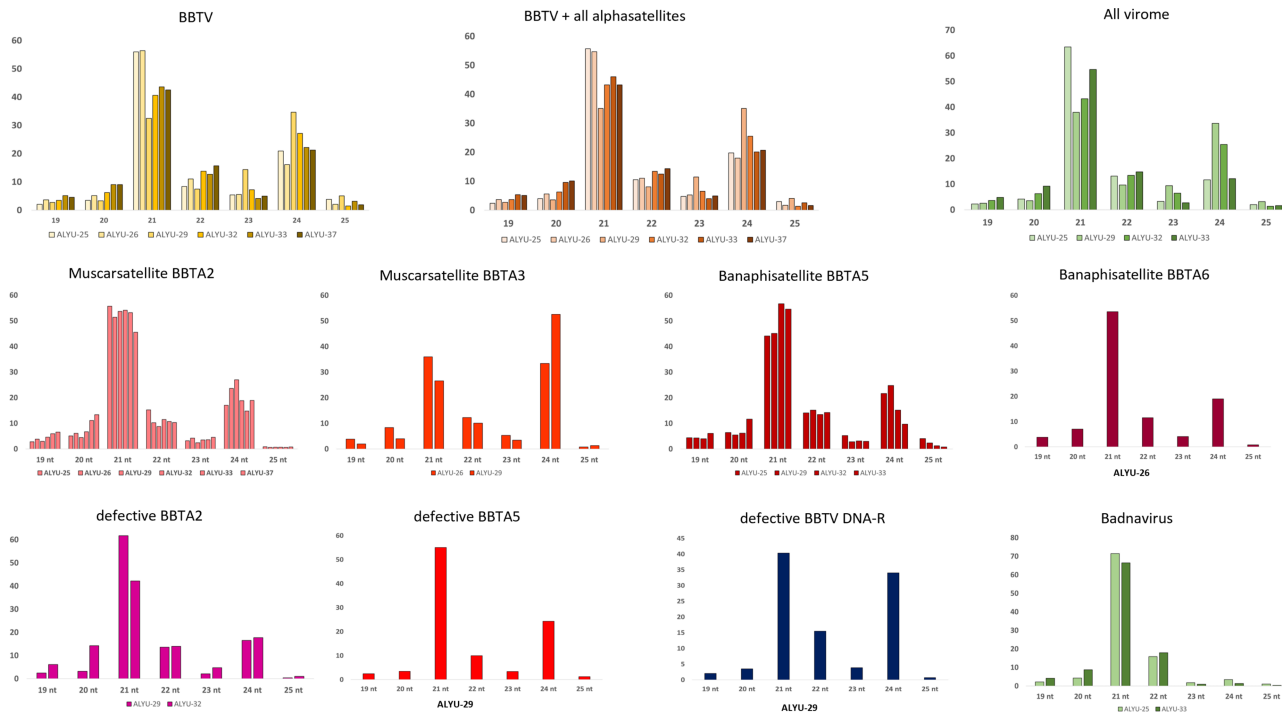


Figure 11. Size class profiles of viral siRNAs. Relative abundance of 19-, 20-, 21-, 22-, 23-, 24-, and 25-nt classes of viral sRNAs (size profile) derived from BBTV, alphasatellites, and other virome components were calculated and plotted in percentage of each size class in total 19–25 nt sRNA reads.

Interestingly, defective derivatives of BBTA2 and BBTA5 present together with defective DNA-R in ALYU-29 spawned sRNAs at low levels, which were about 2–3 times lower than the level of BBTA2 sRNAs, but about 2–3 times higher than the level of BBTA5 sRNAs in the same sample (Fig. 9B). While the sRNA frequency of the three defective molecules and non-defective BBTA2 and DNA-R resembled well the DNA frequency of these five virome components, non-defective BBTA5 being the weakest producer of sRNAs accumulated its DNA at higher levels than any of the latter components (Fig. 10A–B and Supplementary Fig. S10A), indicating its ability to evade RNAi.

Size class, 5′-terminal nucleotide identity and hot spot profiling of viral sRNAs revealed that all BBTV components, alphasatellites, and their defective variants spawn predominantly 21, 22, and 24 nt sRNAs derived from both sense and antisense strands, with sRNA hotspots and peaks being quite similar for each virome component across the samples, and that 21 and 22 nt viral sRNAs are enriched in 5′U and 5′C, while 24 nt viral sRNAs are enriched in 5′A (Fig. 11; Supplementary Fig. S11; Supplementary Datasets S3 and S4). This is consistent with our previous findings for BBTV and banaphisatellite BBTA4 (Guyot et al. 2022) and indicates that the banana homologs of Dicer-like (DCL) family proteins DCL4, DCL2, and DCL3 process viral dsRNA precursors (likely produced by sense and antisense transcription of entire viral DNA) into respectively 21, 22, and 24 nt siRNAs that get preferentially associated with the banana homologs of Argonaute (AGO) family proteins AGO1 (21–22 nt sRNAs with 5′U), AGO5 (21–22 nt sRNAs with 5′C), and AGO4 (24 nt sRNAs with 5′A) to silence BBTV and alphasatellite gene expression at both post-transcriptional and transcriptional levels, as established in model plants (reviewed in Pooggin 2018). Our findings for the defective derivatives of DNA-R, BBTA2, and BBTA5 suggest that these circular non-coding molecules also generate dsRNA precursors of viral siRNAs and may therefore

serve as decoys diverting the antiviral RNAi machinery from BBTV and alphasatellites.

Taken together, the main targets of 21, 22, and 24 nt siRNA-generating antiviral RNAi are BBTV components DNA-U3 and DNA-N and muscarsatellite BBTA2 (present in all six analysed samples) as well as defective DNA-R present in one sample. The muscarsatellite BBTA3 and banaphisatellites BBTA5 and BBTA6 appear to evade antiviral RNAi and thereby accumulate their DNA at higher levels than BBTA2.

Interestingly, the badnaviruses BSMIV and BSVNV were found to spawn siRNAs whose accumulation levels exceeded those of other virome components (Fig. 9B; Supplementary Datasets S3 and S4), likely because genome size differences (Supplementary Dataset S1A). In contrast to BBTV and alphasatellites, both badnaviruses spawn predominantly 21 and 22 nt siRNAs, while 24 nt siRNAs are underrepresented (Fig. 11). These results are consistent with our previous finding for six distinct badnaviruses in persistently infected banana plants, which spawned predominantly 21 and 22 nt siRNAs with the exception for two badnaviruses spawning 24 nt siRNAs at levels exceeding those of 22 nt (but not 21 nt) siRNAs (Rajeswaran et al. 2014). Thus, the banana homolog of DCL3 generating 24 nt siRNAs and thereby mediating transcriptional gene silencing as established in model plants (reviewed in Pooggin 2018) appears to have a more substantial contribution to the defence against BBTV and its alphasatellites which replicate in the nucleus, compared to badnaviruses which replicate in the cytoplasm via reverse transcription.

Conclusions and open questions

Our survey in Vietnam, Laos, and China revealed high prevalence and diversity of BBTV alphasatellites. We reconstructed complete genomes of new genetic variants of the muscarsatellites BBTA2 and BBTA3 (genus *Muscarsatellite*, subfamily *Petromoa-*

phasatellitinae) previously identified in SEA (Vietnam, China, and Taiwan) as well as several variants of two new alphasatellites—BBTA5 and BBTA6—which should be classified as two new species and which are most closely related to BBTA4, a newly emerging alphasatellite associated with BBTV in Congo-DRC (PIO) (Guyot et al. 2022). Based on comparative analysis of nucleotide and Rep protein sequences, we classified BBTV4, BBTA5, and BBTA6 as members of three species in a new genus of the subfamily Nanoalphasatellitinae—Banaphisatellite—and found that banaphisatellites are more related to fabenesatellites (genus *Fabenesatellite*, subfamily Nanoalphasatellitinae) and gosmusatellites (genus *Gosmusatellite*, subfamily Geminialphasatellitinae), both infecting dicot hosts. These findings suggest the provenance of banaphisatellites from a dicot-infecting progenitor, potentially associated with a helper virus from the genus *Nanovirus* of family Nanoviridae (comprising helper viruses of fabenesatellites) and/or the genus *Begomovirus* of family Geminiviridae (comprising helper viruses of gosmusatellites). Independent emergence of banaphisatellites in Africa (BBTA4) and in SEA (BBTA5 and BBTA6) indicates that alphasatellites in their evolution and adaptation to new hosts (dicot vs monocots) may change helper viruses of one family to those of another family.

In the past, BBTV isolates from Vietnam, China, and Taiwan were found to be associated only with muscarsatellites BBTA2 and BBTA3 and babusatellite BBTA1 (genus *Babusatellite*, subfamily Petromoalphasatellitinae) (Supplementary Table S1). This suggests that banaphisatellites BBTA5 and BBTA6 may have emerged in SEA only recently, similar to BBTA4 that was identified only in two isolates sampled in 2012 and 2016 in close vicinity from each other in Congo-DRC, but not in other isolates from that country or other regions of PIO (Stainton et al. 2015; Mukwa et al. 2016; Guyot et al. 2022). In contrast to BBTA4, both BBTA5 and BBTA6 are prevalent in Vietnam and present in its neighbour countries. Our analysis of conserved sequence motifs in the common regions driving replication and gene expression of alphasatellites and their helper BBTV strains from PIO and SEA revealed that BBTA4 (the only satellite from PIO) does not possess CR-M, while the CR-M sequences are well conserved in all other BBTV alphasatellite species. This finding raises a question on the priming mechanism of complementary DNA synthesis in BBTA4 and suggests that following a jump from a putative dicot host, BBTA4 may not have had enough time to co-evolve with its monocot-infecting helper BBTV in Congo-DRC. The high prevalence of banaphisatellites BBTA5 and BBTA6 in SEA would suggest their better adaptation to helper BBTV. In contrast to CR-M, the CR-SL sequences are well conserved between the three species of banaphisatellites that share high sequence identity not only in the stem-loop structure but also in the iterons and the Pol II promoter elements. This implies conserved mechanisms of the initiation of rolling-circle replication by banaphisatellite Rep proteins and of the Pol II transcription of Rep-mRNA, pointing at the potential for trans-replication between alphasatellite species within this genus and highlighting their evolutionary links and common origin. Likewise, the muscarsatellites BBTA2 and BBTA3 share all the conserved CR-SL elements and other sequence features between each other but not with babusatellite BBTA1 or banaphisatellites.

Our analysis of the relative abundances of alphasatellites and helper BBTV components revealed that alphasatellite DNA accumulates at high levels in all plants, constituting 23–80 per cent of total viral DNA (BBTV + alphasatellite). This indicates that autonomously replicating alphasatellites, being unable to mediate trans-replication of helper virus components (as supported by our comparative analysis of CR-SL sequences), compete with the

helper virus for some limiting component(s) of the host replication machinery and likely reduce the helper virus loads, as established for BBTA4 (Guyot et al. 2022). The reduced accumulation of all components of the helper virus would reduce its virulence and thereby increase a lifespan of the BBTV and alphasatellite-infected host plants and in turn increase the chances for both BBTV and alphasatellite to be transmitted to new plants by banana aphids, even in the case of reduced virus acquisition and transmission rates, as was shown for BBTA4 under laboratory conditions (Guyot et al. 2022). Another potential benefit for the helper virus would be the ability of alphasatellite to divert components of the antiviral RNAi machinery generating 21, 22, and 24 nt viral siRNAs that can potentially interfere with viral gene expression at both transcriptional and posttranscriptional levels as demonstrated for BBTA4 (Guyot et al. 2022). Our analysis of siRNA profiles in six selected plant samples from SEA confirmed the previous finding for BBTA4 in that alphasatellite-derived siRNAs of the three major size classes accumulate at high levels, constituting a substantial fraction in total viral (BBTV + alphasatellite) siRNAs. Interestingly, in the plants co-infected with two or more alphasatellites only one of the satellites was a strong producer of viral siRNAs, while the others spawned less abundant siRNAs and accumulated their DNA at higher levels, suggesting evasion of RNAi. The mechanism of RNAi evasion by alphasatellites and their helper viruses remains to be further investigated.

In summary, our findings shed a new light on the provenance of alphasatellites, their co-evolution with helper ssDNA viruses and potential mutual benefits of alphasatellite-helper virus association.

Materials and methods

Surveys in Vietnam, Laos, and China

Banana leaf samples were collected during surveys in Vietnam (2018) and in Laos and China (2019) and locally dried using silica gel.

Total DNA extraction from banana leaf samples

Dried leaf tissue (100 mg) was ground in liquid nitrogen and 500 μ l extraction buffer [100 mM Tris-HCl pH 8.0, 1.4 M NaCl, 20 mM EDTA, 2 per cent alkyltrimethylammonium bromide (MATAB), 1 per cent polyethyleneglycol 6000, 0.5 per cent sodium sulfite] pre-heated at 74°C and supplemented with 0.4 μ l RNase (100 mg/ml) was added to the frozen powder. The mixture was vortexed for 20 s, incubated at 74°C for 20 min and then mixed vigorously with one volume of chloroform-isoamyl alcohol (24:1 v/v) (CIAA), followed by centrifugation at 13,000 rpm for 30 min at 4°C. The supernatant was taken for a second round of extraction with CIAA followed by centrifugation as described above. The supernatant was mixed with one volume of isopropanol pre-cooled at –20°C. The mixture was shaken until appearance of a hank and then spun at 13,000 rpm and 4°C for 30 min. The pellet was washed twice with 500 μ l of 70 per cent ethanol, air dried, and dissolved in 100 μ l of Milli-Q water.

RCA of viral DNA

Circular viral DNA components were amplified by RCA using a TempliPhi RCA kit (GE Healthcare) following the manufacturer protocol. Briefly, 5 μ l sample buffer and 1 μ l total DNA extracted from banana leaf tissues were mixed and heated at 95°C for 3 min. The samples were cooled and 5 μ l reaction buffer and 0.2 μ l enzyme mix were added, followed by incubation at 30°C for 18 h. The enzyme was inactivated by heating at 65°C for 10 min.

Illumina sequencing of RCA products and *de novo* reconstruction of viral genomes

Fifty nanograms of the cleaned undigested RCA products were taken for Illumina sequencing at Fasteris AG (www.fasteris.com). Libraries were prepared using the Nextera XT standard DNA protocol and all libraries were multiplexed and sequenced in one flow cell of HiSeq2500 with a 2× 125-nt paired-end run. Viral genomes were *de novo* reconstructed from the sequencing reads of each library by selecting unique inserts sequenced ≥ 5 , ≥ 10 , ≥ 20 , ≥ 30 , ≥ 40 , or ≥ 50 times and assembling them using Velvet v. 1.2.10 (Zerbino and Birney 2008) with k-mers 77, 79, 83, 87, 91, 95, 99, 103, 107, 111, 113, and 117. All the resulting Velvet contigs were scaffolded using SeqMan Pro v. 7.1.0 (DNASTAR Lasergene). SeqMan contigs of viral origin were identified by BLASTn analysis. The consensus viral genome sequences were verified using SeqMan scaffolds and validated by mapping back the Illumina reads using Burrow–Wheeler Aligner (BWA) 0.7.12 (Burrows and Wheeler 1994) and visualisation using MISIS-2 (Seguin et al. 2016).

Total RNA extraction from banana leaves

RNA extraction was performed using a CTAB-LiCl method as described previously (Golyaev, 2019). Briefly, 0.5 g dried leaf tissue was ground in liquid nitrogen and mixed with 5 mL extraction buffer [300 mM Tris-HCl (pH 7.5), 2 M NaCl, 25 mM EDTA (pH 8.0), 2 per cent CTAB, 2 per cent PVP, and 2 per cent beta-mercaptoethanol]. The mixture was incubated at 65°C for 10 min. Then, 5 mL of CIAA was added and the mixture was centrifuged at 8000 rpm for 10 min at 4°C. The supernatant was mixed vigorously with an equal volume of CIAA, followed by centrifugation at 13,000 rpm for 10 min at 4°C. The supernatant was mixed with 0.6 volume of isopropanol and 0.1 volume of 3 M sodium acetate (pH 5.2) and incubated at –80°C for 30 min, followed by centrifugation at 13,000 rpm for 30 min at 4°C. The pellet was dissolved in 1 mL of RNase free water, 0.3 volume of 10 M LiCl was added, and the mixture was incubated overnight at 4°C, followed by centrifugation at 13,000 rpm for 30 min at 4°C. The pellet was resuspended in 100 μ L of RNase free water, followed by the addition of 0.1 volume 3 metre sodium acetate (pH 5.2) and 2 volume of cold absolute ethanol. The mixture was immediately centrifuged at 13,000 rpm for 20 min at 4°C. The pellet was washed with 200 μ L of 75 per cent ethanol, followed with centrifugation at 13,000 rpm for 10 min at 4°C. The pellet was air-dried and dissolved in 50 μ L RNase-free water.

Illumina sequencing and bioinformatics analysis of viral small RNAome

Integrity of total RNA extracted from banana leaves was verified by capillary electrophoresis on LabChip GX (Perkin Elmer). Illumina sequencing was performed at Fasteris AG (www.fasteris.com) using the Illumina TruSeq small RNA protocol (acrylamide gel size selection of the 18–30 nt range, single strand ligation of the 3' adapter, single strand ligation of the 5' adapter, cDNA synthesis by reverse transcription, amplification of the library). The libraries were multiplexed and sequenced in one flow cell of NovaSeq 6000. Following de-multiplexing and adapter trimming, the resulting reads from each library were mapped using BWA onto the reference sequences of alphasatellites, BBTv, and other components of the virome reconstructed by Illumina sequencing of RCA amplified viral DNA (as described earlier). Mapped viral reads were sorted by polarity (forward, reverse), size (15–34 nts), and 5'-terminal nucleotide identity (5'U, 5'A, 5'G, 5'C) and counted, followed by normalization in RPM of total (viral + non-viral) reads (Supplementary Dataset S3). Relative abundances of viral sRNAs and size

classes of siRNAs from six BBTv components and alphasatellites as well as other virome components were calculated as percentage of reads representing each component in total BBTv and total BBTv + alphasatellite and total virome reads, and the mean values with standard deviations were calculated.

Nucleotide and protein sequence analyses of alphasatellites and helper viruses

Reference sequences of all isolates of alphasatellites available in the NCBI Genbank in February 2022 were downloaded, adjusted to begin with the conserved nonanucleotide sequence TAGTATTAC and aligned to the reconstructed sequences of new alphasatellites using SDT v1.2 with Muscle (Muhire, Varsani, and Martin 2014) to calculate their pairwise identities. A maximum likelihood phylogenetic tree was constructed using Seaview (Gouy, Guindon, and Gascuel 2010) with Muscle and FigTree v1.4.4 (<https://github.com/rambaut/figtree/releases>). Clustering analysis of alphasatellite Rep protein sequences was performed using CLANS (Frickey and Lupas 2004). Since the majority of predicted alphasatellite ORFs encode Rep proteins ranging in size from ca. 290 (Nanoalphasatellitinae) to ca. 315 (Geminialphasatellitinae) amino acids (aa), we excluded from this analysis the following alphasatellites with truncated ORFs: BBTa1 isolate EU366174 (164 aa), BBTv3 isolates U02312 (102 aa) and EU366175 (188 aa), tomato leaf curl alphasatellite isolate KY420167 (216 aa), and okra enation leaf curl alphasatellite isolate HF546575 (217 aa).

Pairwise sequence comparison of the reconstructed sequences of each component of BBTv genome of our SEA and PIO isolates was also performed using SDT v1.2.

CR-SL and CR-M sequences of each BBTv core component from SEA and PIO isolates and of alphasatellites were aligned using SeaView with Muscle. A consensus sequence (logo) was generated using WebLogo v2.8.2 (Crooks et al. 2004).

Data availability

Viral genome sequences were deposited in the NCBI Genbank under the accession numbers ON934232–ON934284 (BBTv isolates ALYU14–24 from PIO), ON959832–ON959993 (BBTv isolates ALYU25–56 from SEA), ON959994–ON960032 (BBTv alphasatellites from SEA), and ON960033–ON960034 (badnaviruses BSVNV and BSMIV).

Supplementary data

Supplementary data is available at Virus Evolution Journal online.

Acknowledgements

We thank Sebastien Ravel for bioinformatics support and Nathalie Laboureau for laboratory management.

Funding

The work was supported by Agropolis Foundation (Montpellier) under the reference ID 1605–011 through the 'Investissements d'avenir' programme (Labex Agro: ANR-10-LABX-0001-01) and the Open Science Call 2016 funds (project BforBB: BSV for banana biodiversity) to M.-L.I.-C. and project partners, the INRAE SPE department funds (project ViroMix) to M.M.P., and a PhD scholarship from Institute Agro (Montpellier) to V.G.

Conflict of interest: The authors declare no competing interests.

References

- Beetham, P. R., Harding, R. M., and Dale, J. L. (1999) 'Banana Bunchy Top Virus DNA-2 to 6 Are Monocistronic', *Archives of Virology*, 144: 89–105.
- Briddon, R. W. et al. (2018) 'Alphasatellitidae: A New Family with Two Subfamilies for the Classification of Geminivirus- and Nanovirus-associated Alphasatellites', *Archives of Virology*, 163: 2587–600.
- Briddon, R. W., and Stanley, J. (2006) 'Subviral Agents Associated with Plant Single-stranded DNA Viruses', *Virology*, 344: 198–210.
- Burns, T. M., Harding, R. M., and Dale, J. L. (1995) 'The Genome Organization of Banana Bunchy Top Virus: Analysis of Six ssDNA Components', *Journal of General Virology*, 76: 1471–82.
- Burrows, M., and Wheeler, D. J. (1994) 'A Block-Sorting Lossless Data Compression Algorithm', Technical Report 124, Digital Equipment Corporation, Digital Equipment Corporation Press: Palo Alto, CA, USA, 1–24.
- Chen, X. et al. (2010) 'SQUAMOSA Promoter-binding Protein-like Transcription Factors: Star Players for Plant Growth and Development', *Journal of Integrative Plant Biology*, 52: 946–51.
- Crooks, G. E. et al. (2004) 'WebLogo: A Sequence Logo Generator', *Genome Research*, 14: 1188–90.
- Dale, J. L. (1987) 'Banana Bunchy Top: An Economically Important Tropical Plant Virus Disease', *Advances in Virus Research*, 33: 301–25.
- Eagle, P. A., Orozco, B. M., and Hanley-Bowdoin, L. (1994) 'A DNA Sequence Required for Geminivirus Replication Also Mediates Transcriptional Regulation', *The Plant Cell*, 6: 1157–70.
- Frickey, T., and Lupas, A. (2004) 'CLANS: A Java Application for Visualizing Protein Families Based on Pairwise Similarity', *Bioinformatics*, 20: 3702–4.
- Gallet, R. et al. (2017) 'The Number of Target Molecules of the Amplification Step Limits Accuracy and Sensitivity in Ultradeep-Sequencing Viral Population Studies', *Journal of Virology*, 91: 561–78.
- Golyaev, V. et al. (2019) 'Plant virome reconstruction and antiviral RNAi characterization by deep sequencing of small RNAs from dried leaves', *Scientific Reports*, 9: 19268.
- Gouy, M., Guindon, S., and Gascuel, O. (2010) 'SeaView Version 4: A Multiplatform Graphical User Interface for Sequence Alignment and Phylogenetic Tree Building', *Molecular Biology and Evolution*, 27: 221–4.
- Guyot, V. et al. (2022) 'A Newly Emerging Alphasatellite Affects Banana Bunchy Top Virus Replication, Transcription, siRNA Production and Transmission by Aphids', *PLoS Pathogens*, 18: e1010448.
- Hafner, G. J. et al. (1997) 'Nicking and Joining Activity of Banana Bunchy Top Virus Replication Protein in Vitro', *Journal of General Virology*, 78: 1795–9.
- Hafner, G. J., Harding, R. M., and Dale, J. L. (1997) 'A DNA Primer Associated with Banana Bunchy Top Virus', *Journal of General Virology*, 78: 479–86.
- Herrera-Valencia, V. A. et al. (2006) 'An Iterated Sequence in the Genome of Banana Bunchy Top Virus Is Essential for Efficient Replication', *Journal of General Virology*, 87: 3409–12.
- Herrera-Valencia, V. A. et al. (2007) 'Mapping the 5' Ends of Banana Bunchy Top Virus Gene Transcripts', *Archives of Virology*, 152: 615–20.
- Horser, C. L., Harding, R. M., and Dale, J. L. (2001) 'Banana Bunchy Top Nanovirus DNA-1 Encodes the 'Master' Replication Initiation Protein', *Journal of General Virology*, 82: 459–64.
- Ji, X. L. et al. (2019) 'Banana Bunchy Top Virus (BBTV) Nuclear Shuttle Protein Interacts and Re-distributes BBTV Coat Protein in *Nicotiana Benthamiana*', *3 Biotech*, 9: 121.
- Karan, M., Harding, R. M., and Dale, J. L. (1994) 'Evidence for Two Groups of Banana Bunchy Top Virus Isolates', *Journal of General Virology*, 75: 3541–6.
- Kruskal, W. H., and Wallis, W. A. (1952) 'Use of Ranks in One-Criterion Variance Analysis', *Journal of the American Statistical Association*, 47: 583–621.
- Kumar, P. L. et al. (2015) 'Biology, Etiology, and Control of Virus Diseases of Banana and Plantain', *Advances in Virus Research*, 91: 229–69.
- Lageix, S. et al. (2007) 'The Nanovirus-Encoded Clink Protein Affects Plant Cell Cycle Regulation through Interaction with the Retinoblastoma-Related Protein', *Journal of Virology*, 81: 4177–85.
- Magee, C. J. P. (1940) 'Transmission Studies on the Banana Bunchy-top Virus', *Journal of the Australian Institute of Agricultural Science*, 6: 109–10.
- Mansourpour, M. et al. (2022) 'Effects of an Alphasatellite on the Life Cycle of the Nanovirus Faba Bean Necrotic Yellows Virus', *Journal of Virology*, 96: e0138821.
- Muhire, B. M., Varsani, A., and Martin, D. P. (2014) 'SDT: A Virus Classification Tool Based on Pairwise Sequence Alignment and Identity Calculation', *PLoS One*, 9: e108277.
- Mukwa, L. F. et al. (2016) 'Low Genetic Diversity of Banana Bunchy Top Virus, with a Sub-regional Pattern of Variation, in Democratic Republic of Congo', *Virus Genes*, 52: 900–5.
- Pooggin, M. M. (2018) 'Small RNA-omics for Plant Virus Identification, Virome Reconstruction, and Antiviral Defense Characterization', *Frontiers in Microbiology*, 9: 2779.
- Qazi, J. (2016) 'Banana Bunchy Top Virus and the Bunchy Top Disease', *Journal of General Plant Pathology*, 82: 2–11.
- Rajeswaran, R. et al. (2014) 'Evasion of Short Interfering RNA-directed Antiviral Silencing in *Musa Acuminata* Persistently Infected with Six Distinct Banana Streak Pararetroviruses', *Journal of Virology*, 88: 11516–28.
- Seguin, J. et al. (2016) 'MISIS-2: A Bioinformatics Tool for In-depth Analysis of Small RNAs and Representation of Consensus Master Genome in Viral Quasispecies', *Journal of Virological Methods*, 233: 37–40.
- Stainton, D. et al. (2015) 'The Global Distribution of Banana Bunchy Top Virus Reveals Little Evidence for Frequent Recent, Human-mediated Long Distance Dispersal Events', *Virus Evolution*, 1: 9.
- Stainton, D. et al. (2017) 'Comparative Analysis of Common Regions Found in Babuviruses and Alphasatellite Molecules', *Archives of Virology*, 162: 849–55.
- Varsani, A. et al. (2021) 'Taxonomy Update for the Family Alphasatellitidae: New Subfamily, Genera, and Species', *Archives of Virology*, 166: 3503–11.
- Wanitchakorn, R. et al. (2000) 'Functional Analysis of Proteins Encoded by Banana Bunchy Top Virus DNA-4 to -6', *Journal of General Virology*, 81: 299–306.
- Wanitchakorn, R., Harding, R. M., and Dale, J. L. (1997) 'Banana Bunchy Top Virus DNA-3 Encodes the Viral Coat Protein', *Archives of Virology*, 142: 1673–80.
- Watanabe, S., Greenwell A. M., and Bressan, A. (2013) 'Localization, concentration, and transmission efficiency of Banana bunchy top

- virus in four asexual lineages of *Pentalonia* aphids', *Viruses*, 5: 758–76.
- Yu, N. T. et al. (2019) 'Independent Modulation of Individual Genomic Component Transcription and a Cis-acting Element Related to High Transcriptional Activity in a Multipartite DNA Virus', *BMC Genomics*, 20: 1–13.
- Zerbino, D. R., and Birney, E. (2008) 'Velvet: Algorithms for de Novo Short Read Assembly Using de Bruijn Graphs', *Genome Research*, 18: 821–9.
- Zhu, H. et al. (2019) 'Banana sRNAome and Degradome Identify microRNAs Functioning in Differential Responses to Temperature Stress', *BMC Genomics*, 20: 33.

Virus Evolution, 2024, **10(1)**, 1–18

DOI: <https://doi.org/10.1093/ve/vead076>

Advance Access Publication 15 December 2023

Research Article

© The Author(s) 2024. Published by Oxford University Press.

This is an Open Access article distributed under the terms of the Creative Commons Attribution-NonCommercial License (<https://creativecommons.org/licenses/by-nc/4.0/>), which permits non-commercial re-use, distribution, and reproduction in any medium, provided the original work is properly cited. For commercial re-use, please contact journals.permissions@oup.com