THÈSE POUR OBTENIR LE GRADE DE DOCTEUR DE MONTPELLIER SUPAGRO

En Biologie des systèmes

École doctorale GAIA – Biodiversité, Agriculture, Alimentation, Environnement, Terre, Eau

PHIM - Plant Health Institute of Montpellier

Interactions moléculaires du banana bunchy top virus et de son alphasatellite avec ses plantes hôtes et les pucerons vecteurs

Présentée par Valentin GUYOT
Le 23 Juin 2022

Sous la direction de Mikhail POOGGIN et Marie-Line ISKRA-CARUANA

Devant le jury composé de

Mme Tatiana TIMCHENKO, Chargée de recherche, CNRS	**Rapporteure**
M. Martin DRUCKER, Directeur de recherche, INRAE	**Rapporteur et Président du jury**
Mme Anne-Sophie GOSSELIN-GRENET, maître de conférences, Université de Montpellier	**Examinatrice**
M. Mikhail POOGGIN, Directeur de recherche, INRAE	**Directeur de thèse**
M. Romain GALLET, Chargé de recherche, INRAE	**Invité**
Mme Marie-Line ISKRA-CARUANA, Directrice de recherche, CIRAD	**Invitée**

UNIVERSITÉ DE MONTPELLIER

L'INSTITUT agro Montpellier

"The eternal mystery of the world is its comprehensibility ...
The fact that it is comprehensible is a miracle."

Albert Einstein, 1936

Remerciements

Enfin voici venue cette partie importante que je vais écrire en français. Sur ce, excusez-moi pour l'ensemble des fautes d'orthographes qui vont suivre, je n'ai malheureusement pas réalisé un parcours scientifique à l'aide de mes capacités linguistiques. Je ne sais pas si je pouvais remercier l'ensemble des acteurs ayant permis la réalisation de ma thèse tant bien ils sont nombreux. J'essaierais donc de faire au mieux !

Avant de commencer cette partie des remerciements qui sont un ressentiment personnel, je tiens à remercier Tatiana/Tania Timtchenko et Martin Drucker qui ont tous les deux accepté d'être rapporteurs de ma thèse ainsi que juristes de celle-ci. Je vous souhaite une bonne lecture ! Je tiens également à remercier Anne-Sophie Gosselin-Grenet et Romain Gallet qui ont tous les deux accepté de faire partie de mon jury de thèse. Enfin, je tiens également à remercier Véronique Brault et Christophe Brugidou qui ont accepté durant ces trois dernières années de participer à mes comités de thèse afin de valider mes travaux de thèses et de m'aider à continuer dans la bonne direction.

Tout d'abord je tiens à remercier l'acteur principal de ma thèse et directeur de thèse Misha (Mikhail Pooggin). Sans lui tous ces travaux n'auraient tout simplement pas vu le jour. Je le remercie pour son optimisme et son expérience scientifique qui a souvent contrebalancer mon pessimisme naturel, mais également le fait qu'il m'a toujours encouragé, sans jamais avoir la moindre remarque négative à mon égard, alors que pourtant, il en avait l'occasion. Je pense que c'est la bonne façon d'encadrer un étudiant en thèse et je pense être chanceux d'avoir fait ma thèse avec Misha, qui plus est, est une thèse différente (de « type anglo-saxon ») des thèses classiques françaises. Ces quatre années de thèse m'ont permis d'acquérir une autonomie presque totale dans mes recherches, mes expérimentations et dans mon organisation du travail… Merci.

Je voudrais également remercier Marie-Line Iskra Caruana, co-directrice de thèse. Bien qu'elle n'ait pas pu être présente en seconde partie de thèse, elle a rapidement compris que je pouvais me débrouiller seul. Je la remercie pour m'avoir également appris à devenir autonome

mais également pour la force dont elle a fait preuve pour maintenir l'équipe 1 pendant tant d'années. Bien que je n'étais pas là depuis le début, quand on fait partie de l'équipe 1 on comprend vite la difficulté qui l'accompagne pour maintenir une certaine cohésion, et cela demande une force de caractère que, je crois bien, seule Marie-Line possède.

Pour continuer avec l'équipe 1, maintenant devenue defensiRNA, je voudrais remercier Nathalie Laboureau pour m'avoir formé aux règles de notre laboratoire en Master 2, mais de manière générale pour sa compassion envers notre environnement ; et je remercie Serge Galzi pour avoir su prendre soin de notre laboratoire et de nos bananiers autant de fois qu'il le pouvait. Je terminerai avec des remerciements pour Rajendran Rajeswaran (Rajesh). Bien que je souhaitais faire ma thèse avec sa présence dans mon encadrement, il a dû malheureusement partir en cours de route. Toutefois, il a toujours été un moteur dans mes expérimentations, surnommé Golden hands, je me suis toujours inspiré de lui afin de faire des manips toujours plus propres et qualitatives. Durant mon stage de Master 2, il a su être d'une aide inoubliable.

Au sein de l'UMR BGPI devenu l'UMR PHIM, je tiens à remercier énormément de personnes, mais j'aimerais d'abord commencer tout particulièrement par mes amis serristes : Romain Ferdinand, Axel Cremades et Remy Habas qui m'ont conseillé tout au long de ma gestion de la cellule humide 2 et des bananiers infectés par le BBTV. Avec un gros remerciement à Remy qui a passé plus de temps à soigner notre cellule en serre K1 contre les thrips que notre équipe elle-même.

Je tiens maintenant à remercier Agnès Delaunay pour sa gentillesse, pour ses cannelés, et tout ce qui l'accompagne ! Ainsi que pour la correction des fautes d'orthographes de mon résumé français ! Je remercie également Dominique Lagrenee pour sa patience et pour m'avoir laissé installer des plantes dans le bâtiment ! Enfin, pour terminer avec l'administration, comment ne pas remercier Michel Amphoux ? Ce grand malade au cœur tendre qui nous a fait découvrir la région de l'Hérault, très bon guide touristique si vous voulez mon avis.

Afin de gagner du temps et de l'espace parce que ce n'est pas non plus l'objectif de ma thèse, je vais continuer à vous remercier en citant vos noms ainsi que quelques mots clés : Claire Neema parce que c'est la famille, Aurélie Ducasse noisettes pour le plaisir, Hilaire Kabore pour la force, Pascal Alonzo la chance, Charlotte Julian de la manade d'Alain, Henri Adreit pour Madagascar, Laurence Blondin et Sonia Guillou pour leur compassion, Jean Carlier pour avoir accepté Thomas en thèse et l'avoir prolonger à PHIM, Stéphane Cociancich pour accepter de

faire les tâches ingrates (rappeler les règles de sécurité), Marlène Souquet parce qu'elle, elle crie tout le temps, Michel Yvon parce que c'est un gars sûr ! Michel Ducamp pour les cacaoyers. Et bien d'autres gens.

Je remercie l'ensemble des chercheurs en virologie pour m'avoir apporté leur aide, merci Philippe Roumaniac de m'avoir montré que, même dans la défaite, on pouvait garder le sourire. Big up à Stéphane Blanc, assurément le meilleur chercheur de PHIM (vous-même vous le savez). Ce fût ainsi un objectif de ma thèse de le battre, peu importe le domaine, mais il faut bien reconnaître que je ne suis pas satisfait du résultat. Je n'ai jamais vraiment pu le battre, même quand la victoire était à portée. Il a préféré chercher le nul, comme un joueur d'échec forçant le pat. Toutefois ce n'est que partie remise, car je crois bien que je me relèverais toujours de mes défaites avec de nouveaux plans maléfiques. Au passage, je cite Yannis Michalakis qui a rejoint la partie en cours de route mais qui a préféré s'allier à Stéphane, triste est le sort qui l'attend.

Enfin, pour en finir avec PHIM, je remercie l'ensemble des étudiants, doctorants, post-doctorants, ingénieurs et techniciens contractuels ou non qui m'ont supporté au quotidien. Tout particulièrement Maud Thierry et Emeline Riccuitiniti avec qui je partageais le bureau au début de ma thèse, triste et heureuse fût la fin de leur thèse en ma compagnie. Maëlle Deshoux et Kevin Gauthier pour leur encouragement. Merci à Jeremy Di Mattia pour toutes nos conversations scientifiques et sa bonne humeur constante ! Thomas Dugratinet, celui qui sentait toujours bon et qui brassait de la bière sur un air de « je m'enfoutisme ». Guillaume Martel pour tout ce qu'il m'a appris, et pour n'avoir jamais désespéré de jouer aux jeux de société avec moi, et il en faut de la patience pour ça. Céline Thiole, pour la gentillesse et la patience dont elle a fait preuve à mon égard, je lui souhaite énormément de bonheur et de bonnes choses, comme le retrait de son permis afin de lui éviter toute catastrophe. Marie Brevet pour sa dévotion à trouver de nouvelles insultes à mon égard. Marie Le Naour, comtesse des effecteurs MAX, que j'espère avoir su divertir. Mélia Bonnamy pour m'avoir supporté lors de nos sorties « course à pied ». Babil Torralba pour sa détermination propre à lui-même. Maël Baudin parce qu'il a la classe. Rémi Pélissier, digne représentant de l'aile gauche et de l'apéro qualitatif. Alexandre Lepagne, mon bon Alex, mon voisin, qui m'a aidé et encouragé jusqu'à la fin de la thèse, heureusement que tu étais là ! Margaux Jammes, alias Zargolito, fille de Méphistophélès et héritière du trône des enfers. Andy Brousses, crypto addict qui m'impressionnera toujours pour son calme et son sens du travail bien fait. Florian Chariat pour sa sympathie et toute l'aide qu'il m'a apporté, et toujours avec le sourire s'il vous plait, en informatique. Je remercie et

j’encourage pour la fin de leur thèse Mathilde Villegas, Laura Mathieu, Julie Ramirez, Lucas Bonometti (fidèle compagnon de bureau), Nutthalak Laksanavilat, Noun Magdy-Ibrahim-Fouad, Yu Fu, Oumaima Moubset et Sarah Yammine. Bon courage à vous.

Je tiens également à remercier les étudiants étrangers avec qui j’ai passé du temps et dont le nom de famille est pour certain trop compliqué : Karthik, Victor Golyaev, Nathalie Sukhikh et Silvia. J’ai vraiment apprécié passer du temps avec eux, dont mes premières soirées entièrement anglophones, c’était vraiment impressionnant. Merci d’avoir pris le temps de comprendre mon anglais.

Pour finir, je tiens à remercier ces personnes en dehors de PHIM à qui je tiens énormément et qui ont été un soutient indispensable pendant ces quatre dernières années. Mes parents sans qui cette thèse n’aurait pu être possible, mon père sans qui mes études n’auraient jamais pu être possible et ma mère pour ses encouragements tout au long de ma thèse, Marie, ma belle-mère, pour son attitude toujours positive incroyable. Je tiens également à remercier l’ensemble de mes amis troyens qui m’ont vu grandir et félicités pour le parcours accompli, ainsi que ma copine Cindy qui m’a supporté et encouragé à travailler pendant toute la rédaction de ma thèse. Merci à vous.

Voilà, j’espère n’avoir oublier personne. En tout cas merci ! C’était bien cool mais je suis aussi content que ma thèse touche à sa fin. De nouvelles aventures arriveront, bien entendu, et j’espère qu’elles seront scientifiques. Toutefois n’oubliez pas que seuls les *Lithops* sont aussi absolus.

Sur ce, Bonne lecture à tous !

Table of content

Résumé des travaux de thèse en français 1

Introduction 11

Bananas: the most consumable fruit in the world 11

Physiology and taxonomy of bananas 13

Banana pests and pathogens 18

Banana bunchy top disease 19

BBTD transmission and control 23

Banana bunchy top virus 24

Multipartite nature of BBTV 33

Spoiler alert: Alphasatellitidae (review) 34

RNA interference-based antiviral defences 35

Context and main objectives of the PhD thesis work 37

Chapter 1: Role of alphasatellites in helper virus replication, gene expression, antiviral RNAi evasion and transmission by insect vectors 39

Abstract 40

Introduction 42

Results and Discussion 46

Conclusions and open questions 73

Materials and Methods 76

References 86

Chapter 2: Study of the prevalence and diversity of alphasatellites associated with banana bunchy top virus 97

Abstract 98

Introduction 99

Results and Discussion 101

Conclusions and open questions 127

Materials &Methods 129
References 133
Chapter 3: Host range of banana bunchy top virus and its alphasatellites 137
References 144
Chapter 4: Review on alphasatellites of plant ssDNA viruses 147
Introduction 147
History, classification and host range of alphasatellites 149
Genome organisation and replication of alphasatellites and their helper viruses 151
Diversity of geminialphasatellites and their impact on helper geminivirids 153
Role of alphasatellites associated with nanovirids 158
Coconut foliar decay virus 160
References 161
Discussion and Perspectives 171
References of Introduction & Discussion 176
Supplementary data 189
Supplementary data for Chapter 1: article 1: 191
Supplementary data for Chapter 2, article 2: 218
Supplementary data for Chapter 3, article 3 (short communication): 233

List of frequent abbreviations

BBTD: banana bunchy top disease

BBTV: *Banana bunchy top virus*

PIO: Pacific and Indian Oceans

SEA: South-East Asia

FBNYV: *Faba bean necrotic yellows virus*

HV: helper virus

DRC: Democratic Republic of the Congo

DRC-alphasatellite/DNA-alpha: alphasatellite associated with BBTV isolates from DRC

siRNA: small interfering RNA

Gmb: Gembloux

IC: immuno-capture

PCR: polymerase chain reaction

RCA: rolling circle amplification

bp: base pair; nt: nucleotide

Kb: kilobase

CR-CL: common region stem-loop

CR-M: common region major

RT: room temperature

EtBr: ethidium bromide

BBTA: Banana bunchy top alphasatellite

Résumé des travaux de thèse en français

Important : Les informations fournis dans ce résumé sont détaillées et/ou citées plus loin dans les différentes parties du manuscrit (en anglais).

Introduction

La banane est considérée comme la 4ème plus grande culture mondiale et la première culture fruitière au monde. Elle est produite à partir des plantes de bananiers qui sont des monocotylédones de la famille des Musaceae. Elles sont cultivées sur de grandes surfaces en monoculture intensive ou en culture associée et majoritairement récoltées sous forme de banane dessert et/ou de banane à cuire. Sa production intensive la rend sensible à de nombreux pathogènes et ravageurs, dont les principaux sont les champignons et les virus. Parmi les virus, le banana bunchy top virus (BBTV) est le virus le plus dommageable à la culture de la banane et est considéré comme un des virus les plus dangereux pour le monde agricole, toute culture confondue. Il est responsable de la banana bunchy top disease (BBTD), maladie virale qui va induire chez les plantes infectées l'apparition de tirets verts foncés ou « morse code » sur les nervures de la plante, une chlorose sur la marge des feuilles, mais également un nanisme des plantes qui se traduit par des feuilles plus petites, resserrées et craquantes. Les plantes infectées ne peuvent pas guérir de la maladie et finissent par mourir. De plus, elles ne peuvent plus fleurir et donc ne produisent plus de fruit. Les conséquences sur la production de la banane peuvent conduire jusqu'à une perte de rendement de plus de 95 % à l'échelle d'un pays entier, créant ainsi de véritable épidémie de BBTD et conduisant à des catastrophes économiques majeurs pour les agriculteurs locaux mais également les consommateurs. La BBTD est capable d'infecter les plantes de la famille des Musaceae et donc toutes les espèces de bananiers cultivées. Elle est aujourd'hui présente en Afrique, en Asie, en Océanie et sur les îles de l'océan Indien. Seuls l'Europe, où la culture de la banane est presque absente, et le continent américain ne présentent pas de signes de la présence de la maladie à ce jour. Cependant, la propagation de plus en plus rapide, notamment en Afrique et en Asie depuis le XXème siècle alerte sur l'importance d'une gestion de cette maladie virale. Or, l'unique manière de l'éradiquer efficacement à ce jour est sa détection rapide, puis l'arrachage des plantes infectées ainsi que celles l'entourant dans un rayon de 15 mètres. Cependant, les symptômes de la maladie virale peuvent mettre jusqu'à 6 mois avant d'apparaitre sur une plante contaminée, rendant difficile la détection de la maladie avant sa propagation. La maladie virale et le virus sont transmis de plante à plante par le puceron noir du bananier *Pentalonia nigronervosa*, unique vecteur du

BBTV connu à ce jour, selon le mode circulant non propagatif. Ce puceron présent dans l'ensemble des régions où est cultivée la banane ne peut être éradiqué, d'autant que cela induirait une forte érosion génétique. En effet pour le détruire, il faudrait employer des insecticides qui ne sont pas que spécifiques de ce puceron. Il est donc urgent de mieux comprendre comment se déroule l'infection virale et comment il serait possible de mieux la prévenir.

Concernant le BBTV, il s'agit d'un virus de la famille des *Nanoviridae* et du genre *Babuvirus*. C'est un virus circulaire à ADN simple brin (ssDNA) multipartite qui est composé de 6 ADN différents, chacun encapsidé par la même protéine de capside et codant pour une seule protéine. Il est composé de l'ADN-C codant pour la « cell cycle protein » (Clink) qui est impliquée dans le changement de phase cellulaire d'une cellule en la forçant à rentrer en phase S (réplication de l'ADN génomique). L'ADN-M code pour la protéine de mouvement (MP) qui permet au virus de se déplacer dans la plante. L'ADN-N code pour la « nuclear shuttle protein » (NSP) qui permet au virus de rentrer dans le noyau des cellules végétales mais qui sert également de facteur assistant essentiel à la transmission du virus par le puceron vecteur. L'ADN-R code pour la « master replication initiator protein » (M-Rep) qui est impliquée dans la réplication de chaque ADN viral du BBTV. L'ADN-S code pour la protéine de capside (CP) qui permet l'encapsidation des différents ADN viraux et la formation du virion. Enfin, l'ADN-U3 code pour une petite protéine à la fonction inconnue. L'ensemble de ces six composants viraux forment le BBTV et permettent la mise en place de la maladie virale. Chacun des six composants viraux font environ 1 à 1,1 kilobase (kb) et possède deux régions communes : la « common-region stem-loop » (CR-SL) et la « common-region major » (CR-M). La CR-SL fait 69 nucléotides et contient la séquence 'TATTATTAC' qui forme une boucle d'ADN, origine de réplication des ADN viraux. La M-Rep va reconnaitre cette séquence et la scinder afin de permettre à une ADN polymérase de la plante de répliquer l'ADN viral. La CR-M fait environ 92 nucléotides et forme une région commune aux différents composants du BBTV supposée être la région de priming permettant la synthèse des brins complémentaires d'ADN de chaque composant viraux. Les ADN double brin formés seront la matrice de l'ARN polymérase (Pol II) de l'hôte et ainsi permettront la production d'ARN messagers (ARNm) viraux. A ce jour, les mécanismes de réplication du BBTV sont peu connus et se basent principalement sur les connaissances d'une autre famille de virus de plante à ADN circulaire simple brin : les G*eminiviridae*.

En plus des six composants du BBTV, il peut être retrouvé associé avec des ADN circulaires simple brin d'environ 1,1 kb appelé alphasatellites. Les alphasatellites sont des molécules virales similaires à l'ADN-R du BBTV appartenant à la famille des *Alphasatellitidae* codant pour une seule protéine Rep qui leur permet de s'autorépliquer. Toutefois, ils ont besoin d'un virus assistant, ici le BBTV, afin de se déplacer dans la plante, de s'encapsider et d'être transmis de plante à plante par le puceron vecteur. A ce jour, les alphasatellites ont été découverts associés uniquement avec des virus de plante à ADN circulaire simple brin des familles des *Nanoviridae* (genres *Nanovirus* et *Babuvirus*), *Metaxyviridae* (genre *Cofedevirus*) et *Geminiviridae* (genres *Begomovirus* et *Mastrevirus*). Leur présence lors de l'infection virale n'est pas systématique, elle n'est donc pas nécessaire à la mise en place de la maladie virale. De plus, bien que l'on puisse penser que de telles molécules apportent des bénéfices à leur virus assistant, les alphasatellites ne sont pas capables de transrépliquer les composants de leur virus assistant et leur rôle dans l'infection n'a pas été bien établi. Chez les *Nanoviridae*, très peu d'études existent à leur sujet. Chez les *Geminivirus*, certaines études ont montré que les alphasatellites avaient un impact négatif sur leur virus assistant, en réduisant leur accumulation virale. Mais d'autres ont montré que la présence des alphasatellites augmentait la sévérité des symptômes viraux ou permettait de passer outre les défenses antivirales de la plante basées sur le mécanisme des petits ARNs interférents.

Le mécanisme de l'ARN interférence chez les plantes consiste à la production de petits ARNs majoritairement de taille de 21, 22 et 24 nucléotides (nt) suite à la découpe d'un ARN double brin par des ribonucléases appelées DICER. Ces petits ARNs vont ensuite être pris en charge par des protéines nommées Argonaute et former le complexe RISC pour RNA induced silencing complex. Le complexe RISC ainsi formé d'un petit ARN qui va servir de guide et d'une protéine Argonaute va pouvoir cibler précisément soit des ADN et permettre la mise en place de marques épigénétiques qui vont empêcher la transcription de nouveaux ARN messagers, soit de cibler des ARN messagers afin de les empêcher d'être traduit en protéine en les dégradant ou en bloquant leur traduction. Ce système de l'ARN interférence est à l'origine un mécanisme complexe de régulation des gènes impliqué dans le développement des plantes. Mais il sert également en tant que moyen de défense contre les acides nucléiques étrangers, tel que les virus. Ainsi, à partir d'ARN double brin viraux, il y aura production de petit ARN viraux de 21, 22 et 24 nt qui vont cibler l'ARN viral ou l'ADN viral bloquant sa transcription au travers de la RNA dependant DNA Methylation (RdDM). Ce mécanisme, qui a été mis en évidence au début des années 2000, est aujourd'hui très étudié en virologie des plantes afin de comprendre

comment elles se comportent face à une infection virale. Concernant le BBTV et les alphasatellites, l'étude des mécanismes de défenses des plantes en réponse à de telles molécules n'ont pas encore été étudiés en détail.

Objectif de la thèse

En 2016, des pucerons porteurs du BBTV provenant de la République Démocratique du Congo (RDC) ont été ramené en France au laboratoire de l'Unité Mixte de Recherche - Plant Health Institut of Montpellier (UMR PHIM). Ils ont été nourris sur des bananiers sains et ont ainsi transmis leur isolat du BBTV. L'étude par séquençage haut débit avec la technologie Illumina d'ADN de pucerons infectés et de bananiers infectés par le BBTV a permis de mettre en évidence la présence des six composants du BBTV dans les isolats viraux mais également la présence d'un alphasatellite. Cet alphasatellite est le premier découvert associé au BBTV en RDC, les autres précédemment découverts associés au BBTV provenaient de Chine, du Vietnam et de Taiwan. L'analyse de la séquence complète de cet alphasatellite comparée aux autres séquences d'alphasatellites disponibles sur Genbank a montré qu'il appartenait à un genre d'alphasatellite unique. De plus, nous avons montré qu'il ne modifiait pas les symptômes de la maladie virale et qu'il pouvait être perdu lors de la transmission du virus de plante à plante. L'objectif majeur de ma thèse a ainsi été de caractériser cet alphasatellite en laboratoire et de déterminer s'il avait un impact sur son virus assistant. Une collecte d'échantillons de BBTV en 2018 et 2019 au Vietnam, Laos et Chine, ainsi que des échantillons de BBTV provenant de RDC, du Gabon, de Malawi, Nouvelle-Calédonie et du Bénin disponibles au laboratoire a conduit à mon second objectif de thèse : celui de caractériser par séquençage haut débit la diversité des isolats de BBTV et s'il y avait présence d'alphasatellite dans ces échantillons. Enfin, le dernier objectif de ma thèse a été la recherche d'un hôte alternatif du BBTV. Des plantes autres que des bananiers et présentant des symptômes atypiques, proches des bananiers infectés par le BBTV au Vietnam et au Laos, ont été prélevées et analysées par séquençage haut débit.

Résultats

Premier objectif de thèse : analyse de l'alphasatellite découvert en République Démocratique du Congo

L'analyse en laboratoire de cet alphasatellite a permis de publier un article scientifique dans la revue *PLoS Pathogens* et portant le titre « A newly emerging alphasatellite affects banana bunchy top virus replication, transcription, siRNA production and transmission by aphids ». Cet article détaille la caractérisation de l'alphasatellite associé au BBTV en RDC. Tout d'abord, suite à la découverte de l'alphasatellite, nous avons analysé l'ensemble des échantillons de BBTV disponible dans notre laboratoire afin de retrouver cet alphasatellite dans d'autres échantillons. Nous l'avons ainsi détecté dans un seul échantillon datant de 2012 et provenant également de RDC, ce qui nous a permis de conclure que cet alphasatellite était émergent en RDC. Nous ne l'avons détecté dans aucun autre échantillon provenant d'autres pays. Nous avons également montré que l'alphasatellite était encapsidé par son virus assistant, qu'il impactait négativement la transmission de la maladie virale en réduisant son taux de transmission de plante à plante. De plus, par q-PCR nous avons montré que la présence de l'alphasatellite réduisait la charge virale du BBTV à la fois dans le bananier et dans le puceron. La présence de l'alphasatellite de RDC a donc un impact négatif sur son virus assistant au niveau de l'ADN. De plus, nous avons également séquencé les ARN messagers totaux et les petits ARNs totaux de trois échantillons infectés avec le BBTV et trois échantillons infectés avec le BBTV en présence de l'alphasatellite. L'analyse des ARN messagers viraux entre les deux conditions a permis d'observer que la quantité d'ARN messagers du BBTV diminuait en présence de l'alphasatellite, avec un impact particulièrement négatif pour la transcription d'ADN-N. Concernant les petits ARNs, la présence de l'alphasatellite induit également une baisse des petits ARNs viraux pour l'ensemble des composants du BBTV, excepté pour le composant R. Concernant l'alphasatellite, la q-PCR a permis de montrer que c'était l'ADN viral le plus abondant dans les échantillons de bananiers infectés et un des plus importants chez les pucerons infectés. Toutefois la quantité d'ARNs messagers de l'alphasatellite est très faible, signifiant qu'il est faiblement transcrit. C'est également un des composants viraux le plus ciblé par les petits ARNs viraux. La q-PCR nous a également permis d'établir l'abondance relative de chacun des composants du BBTV et de l'alphasatellite et ainsi d'établir la formule génomique virale. En comparaison, nous avons également reconstruit la formule des ARN messagers viraux et des petits ARNs viraux. La formule des ARNs messagers est très différente de la formule génomique (ADN) et de la formule des petits ARNs viraux, qui elles sont très

ressemblantes. Nous avons pu ainsi conclure que la quantité de petits ARNs viraux produit était proportionnelle à la quantité d'ADN viral et ainsi que l'alphasatellite était capable d'échapper au mécanisme de défense de la plante basé sur les petits ARNs. De plus, le séquençage des petits ARNs viraux a permis de mettre en évidence que non seulement tous les composants du BBTV et l'alphasatellite sont la cible des mécanismes de défense de la plante basé sur les petits ARNs, mais également que l'entièreté de chaque séquence virale était ciblée sur leur deux polarités (sens et anti-sens) par des petits ARNs des tailles de 21, 22 et 24 nt, qui est un signe distinctif des virus à ADN. Concernant les ARN messagers, nous avons également pu observer la présence d'ARN dans les deux polarités sur l'ensemble de chaque séquence, signe qu'il y a bien formation d'ARN double brin précurseur des petits ARNs. Enfin, nous avons cloné l'ensemble des composants du BBTV et l'alphasatellite à partir de l'isolat original de RDC et créé des dimères partiels de chaque composant viraux. Ces dimères partiels ont ensuite été introduits dans un vecteur binaire puis dans une souche d'*Agrobacterium tumefasiens* afin d'être utilisés pour agro-inoculer des plantes de bananier et de *Nicotiana benthamiana*. L'inoculation de plantes de *N. benthamiana* a permis d'obtenir des plantes infectées avec le BBTV et l'alphasatellite, et de montrer que l'alphasatellite était encapsidé par son virus assistant, trois semaines post-infection. Toutefois, après deux mois d'infection, les plantes se sont rétablies de leurs symptômes viraux et leur analyse a montré que dans certaines plantes le virus avait disparu mais que l'alphasatellite était toujours présent dans les feuilles systémiques, preuve qu'il est capable de s'autorépliquer et de se déplacer dans la plante de manière autonome. Cette étude montre donc que l'alphasatellite de DRC impacte négativement son virus assistant en réduisant l'accumulation des ADN et ARN messagers viraux, mais également qu'il impacte la production des petits ARNs viraux. Enfin, une analyse phylogénétique de cet alphasatellite avec l'ensemble des alphasatellites découverts, comprenant les précédents alphasatellites associés au BBTV, a dévoilé que cet alphasatellite était plus proche phylogénétiquement d'alphasatellites associés à des nanovirus qui infectent les légumineuses de la famille des Fabaceae (dicotylédones) que des alphasatellites précédemment découvert en Asie associé avec le BBTV. Il forme ainsi un genre d'alphasatellite encore inconnu, infectant une monocotylédone. Nous avons alors proposé de le nommé *Banaphisatellite musae1* formant le genre des *Banaphisatellite* pour banana aphid alphasatellite.

Deuxième objectif de thèse : analyse de la diversité du BBTV et leurs alphasatellites provenant de différents pays.

L'analyse d'échantillons de bananiers infectés par le BBTV provenant du Vietnam (n = 11), du Laos (n = 10), de Chine (n = 5), de RDC (n = 5), du Gabon (n = 3), du Benin (n = 1), de Malawi (n = 1) et de Nouvelle Calédonie (n = 1) par séquençage haut débit avec la technologie Illumina a permis la reconstruction des six composants du BBTV dans l'intégralité des échantillons infectés (sauf un où l'ADN-U3 est absent) mais également d'un total de 29 nouvelles séquences d'alphasatellites. L'ensemble de ces résultats, décrit dans le chapitre 2 de ma thèse, formeront un deuxième article scientifique actuellement nommé « A new genus of alphasatellites associated with banana bunchy top virus in South-East Asia ». L'analyse des séquences a permis de confirmer les deux groupes phylogénétiques du BBTV que constitue l'Asie du Sud-Est (SEA) et les océans Pacifique et Indien (PIO). Le séquençage a permis de reconstruire pour chaque échantillon les séquences virales, qui forment ainsi une ressource génétique inestimable, car comprenant de nombreux isolats de pays où aucune information sur le BBTV n'était encore disponible. De plus, nous avons également étudié en détail les régions communes du BBTV : la CR-SL et la CR-M afin de mettre en évidence les différences entre les groupes phylogénétiques et les composants viraux, mais le plus intéressant reste l'analyse des séquences d'alphasatellites reconstruites. En effet, dans l'ensemble des échantillons, seulement un alphasatellite a été reconstruit dans un échantillon de RDC provenant du groupe PIO, qui avait été également étudié dans la partie 1 (chapitre 1). Ainsi, 28 nouvelles séquences d'alphasatellites ont été reconstruites dans les échantillons du Vietnam, du Laos et de Chine avec dans certains échantillons jusqu'à trois alphasatellites différents. Parmi les alphasatellites reconstruits, 15 séquences proviennent de deux espèces d'alphasatellites déjà décrites dans la littérature associée au BBTV en Chine et au Vietnam. Quatorze autres séquences quant à elles correspondent à deux nouvelles espèces du genre *Banaphisatellite* décrites dans la partie 1 de la thèse. Nous les avons analysées phylogénétiquement avec l'ensemble des espèces types d'alphasatellites précédemment décrites mais également par clustering de leur séquence codante pour la Rep sous forme d'acide aminé. Nous avons ainsi observé que ces alphasatellites étaient plus proche des fabenesatellites (alphasatellites associés à des nanovirus) et des gosmusatellites (alphasatellites associés à des geminivirus) qui infectent des dicotylédones. Ces résultats appuieraient le fait que ces nouveaux alphasatellites auraient fait un bond de plantes dicotylédones à des plantes monocotylédones et se seraient adaptés au BBTV. De plus, nous avons analysés leur CR-M et CR-SL, et mis en évidence les différences entre espèces

d'alphasatellite ainsi que leurs potentielles séquences iterons impliquées dans la réplication virale. Enfin, nous avons analysé la formule génomique de l'ensemble des échantillons séquencés, et observé que les alphasatellites étaient toujours les composants viraux majoritaires, et que plus il y avait d'alphasatellite dans les échantillons, plus la fréquence du BBTV diminuait de manière importante. Nous avons également séquencé les petits ARNs de 6 échantillons possédant des alphasatellites et observé que le taux de petits ARNs variait en fonction des alphasatellites présents, mais ensemble, ces petits ARNs produits à partir de tous les alphasatellites présents dans chaque échantillon constituent une proportion élevée des petits ARNs totaux du virus. Pour certaines espèces d'alphasatellites présents ensemble dans de nombreux échantillons, le taux de petits ARNs produits est très important alors que pour d'autres le taux de petits ARNs est faible, indiquant deux systèmes différents d'évasion des mécanismes de défense de la plante, toujours sans avoir d'impact sur l'accumulation virale.

Troisième objectif de thèse : analyse d'un potentiel hôte alternatif du BBTV

L'analyse des plantes voisines à certaines plantes infectées par le BBTV au Vietnam et au Laos a permis de mettre en évidence deux plantes potentiellement infectées par le BBTV : *Chromolaena odorata* (Asteraceae) et *Bidens pilosa* (Asteraceae). En effet, le séquençage haut débit des ADN viraux de ces échantillons a permis dans la première plante de reconstruire le BBTV (5 composants) ainsi qu'un alphasatellite et dans la deuxième de reconstruire le BBTV au complet ainsi que quatre alphasatellites différents. De plus, l'analyses précise de *C. odorata* a montré que les séquences virales étaient différentes de celles de sa plante voisine, mais également que la formule génomique était différente. Toutefois, le taux de reads (courtes séquences nucléotidiques obtenues par séquençage Illumina) viraux est faible dans cet échantillon, témoignant possiblement que le BBTV ne se réplique pas efficacement dans cette plante. Concernant *B. pilosa*, l'ensemble des séquences du BBTV et d'un alphasatellite sont les mêmes que sa plante voisine, mais les trois autres alphasatellites correspondent à des séquences d'alphasatellites retrouvées dans d'autres échantillons de BBTV. Il n'y a donc aucune nouvelle séquence dans cet échantillon, mais le nombre de reads est très élevé, plus que dans certains échantillons de bananiers infectés par le BBTV. De plus, nous avons observé également une formule génomique totalement différente de celle de sa plante voisine ou d'un autre bananier infecté par le BBTV. Enfin, des échantillons sains séquencés ont permis d'établir un seuil de contamination du séquençage illumina, permettant d'écarter cette hypothèse. De plus, un

échantillon de laboratoire contaminé a également permis d'écarter cette hypothèse car dans le cas d'une contamination en laboratoire, les séquences virales sont identiques et la distribution des reads est presque identique sur la séquence de référence. Nous démontrons que ces deux plantes pourraient être des hôtes alternatifs du BBTV. De manière très intéressante, ces deux plantes sont des dicotylédones, or, jusqu'à maintenant le BBTV n'a été détecté que chez les plantes de la famille des Musaceae et chez quelques plantes de la famille des Zingeberaceae qui sont des monocotylédones. Ce serait donc la première fois qu'un hôte alternatif naturel du BBTV serait une dicotylédone, et le fait que des alphasatellites ont également été retrouvés confirmerait d'autant plus leur origine d'une dicotylédone.

Discussions et Conclusions des travaux de thèse

Ces trois années et demi de thèse ont permis de mettre à jour la découverte d'un nouveau genre d'alphasatellite : les banaphisatellites. Nous avons essayé de pousser au maximum l'analyse de ce nouveau genre et de caractériser ce nouvel alphasatellite découvert en RDC, ce qui a permis de conduire à une étude très complète de l'analyse des alphasatellites. De plus, l'analyse de la prévalence des alphasatellites dans des isolats de BBTV provenant de différents pays a permis d'apporter une ressource génétique inestimable pour l'ensemble des équipes de recherches travaillant sur ce virus, aussi bien en terme de séquences du BBTV qu'en terme de séquences d'alphasatellite. Ces séquences d'alphasatellites ont permis de découvrir le genre des *Banaphisatellite*, encore inconnu et pourtant fréquent dans les échantillons d'Asie du Sud-est. Une revue sur les alphasatellites a également été produite (chapitre 4 de la thèse) et sera publiée afin d'apporter une synthèse de l'ensemble des connaissances sur les alphasatellites mais également d'apporter des idées d'amélioration de la classification, notamment basée sur la séquence d'acide aminé de la protéine Rep codée par les alphasatellites. Ces idées permettent de mettre en évidence des divergences non présentes lors de l'analyse de la séquence entière des alphasatellites. Enfin, l'analyse de plantes voisines de bananiers infectés par le BBTV a permis de compléter l'ensemble de mes recherches, en appuyant l'hypothèse d'un saut de dicotylédone à monocotylédone de l'alphasatellite, tout en apportant de nouvelle piste pour un hôte alternatif du BBTV. Ces résultats sont également appuyés par le clone infectieux du BBTV fonctionnel chez *N. benthamiana,* qui est une dicotylédone. En conclusion, ces travaux de thèse ont permis d'apporter des informations essentielles sur l'étude du BBTV et des alphasatellites, mais également de nouvelles perspectives scientifiques. Une hypothèse générale que nous

garderons suite à ces travaux, est, au vu de la diversité des alphasatellites et leurs multiples rôles dans les différentes infections virales, il est possible qu'ils soient capables de recruter des protéines essentielles à la fonction virales : tel que le composant S pour la protéine de capside et le composant M pour la protéine de mouvement afin de former un nouveau virus à ADN simple brin circulaire. Cette hypothèse s'appuie sur l'existence des *Metaxyviridae*, une famille de virus à ADN simple brin circulaire unique où la M-Rep est codée par un alphasatellite, et où plus d'alphasatellite que de composant viraux sont retrouvés dans une infection virale. Au final, nul doute que la recherche sur les alphasatellites réserve encore de nombreuses surprises, et que ces molécules, à l'apparence de parasite de virus, pourraient cacher un rôle dans l'évolution des virus de plante à ADN bien plus important.

Introduction

Bananas: the most consumable fruit in the world

There are about 1200 varieties or cultivars (cultivated varieties) of bananas (genus *Musa*, family Musaceae) in the world that are currently produced mainly in two forms: dessert bananas, which are rich in sugar when ripe and very digestible and largely represented by the cultivar Cavendish (*Musa acuminata* triploid with AAA genotype), and cooking bananas, which are firmer and rich in starch, must be cooked before consumption and mostly represented by plantains (*M. acuminata* x *M. balbisiana* hybrid triploid with AAB genotype, see below) (Lescot, 2020). Over 116 million tons of dessert bananas were produced in 2019 from the production area of more than 5.1 million hectares mainly distributed in Asia (54.1%), America (25.5%) and Africa (18.4%) but also to a lesser extent in Oceania (1.5%) and in Europe (0.5%) (FAOSTAT 2019, Figure 1A). India is the first worldwide producer of dessert bananas > 30 million tons, followed by China (ca. 11.5 million tons) and Indonesia (ca. 7.2 million tons). Over 41 million tons of plantains and others cooking bananas were produced in 2019 from the production area of 5.7 million hectares according to FAOSTAT, being distributed only in three continents: Africa, with more than half of the worldwide production (64.2%), America with 23.3% and Asia with 12.4%. The three first countries producing cooking bananas are Ghana, Democratic Republic of the Congo and Cameroon, with a total production almost identical ranging between 4,870 and 4,780 million tons in 2019 (Figure 1B). Finally, with a worldwide production about 157 million tons per year over a cultivated area of more than 10.8 million hectares, of which 15 % is used for export (Lescot, 2015; Kumar *et al.*, 2015) and according to "Food and Agriculture Organization of the United Nations" (FAOSTAT, 2019), bananas are the most economically important world fruit crops mainly cultivated in humid sub-tropical area (Price, 1995; Heslop-Harrison & Schwarzacher, 2007). Furthermore, banana cultivation is one of the most important sectors for maintaining global food security providing staple food for over 130 countries and feeds more than 400 million people in tropical countries (Fouré & Tézenas Du Montcel, 2000).

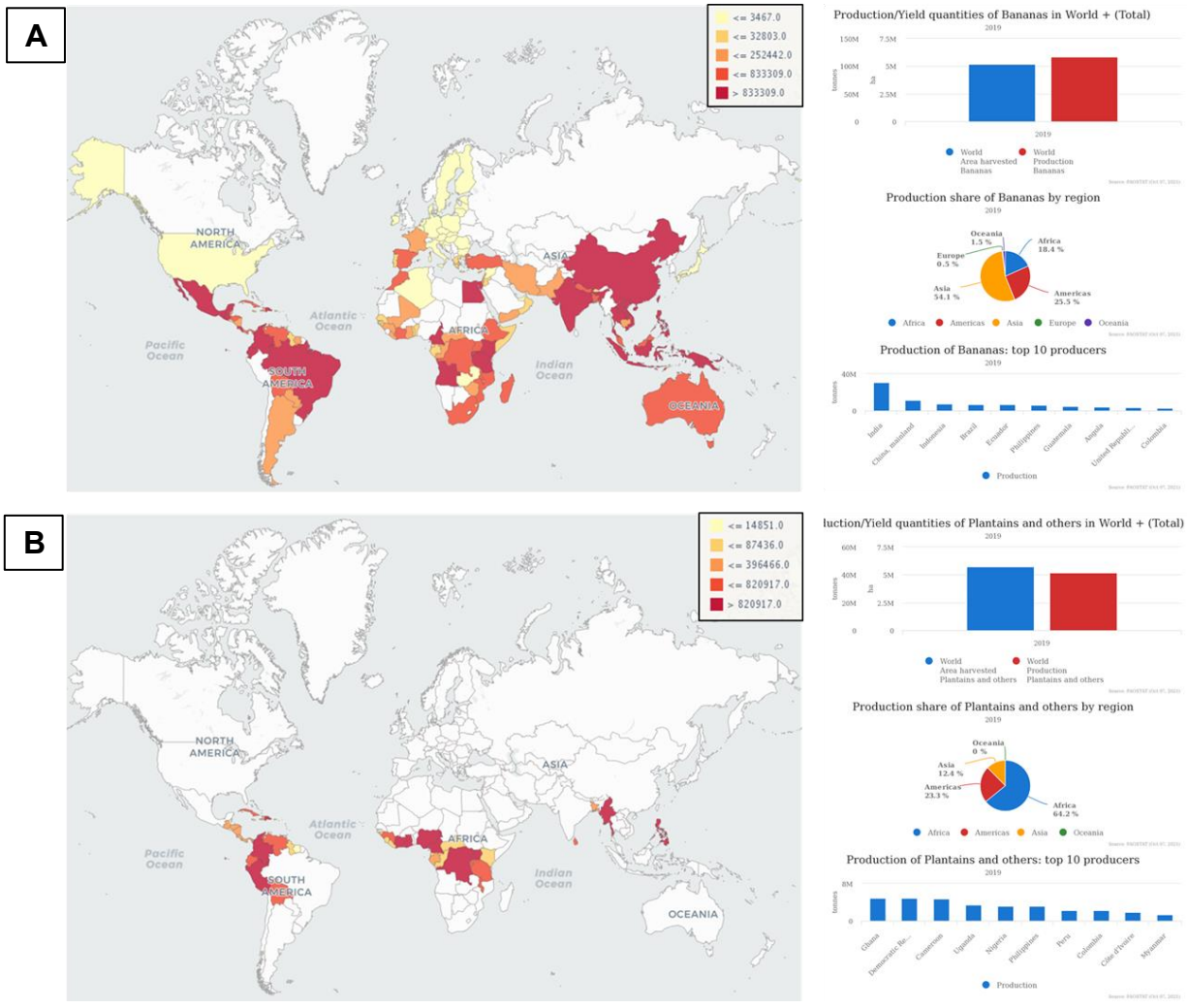


Figure 1. Worldwide production of bananas (A) and plantains (B).

The Figure from FAOSTAT, 2019.

Physiology and taxonomy of bananas

Bananas and plantains are giant herbaceous plants, consisting of branched underground organ, named rhizome or corm, with roots and vegetative buds, and an erect pseudostem composed of tightly packed leaf bases, which can be 1 to 8 meters high (Figure 2 - Jones, 2019). The apical meristem, which is located in the centre of the pseudostem at about soil level, gives rise to a succession of lead primordia. Each primordium grows upwards and differentiates into a leaf base, a petiole and a lamina. The petiole and tightly rolled lamina emerge at the top of the pseudostem in the centre of the cylinder formed by older leaf bases. The lamina starts to unfold when fully emerged. In tropical countries, the first phase of growth can last from 6 to 8 months, with a continuous foliar production and in general 10 to 15 functional leaves simultaneously. At a certain stage of plant development, usually after about 25 to 50 leaves have been produced, the apical growing point stops producing leaves and develops an inflorescence, which marks the phase of floral initiation. The inflorescence is forced out through the top of the true stem elongates. The portion of the true stem with the inflorescence that protrudes beyond the base of the uppermost leaf at flowering is known as the transitional peduncle. The bracts on the inflorescence lift to expose double layers of female nodes with tightly packed fruits known as “hands”. Individual fruits that arise from each flower are often referred to as “fingers” (Figure 2). Bracts later dry and fall off. This is followed along the peduncle by two or three clusters of neutral flowers, which commonly abscise. Clusters of male flowers then follow to the end of the male peduncle where the male bud is located, which contains the remaining male flowers tightly enclosed in their bracts. The male bud is sometimes called the bell and sometimes consumed. The bunch is the collective name for the hands of fruit attached to the female peduncle. In the tropics, bunches can be harvested between 85 and 110 days after flowering, but it’s up to 210 days in cool subtropics. After harvest, the pseudostem dies. The plant propagates by forming a sucker, which is an outgrowth of a vegetative bud on the rhizome. Several suckers can arise from one rhizome. One sucker is usually selected by the farmer as a follower to grow on and regenerate the plant (Lassoudière, 2007). This mode of vegetative multiplication makes it possible to maintain the same banana tree during several years and cycles of culture.

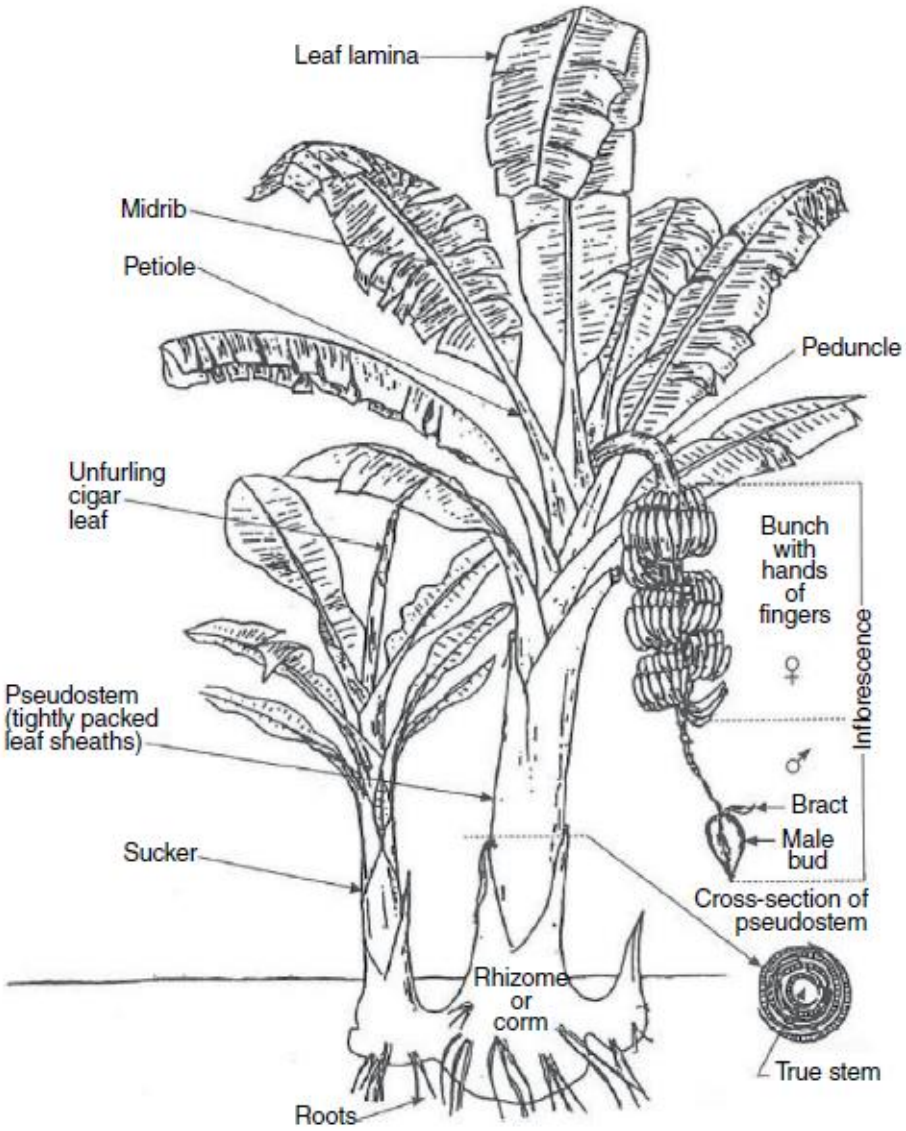


Figure 2. Banana morphology (picture from Jones, 2019)

Bananas and plantains are monocotyledonous (monocot) plant species classified in the genus *Musa* of family Musaceae in the order Zingerberales (Ploetz *et al.*, 2007). Musaceae also contains *Ensete* and *Musella* genera which are morphologically similar to *Musa* plants. The genus *Musa* contains more than 50 banana species, classified since 1947 (Argent, 1976) into 5 sections based on morphological characteristics and chromosome number: *Ingentimusa* (2n = 14), *Callimusa* (2n = 20), *Australimusa* (2n = 20), *Eumusa* (2n = 22), and *Rhodochlamys* (2n = 22). Interestingly, a new classification was established in 2013 to proposes 2 sections, *Musa* (grouping *Eumusa* and *Rhodochlamys*) and *Callimusa* (grouping *Australimusa, Callimusa* and *Ingentimusa*), but is not yet globally used (Häkkinen, 2013). Most of edible bananas and plantains which are grown for consumption come from the *Eumusa* section and are derived from intra- and inter-specific crosses of the seedy diploid species *Musa acuminata* (denoted as genome A) and *Musa balbisiana* (denoted as genome B) (Ploetz *et al.*, 2007; Heslop-Harrison & Schwarzacher, 2007). Both species are thought to be domesticated between 7,000 and 10,000 years ago and have a wild origin in the "South Pacific Triangle" that extends from eastern India to southern China and as far south as the islands of Indonesia and New Guinea (Figure 3A - Simmonds, 1962; Denham *et al.*, 2003). Around 180 wild diploids of *M. acuminata* (denoted AA or AAw, w = wild) and *M. balbisiana* (denoted BB or BBw produce fruits of little interest for human consumption, although they can be used for feeding livestock, as can the leaves and fibres for textiles (Lescot, 2020).

Random natural fertilization events, between an unreduced gamete (2n) and a reduced gamete (n) from genetically distant parents, have led to the appearance of intra- (AAA) and interspecific triploids (AAB, ABB). These triploid banana plants have greater vigour and growth (heterosis effect), and better adaptation to diverse environments, probably due to the contribution of an additional allele for each gene (Perrier *et al.*, 2009). Their cultivation was exported beyond the original centre of diversity (Figure 3B - Perrier *et al.*, 2011). Domestication of bananas was based on the selection of key traits such as parthenocarpy (absence of seeds in the fruit) and polyploidy (mainly triploidy) often associated with more vigorous growth and larger fruit production. The genetic mutation leading to the parthenocarpy trait probably arose in the species *M. acuminata*, as no diploid BB individuals have been reported to date to carry this trait. The vegetative mode of banana propagation through suckers allows the maintenance of seedless individuals over time.

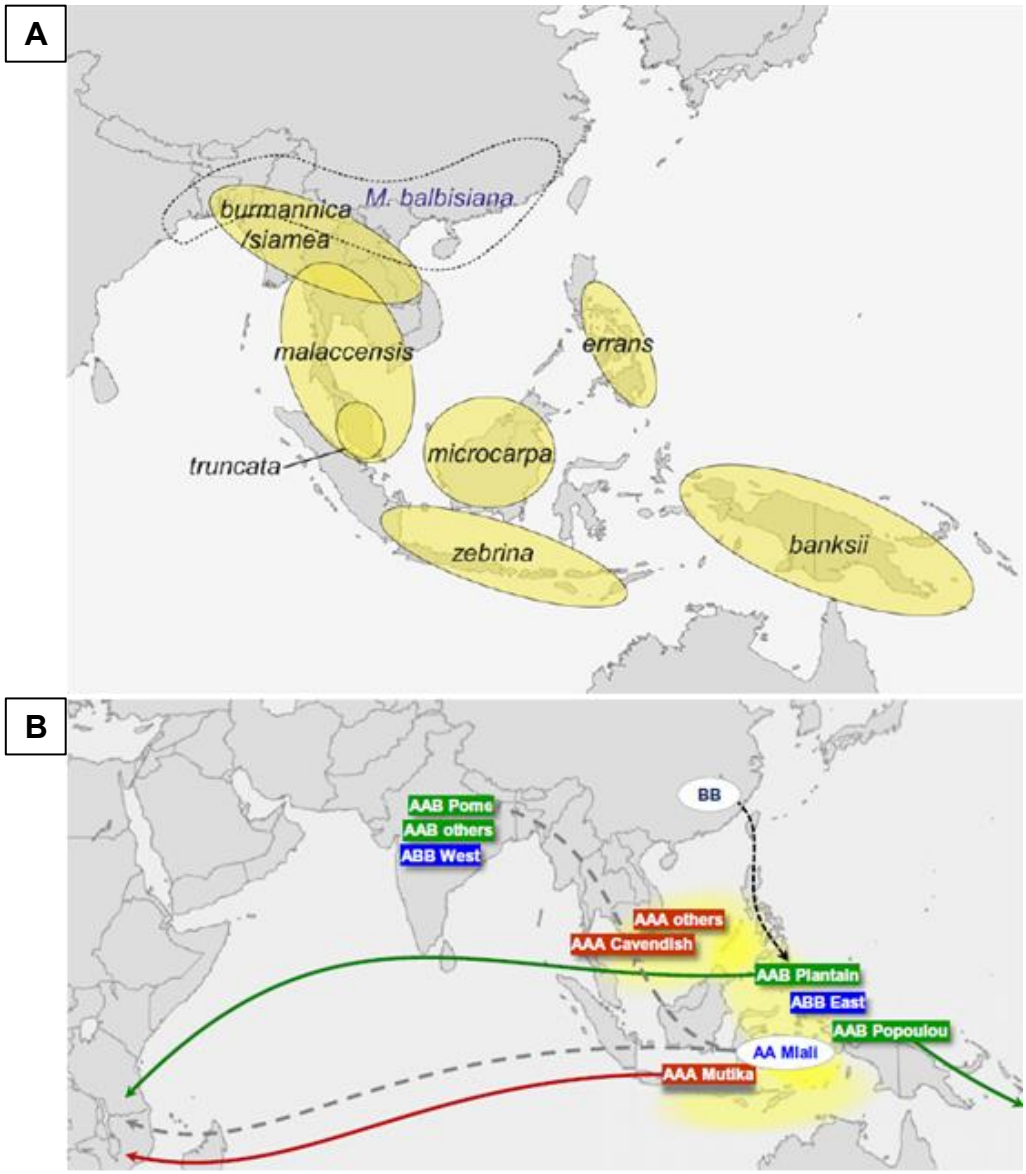


Figure 3. Origin and exportation of *Musa acuminata* and *Musa balbisiana*

(A): Geographical distribution of *M. balbisiana* (diploid BB) and subspecies of *M. acuminata* (diploid AA) in yellow circle. (B): Migration of the main triploid subgroups which occurs naturally. Gray dotted arrows indicate (i) the migrations of Mlali AAcv subgroup, which is not found in ISEA today, to mainland southeast Asia, where it contributed to AAA Cavendish, then to India, where it met *M. balbisiana* to give AAB Pome; and (ii) migrations of the Mlali subgroup to the East African coast. Black dotted arrows indicate the route of *M. balbisiana* from south China to NG over Taiwan and the Philippines, if Austronesian speakers were instrumental in the dispersal of this species. Both figures were adapted from Perrier et al. (2011).

Today, among all cultivars and varieties, 4 major groups are distinguished: 1) diploid AA and AB, 2) triploid AAA, 3) triploid AAB and 4) triploid ABB (Figure 4, Simmonds & Shepherd, 1955). There are naturally occurring tetraploid varieties (AAAA, AAAB, AABB, and ABBB) that are quite rare, but some cultivated tetraploid cultivars which are hybrids were created by breeding programs such as the "Fundación Hondureña de Investigación Agrícola" (FHIA) (Jones, 2019). Plantains, which belong to a subgroup of AAB varieties, have significant diversity with more than 200 accessions (Jones, 2019). Dessert bananas are part of the AAA group and the best known cultivars are the clones of 'Gros Michel' and 'Grande naine' groups. Today, the 'Cavendish' group, belonging to the 'Large Dwarf' variety, accounts for more than 50 % of the world's production. Nevertheless, and regardless to mineral deficiencies, chemical injury, adverse climate and extreme weather, the low genetic diversity of AAA group and the extent of banana cultivation make this culture susceptible to many pest and diseases.

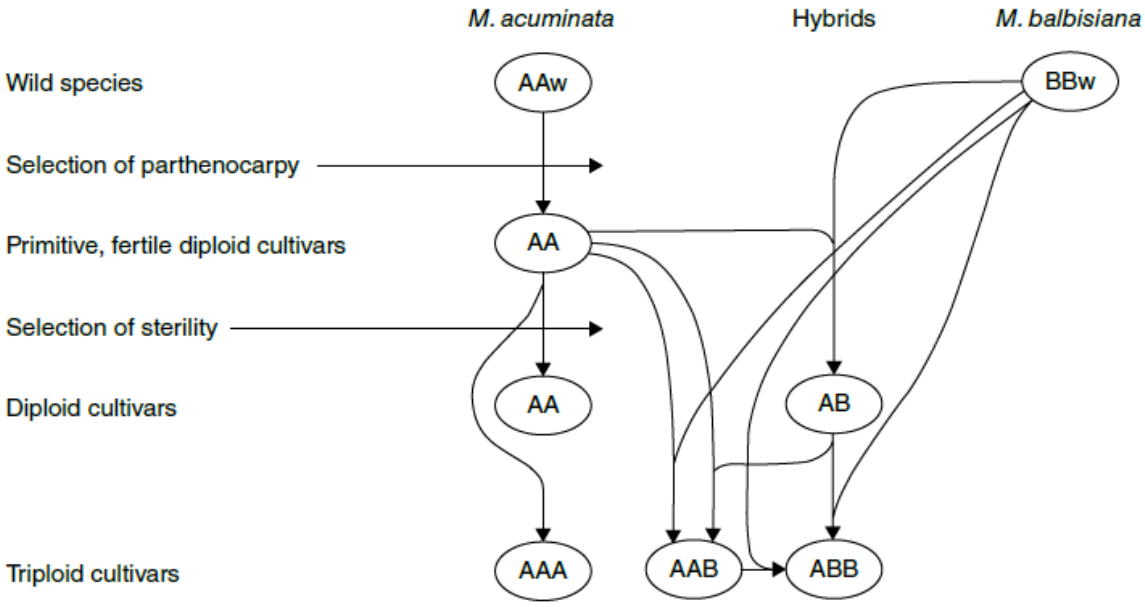


Figure 4. Evolution of the main genomic groups of edible banana cultivars and *Eumusa* series.

From Jones, 2019. “w” = wild type. A = *Musa acuminata* genome. B = *Musa balbisiana* genome.

Banana pests and pathogens

Intensive monoculture is a profitable and efficient agronomical production system allowing high yields of the same type of culture but it exerts significant pressure on pathogen populations. The lack of rotation and diversity in monocultures favors the appearance of adapted genotypes in pathogens that can lead to yield losses and press for variety change. Regarding to perennial crops, changing varieties regularly is not economically viable or sometimes impossible. Therefore, the inoculum pressure in this type of crops is constant because it is uninterrupted between two agricultural cycles, and this is particularly true in tropical regions that require a large quantity of phytosanitary products (Gowen, 1995). With the world global climate change and, in this context, the reduction of pesticide use, the impact of diseases and pests may increase in future years (Fones *et al.*, 2020). Banana production for export relies mainly on this monoculture model and on varieties of the Cavendish group, which are particularly susceptible to many banana diseases and pests (Ploetz *et al.*, 2015; Jones, 2019).

Among pests causing destructive damage to cultivated bananas, the most recognized are nematodes which are separate in 4 types: burrowing nematodes, spiral nematodes, root-lesion nematodes and root knot nematodes caused by *Radopholus similis*, *Helicotylenchus multicinctus*, *Pratylenchus spp* and *Meloidogyne spp*, respectively (Jones, 2019). Considering diseases impacting the banana cultivation, there are many fungal diseases split in three types. Diseases affecting the roots, corm and pseudostem are mainly caused by the fusarium wilt or a so-called Panama disease (Stover, 1962). This disease is caused by the fungal agent *Fusarium odoratissimum* previously known as *Fusarium oxysporum* f. sp. *cubense* (Ploetz, 2006). This fungal agent possesses several variants of which the tropical race 4 (TR4) is the most destructive on banana varieties including Cavendish (Ploetz, 2015). Previously restricted to the Eastern Hemisphere, this threat was detected in northeast Columbia in 2018 and now put in danger the whole New World continent with dramatic crop losses (García-Bastidas *et al.*, 2020).

Other fungal diseases affecting banana fruits are the anthracnose fruit rot caused by *Colletotrichum musae* (Dring & Wardlaw, 1974; Zakaria, 2021), and the brown spot caused by *Cercospora hayi*, (Crous & Braun, 2003), while those affecting leaves include the most damageable complex of three pathogens: *Pseudocercospora fijiensis*, *P. musae* and *P. eumusae* (also known as *Mycosphaerella sp.*) grouped in “Sigatoka leaf spot disease”. In addition to

fungi, there are also bacteria and phytoplasmas that cause banana diseases, such as *Xanthomonas campestris pv. Musacearum*, (Xcm), the destructive pathogen in eastern and central Africa (Jones, 2019). The last but not least are banana diseases caused by viruses.

Banana plants were recorded to be susceptible to 7 different viruses of which 5 viruses are responsible for bananas diseases (Jones, 2019). The bract mosaic disease is caused by banana mosaic bract virus (BBrMV, +ssRNA virus, genus *Potyvirus*, family *Potyviridae*) and is geographically restricted to Philippines, India and Sri Lanka. The banana mosaic, a widespread disease, usually everywhere bananas are cultivated, is caused by cucumber mosaic virus (CMV, +ssRNA virus, genus *Cucumovirus*, family *Bromoviridae*) with a broad host range. Nevertheless, it is usually only a nuisance to growers establishing new plantings using corms or suckers. The banana streak disease is also one of the most widespread banana diseases caused by several species of banana streak virus (BSV, dsDNA-RT virus, genus *Badnavirus*, family *Caulimoviridae*). This virus has a large range of incidence, which in some places can be a serious menace for farmers. The banana mild mosaic disease, with a low economic impact, is caused by banana mild mosaic virus (BanMMV, +ssRNA virus, unassigned genus, family *Betaflixiviridae*). Banana virus X (BVX, +ssRNA virus, unassigned genus, family *Betaflixiviridae*) and ampeloviruses (+ssRNA viruses, genus *Ampelovirus*, family *Closteroviridae*) were also discovered in banana germplasm collections, but infected bananas are symptomless and thus these viruses have no agronomical and economic impact. And the last virus, subject of this PhD thesis, is banana bunchy top virus (BBTV, ssDNA virus, genus *Nanovirus*, family *Nanoviridae*), the most damaging viral agent of bananas and plantains, responsible of the banana bunchy top disease.

Banana bunchy top disease

Banana bunchy top disease (BBTD) was first recorded in 1889 in Fiji (Magee, 1927), but was already observed there in 1879. The origin of BBTD in Fiji is not clear, and as described above, the centre of origin of the genus *Musa* is the South and Southeast Asian-Australasian region and it is probable that the causal agent of BBTD also originated somewhere in this broad geographical area (Stainton *et al.*, 2012), which is confirmed by phylogenetic analysis of sequenced BBTV isolates (Stainton *et al.*, 2015). This disease is considered to be the most economically destructive for cultivation of bananas and plantains and among the top 100 of the

"World's Worst" invaders (Lowe *et al.*, 2020). It was responsible for dramatic crops loss, with some historical epidemies. The most known is in Australia where BBTV was recorded in 1913. At that time, the Australian banana industry grew rapidly with a pick of production in 1922, but 3 years later, it dropped down by 90-95 % and every plantation was affected by bunchy top in 1925. In addition, the total production area was decreased by 90% (Magee, 1927). C.J.P. Magee described at that time: 'It would be difficult for anyone who has not visited these devastated areas to visualize the completeness of the destruction wrought in such a short time by a plant disease' (Magee, 1927). Other dramatic epidemies occurred in Fiji (1900s), in Tamul Nadu (India) during the 1960s, and Pakistan during the 1990s, causing up to 100% losses (Kesavamoorthy, 1980; Soomro *et al.*, 1992; Selvarajan & Balasubramanian, 2014; Kumar *et al.*, 2015).

During the 20th century BBTD has spread quickly in the Old World by movement of planting material by humans, traders and returning soldiers (Magee, 1953; Qazi, 2016). Historically, after its discovery in Fiji islands, the disease was recorded early in 1900 in Taiwan and in 1901 in Egypt but the origins of these outbreaks are unknown (Magee, 1953). Then, the disease has spread in Australia and Sri Lanka in 1913, probably from Fiji (Dale, 1987). It was then recorded in Bonin, Ellice and Wallis islands during the 1920s. In 1940 the disease was first recorded in India (probably brought from Sri Lanka) and in Philippines in 1960 (Dale, 1987). In the same time, BBTD was introduced in Democratic Republic of the Congo (DRC) in the 1950s and recorded in Kisangani region (Kumar *et al.*, 2011). In the next 70 years, BBTD has spread across countries bordering DRC, India, Taiwan and islands around Australia. Today this disease is present in more than 40 countries covering Sub-Saharan, South and Est-Mediterranean Africa, South and Southeast Asian-Pacific, and Australia (Figure 5, Table 1, Jones, 2019). Only the American continent where bananas are cultivated at large scale stays free of the disease, except Hawaii islands. But the disease continues to spread between countries and more recently it was reported in Tanzania (Mpoki *et al.*, 2021). Although spread of BBTD is quite well documented and rigorously followed in African countries with the Alliance for banana bunchy top disease control in Africa (https://www.bbtvalliance.org/index.php/bbtv/), epidemiology of BBTD in Asian-Australian countries is not well documented, and most available information is blurry, false or come from personal communication. The most complete records of the BBTD spread come from Dale, 1987, Kumar et al., 2011 and Jones, 2019, but

some references given by Jones, 2019 appear to be inaccurate. Table 1 and Figure 5 represent the most up-to-date information and world map of BBTD, respectively.

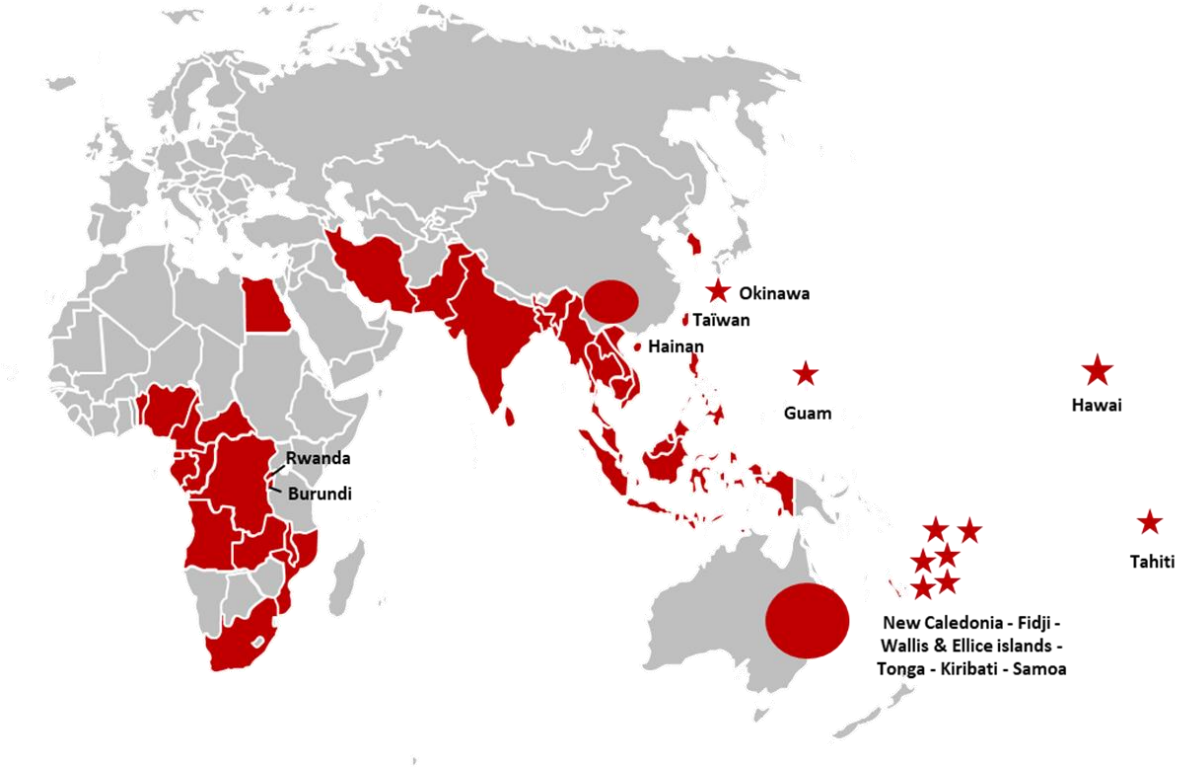


Figure 5. Banana bunchy top disease geographical distribution.

Red stars represent islands, but their location is approximate. Names of small countries or regions are indicated in black.

Free editable world map was taken from https://free-editable-worldmap-for-powerpoint-slides.fr.softonic.com/

Region	Countries and reccording date if available
Africa	**Egypt** (1901) - **Democratic Republic of the Congo** (1950s) - **Central African Republic** (before 1960s) - **Republic of the Congo** (before 1960s) - **Gabon** (1978) - **Rwanda** (1980s) - **Burundi** (1980s) - **Equatorial Guinea** (1980s) - **Malawi** (1990s) - **Zambia** (1990s) - **Angola** (2008) - **Cameroon** (2008) - **Benin** (2011) - **Nigeria** (2012) - **Mozambique** (2015) - **South Africa** (2015)
Asia	**Taïwan** (1900) - **Sri Lanka** (1913) - **Japan** (Bonins islands 1920s - Okinawa 1992) - **India** (1940s) - **Philippines** (1960) - **Vietnam** (1992, first repport in 1969) - **Indonesia** (1978) - **Pakistan** (1989) - **Malaysia** (1991) - **Thailand** (1991) - **Iran** (2004) - **Bangladesh, Cambodia, Vietnam, China** (Hainan-Yunnan), **Laos, Myanmar, Korea** (no estimated date available)
Oceania	**Fidji** (1879) - **Australia** (1913) - **Tonga** (1920s) - **Samoa** (1920s) - **Guam islands** (1920s) - **Ellice and Wallis Islands** (1920s) - **Hawaii** (1989) - **New Caledonia** (1999) - **Tahiti** (2017) - **Kiribati** (no estimated date available)

Table 1. Geographical distribution of banana bunchy top disease

Symptoms of BBTD are very distinctive (Magee, 1953; Hooks *et al.*, 2008). The first symptoms on infected *Musa* plants appear on a newly emerging leaf and are described as discontinuous dark-green streaks or dots along the veins (called Morse code) (Figure 6A). Successive leaves become smaller, both in length and in width of the lamina, and display chlorotic and upturned margins (Figure 6B&C). The leaves become harsh and brittle, they stand straighter than normal, giving rise to a bunch of small leaves at the top of the plant, called "bunchy top", and leading to a stunting of the plant (Figure 6D). Suckers of infected plants show symptoms directly in the first leaves to emerge. Early infected plants do not produce fruits. Interestingly, plants infected late in the growing cycle may produce fruits once, but the bunch stalk and the fruits will be small and distorted (Nelson, 2004). Although infected plants survive BBTD, they are more sensitive to abiotic and other biotic stress, and they often die in one or two years. Nonetheless, symptomless strains in Cavendish plants were reported from Taiwan (Su *et al.*, 1993). For more pictures of BBTD symptoms compare to healthy banana plants: see supplementary data S1 (p240)

Although first reported in 1889, the viral origin of BBTD pathogen agent was discovered by the ground-breaking work of Magee in 1927. In only three years, he established that the causal agent was a virus, named banana bunchy top virus, and its insect vector of the banana aphid *Pentalonia nigronervosa*.

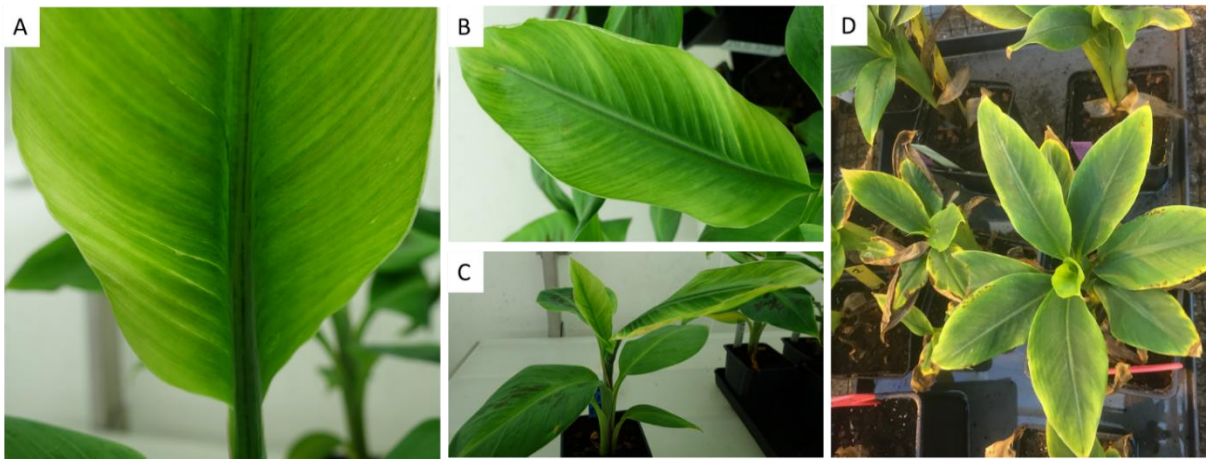


Figure 6. BBTD symptoms.

(A) The first emerging BBTD symptoms: dark-green dotted streaks or "Morse code". **(B)** Curling of infected leaves. **(C)** Emergence of the first systemically-infected leaf: smaller in both length and width and with chlorotic margins. **(D)** fully infected plant, displaying dwarfism and the bunch of leaves on the top.

BBTD transmission and control

BBTD is transmitted from plant to plant by the banana aphid *Pentalonia nigronervosa* Coquerel (family *Aphididae*, order Hemiptera) in a persistent circulative non-propagative manner, and is restricted to phloem tissues (Magee, 1940; Hafner *et al.*, 1995; Hu *et al.*, 1996; Watanabe & Bressan, 2013). Each developmental stage of aphids (winged aphids, wingless aphids and each of the four nymphal instars preceding the adult form) can uptake and transmit the virus (Magee, 1940). To transmit the virus, a minimum acquisition-access period of 4 hours is needed to acquire the virus by banana aphids and a minimum inoculation-access period of 15 min to transmit it (Hu *et al.*, 1996). But the temperature and aphid development stage also play an important role in disease transmission, with a better transmission at 25°C and 30°C than 20°C, and a better transmission with adult aphids than previous instar nymphs (Anhalt & Almeida, 2008). In addition to *P. nigronervosa*, *Pentalonia caladii* was shown to be a competent vector for BBTV (Watanabe *et al.*, 2013). The rate of BBTD transmission is not well defined, mainly because it depends on the virus strain virulence, the aphid lineage, the plant variety and age, and a lot of biotic and abiotic factors that could not be measured or controlled. Finally, it was recently shown that BBTD-infected plants attract aphids through volatile organic compounds (VOCs) emission (Safari Murhububa *et al.*, 2021). The increasing production of VOCs by virus-infected plants is a typical example of vector manipulation by viruses (reviewed in Blanc & Michalakis, 2016).

Once BBTD is established the plant cannot recover from the disease, even if there is some partial resistance (Thomas *et al.*, 2003). Today, the only known way to wrestle or control BBTD is the prompt identification and eradication of infected plants as early as possible, but also the destruction of banana plants around 20 meters from the infected one, and finally replanting virus-free plants. Although tedious, it was done in Australia to control the disease (Dale, 1987), and the rapid detection of infected plants could allow to eradicate it. Likewise, in Togo where the disease was recently spotted, it was immediately controlled (Kolombia *et al.*, 2021). But this method is poorly efficient due to the continuous spread of the disease across the world (Figure 5). In addition, because of the repartition of banana aphids everywhere bananas grow, control of the vector by chemical treatment is sub-optimal and not efficient in the long term. Thus, further studies of the causal agent and its vector to better understand BBTV infection and transmission cycles will help to develop new strategies against the banana bunchy top disease.

Banana bunchy top virus

The causal agent of BBTD is banana bunchy top virus (BBTV). Although Koch postulates were not formally proven using infectious clones of the virus that would recreate the disease via agroinfiltration or biolistic delivery, characteristic BBTV virions were always detected associated with the disease (Dietzgen & Thomas, 1991). BBTV host range is restricted to mainly *Musa* species including *M. acuminata*, *M. balbisiana*, *M. coccinea*, *M. jackeyi*, *M. ornata*, *M. textilis* and *M. velutina* (Thomas *et al.*, 2003). But it was also found to infect the closely related monocot *Ensete ventricosum* (Thomas & Iskra Caruana, 2000). According to Qazi, 2016, there are potentially alternative hosts of this virus outside Musaceae family which required further confirmation (Espino *et al.*, 1993; Geering & Thomas, 1997). Interestingly, the host range of *P. nigronervosa* & *P. caladii* is restricted to the monocot families Araceae, Cannaceae, Commelinaceae, Heliconiaceae, Musaceae, Strelitzeaceae and Zingiberaceae (Blackman & Eastop, 1994; Suparman *et al.*, 2017; Bagariang *et al.*, 2019). Tests of BBTV transmission by viruliferous banana aphids to the monocot plants taro (*Colocasia esculenta*, Araceae) and ginger (*Alpinia purpurata*, *A. purpurata* and *A. purpurata*, Zingiberaceae) was unsuccessful (Hu *et al.*, 1996). But a more recent work of Pinili et al. (2013) showed that the monocot plants *Alpinia zerumbet* (Zingiberaceae), *Colocasia esculenta* (*satoimo*, Araceae) *and Canna indica* (Cannaceae) could serve as alternative hosts upon transmission by *P. nigronervosa* with the BBTV isolates from Japan and Philippines. Furthermore, Hamim et al. (2017) detected BBTV in the monocot *Heliconia aurantiaca* (Heliconiaceae) in Hawaii. Most recently, BBTV was confirmed to infect *M. textilis*, *M. velutina* and *M. acuminata subsp. Sumatrana* but also the three monocot plants: ginger (*Zingiber officinale*), turmeric (*Curcuma longa*) and aromatic ginger (*Kaempferia galanga*) from the family Zingiberaceae in Indonesia, which are intercropped with bananas (Rahayuniati *et al.*, 2021). Notably, studies of alternative hosts were done mainly on plant species which are grown in co-culture with banana plants and which are the hosts for *Pentalonia* aphids. Thus, these studies concluded, in addition to Musaceae hosts, alternative hosts of BBTV are the monocots mostly from the family Zingiberaceae.

Although BBTD was first recorded in 1889, molecular studies of the virus started in the 1980s, and BBTV was initially falsely characterized as a double stranded (ds) RNA virus (Dale *et al.*, 1986) and was classified as a luteovirus (Wu & Su, 1990). It is only in early 1990s the multipartite circular single-stranded (ss) DNA nature of BBTV genome was brought to the light (Harding *et al.*, 1991; Thomas & Dietzgen, 1991; Burns *et al.*, 1994, 1995). Today, this virus is classified in the realm *Monodnaviria*, kingdom *Shotokuvirae*, phylum *Cressdnaviricota*, class *Arfiviricetes*, order *Mulpavirales*, family *Nanoviridae*, genus *Babuvirus* (Krupovic, 2013; Kazlauskas *et al.*, 2019; Krupovic *et al.*, 2020; Koonin *et al.*, 2021). *Nanoviridae* have two genera: *Nanovirus* which now contains 11 species, and *Babuvirus* which contains three species: *Abaca bunchy top virus* (ABTV), the causative agent of abaca bunchy top disease, a disease similar to BBTD but recorded only in Philippines (first in 1910, in contrast to BBTD recorded only in 1960 in the same country) and Malaysia (Sharman *et al.*, 2008), *Cardamom bunchy dwarf virus*, the causative agent of Foorkey disease in eastern sub-Himalayan mountains (Mandal *et al.*, 2004, 2013) and BBTV. All members of *Nanoviridae* have a circular ssDNA multipartite genome organization, are transmitted by hymenoptera insect vectors and are restricted to phloem tissues (Mandal, 2010; King *et al.*, 2011).

BBTV is composed of six circular ssDNA components, each of 1.0 to 1.1 Kilobase (Kb) in length (Figure 7A) and each individually encapsidated by viral coat protein within isometric particles (virions) with a diameter of 18 to 20 nm (Wu & Su, 1990; Burns *et al.*, 1995). The viral replication occurs in the nucleus via a rolling circle replication (RCR) mechanism by the host machinery recruited by viral replication initiator protein (Rep). The RCR occurs on a circular dsDNA intermediate which also serves as a template for viral transcription by the host DNA-dependent RNA Polymerase II (Pol II) (described below). Each of the six BBTV components, namely **DNA-C** (originally designated DNA-5), **DNA-M** (DNA-4), **DNA-N** (DNA-6), **DNA-R** (DNA-1), **DNA-S** (DNA-3), and **DNA-U3** (DNA-2), possesses a single open reading frame (ORF) on the virion strand coding for one protein (monocistronic system). Each ORF is preceded by the Pol II promoter region with a TATA box and followed by a Pol II terminator region with a polyadenylation signal (Beetham et al., 1999). DNA-R codes for the master replication initiator protein (**M-Rep**; 33.6 kiloDalton (kDa)), an endonuclease whose role is to initiate viral DNA replication of all BBTV components in the host cell nucleus. DNA-R is the minimal replicative unit of BBTV (Harding et al., 1993; Horser et al., 2001). Moreover, DNA-R was also found to possess a small ORF within the M-Rep ORF, which codes for a

putative 5 kDa protein of unknown function (Beetham *et al.*, 1997; Herrera-Valencia *et al.*, 2007). DNA-S codes for the coat protein (**CP**; 19.3 kDa) encapsidating each BBTV DNA component and form the virion (Wanitchakorn *et al.*, 1997). DNA-M codes for a movement protein (**MP**; 13 kDa) involved in cell-to-cell movement, and DNA-N for the nuclear shuttle protein (**NSP**; 20 kDa) that allows viral particles to shuttle in and out of the nucleus (Wanitchakorn *et al.*, 2000a; Ji *et al.*, 2019). More recently, it was demonstrated that DNA-N is also essential for aphid transmission of nanoviruses (genus *Nanovirus*) (Grigoras *et al.*, 2018; Di Mattia *et al.*, 2020). DNA-C codes for a cell-cycle link protein (**Clink**; 20 kDa) that turns the host cell cycle into S phase and thus enhances virus replication (Lageix *et al.*, 2007). DNA-U3 codes for a small protein (10.3 kDa) of unknown function (Beetham et al., 1999). All the six components of BBTV genome share two common regions: a common region stem-loop (CR-SL) and a common region major (CR-M) (Figure 7 - Burns *et al.*, 1995). The CR-SL is up to 69 nt long, 62 % identical among all the components and contains an inverted repeat sequences forming a stem-loop secondary structure with the nonanucleotide loop sequence **TATTATTAC** in the loop. This stem-loop structure is the origin of replication, well conserved among *Nanoviridae*, having the nonanucleotide sequence almost invariable (with only the 3rd nucleotide being different in some cases). Based on analogy with other ssDNA viruses (see below), to initiate RCR the M-Rep nicks the nonanucleotide sequence between 7th and 8th nucleotides: **TATTATT//AC** and binds the 5' end of the 3' cleavage product to open the dsDNA intermediate and allow the host DNA polymerase to initiate strand synthesis using the complementary strand as the template (Hafner *et al.*, 1997b). Both catalytic and binding activity of the M-Rep are ATP-independent but need two specific bivalent ions: magnesium (Mg^{2+}) and manganese (Mn^{2+}). The CR-SL also possesses 3 iterated sequence repeats (called iterons) of 5 nts (**GGGAC**), two in forward (**F1**, **F2)** and one in reverse (**R**) orientation, which are presumed to be binding sites for the M-Rep protein (Hafner *et al.*, 1997b; Londoño *et al.*, 2010). Iterons F1 and F2 are both located in tandem on the virion strand, 3' of the stem-loop, whereas the iteron R is located 5' of the stem-loop. The iterons were shown to be required for trans-replication of BBTV DNA-N mediated by M-Rep: mutation of the iterons F1 and R reduced its trans-replication by 42 and 62 %, respectively, while mutation of the iteron F2 or iterons F1+F2 abolished the trans-replication (Herrera-Valencia *et al.*, 2006). These data indicate that iterons F1 and R act in synergy as important enhancers of the replication process, whereas the iteron F2 is essential for replication and recognition by M-Rep. The CR-M is a primer binding site for synthesis of the complementary strand by the host DNA polymerase (Hafner *et al.*, 1997a). CR-M is ca. 66 to 92 nt long, with a 26 nt deletion in DNA-R, and contains in its 3' region a 15 nt

GC-rich sequence with an inverted repeat, well conserved among all the BBTV components (**GGGCC<u>G[A/T/C]A</u>GGCCCGT**). This GC-rich sequence is hypothesized to form a small stem-loop structure (the 3nt loop sequence underlined) similar to the single-strand initiation (ssi) signals in some bacterial plasmids (Lin & Meyer, 1987; Yoichi *et al.*, 1988). In addition, Hafner et al. (1997a) have shown that DNA primers are encapsidated in the virions, since purified virions were able to synthesize the complementary strand in the presence of a DNA polymerase to form the dsDNA template for replication and transcription. These DNA primers of around 80 nts in length appeared to be derived from DNA-C and were able to initiate complementary strand synthesis on all the BBTV components, except DNA-S (Hafner *et al.*, 1997a). This might be due to low relative-abundance of DNA-S (see below the results of this PhD work) in order to limit encapsidation when BBTV invades the host cell. Further investigation is required to better understand the role of *cis*-acting elements in the CR-SL and CR-M in the replication cycle of BBTV.

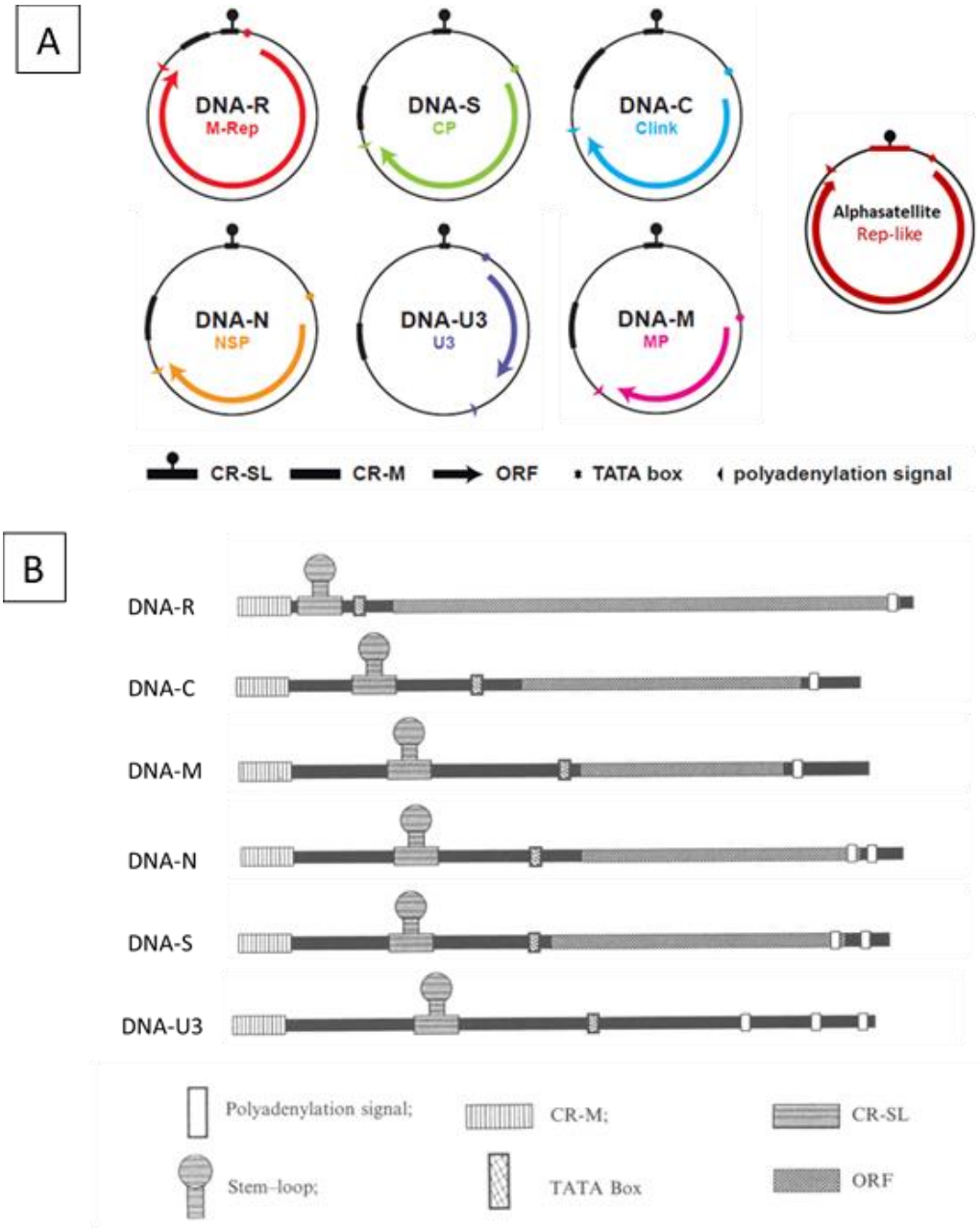


Figure 7. Genome organization of banana bunchy top virus (BBTV) and BBTV alphasatellites.

BBTV is composed of six circular single-stranded DNA components. Each component contains one ORF coding for one viral protein, a common region stem loop (CR-SL) and a common region major (CR-M) which are conserved in all six components. Arrows indicate the location and approximate size of the ORFs and the direction of transcription and translation. BBTV alphasatellite codes for a Rep-like protein and has a CR-SL different from its helper virus. Each BBTV component and alphasatellite have a size ranging from 1.0 to 1.1 Kb. Taken from King et al. (2011) **(A)** and Bruns et al. (1995) **(B)**.

The mechanism of rolling-circle replication (RCR) was not investigated in sufficient detail for BBTV or other nanovirids and it is assumed to be similar to that of geminivirids (Gronenborn, 2004; Jeske, 2009; Pooggin, 2013). In geminivirids, viral ssDNA enters in the host cell nucleus and the complementary strand is then synthesized by the host DNA polymerase using the primer potentially encapsidated with circular ssDNA or via auto-priming, which yields the circular dsDNA intermediate serving as a template for both RCR and transcription of viral mRNA by the host RNA polymerase II (Pol II). Viral Rep protein is translated from its mRNA transported from the nucleus to the cytoplasm and comes back into the nucleus to initiate the replication of viral dsDNA (from which Rep-mRNA was transcribed). To this end, Rep binds the stem-loop regions in the origin of replication. Interestingly, geminiviral Rep has a higher affinity for ssDNA than dsDNA (Thömmes *et al.*, 1993) and binds preferentially the virion strand of dsDNA (Laufs *et al.*, 1995). This binding is guided by recognition of the flanking region of the stem-loop containing iterons sequences (Argüello-Astorga *et al.*, 1994), which form a common region specific to each viral species (Fontes *et al.*, 1992, 1994; Hanley-Bowdoin *et al.*, 2000). Then, M-Rep nicks the stem-loop structure in the nonanucleotide sequence of the loop (as described above for BBTV), gets covalently linked to the 5’ end of the nicked strand, making the released 3’-OH end available to prime the synthesis of a new virion strand by the host DNA polymerase. Opening of the nicked dsDNA and allowing accessibility of the complementary DNA template is expected to be performed by the ATP-dependent helicase activity of Rep as demonstrated by Clérot & Bernardi (2006). At some time after the synthesis of the new virion strand begins, Rep makes a second nick at the same position in the nonanucleotide sequence. At the end of each cycle of RCR, the displaced virion strand is circularized by the ligase (end-joining) activity of Rep and released from the RCR complex. The released circular ssDNA can be used for synthesis of a new dsDNA intermediate for RCR or to be packaged by viral coat protein into a virion. Based on the mechanistic details of RCR available for geminivirids, BBTV and other nanovirids, a model for Rep-mediated rolling-circle replication of nanovirids and alphasatellites (see below) can be proposed (Figure 8A; Pooggin unpublished).

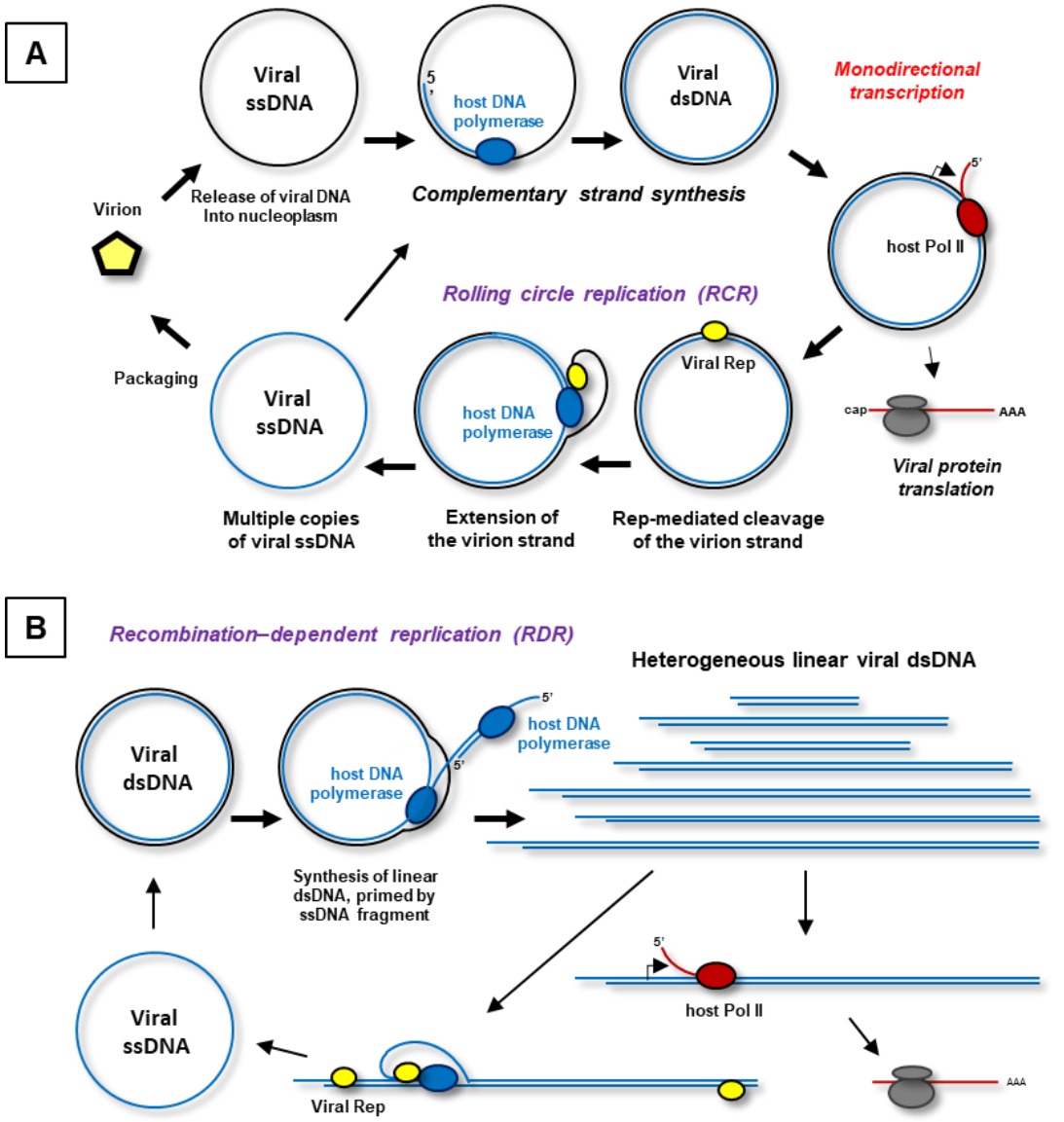


Figure 8. Models for rolling-circle replication (RCR) and recombination-dependent replication (RDR) of nanovirids and alphasatellites.

(A) Model for Rep-dependent RCR: 1) circular ssDNA is released from the virion into the host cell nucleus; 2) the complementary strand is synthesized by the host DNA polymerase; 3) the resulting circular dsDNA is used as a template for the host RNA polymerase II (Pol II) to transcribe viral mRNA which is transported to the cytoplasm for translation into viral protein (Rep in the case of DNA-R and alphasatellite DNA); 4) Rep protein is transported back to the nucleus to initiate replication of dsDNA (DNA-R or alphasatellite DNA) from which Rep-mRNA was transcribed and, in the case of nanovirids, trans-replication of other components. To this end, Rep binds the CR-SL and nicks the nonanucleotide sequence on the virion strand and gets attached to the 5' end of the nicked strand. The host DNA polymerase synthesizes a new virion strand by extending the existing virion strand at the 3'-end, using the complementary strand as a template. At each cycle of RCR, Rep makes a second nick at the nonanucleotide sequence and then re-ligates the displaced virion strand. The resulting multiple copies of circular ssDNA are either packaged in new virions or be converted to circular dsDNA for transcription and replication.

(B) Model for Rep-independent RDR: Figure adapted from Pooggin 2013. For legend, see the text. For more details, see Pooggin 2013.

In addition to RCR, geminivirids can also replicate via recombination-dependent replication (RDR) (Jeske *et al.*, 2001; Preiss & Jeske, 2003; Alberter *et al.*, 2005; Jovel *et al.*, 2007; Erdmann *et al.*, 2010). RDR is initiated by insertion of a small fragment of viral ssDNA into a circular dsDNA intermediate with the help of host recombination enzymes and is independent of viral Rep (Figure 8B). The fragment anneals to one of the strands of unwound dsDNA and serves as a primer for the host DNA polymerase synthesizing a daughter strand on the circular template strand for one or more rounds. The newly synthesized ssDNA is then converted to dsDNA by the host DNA polymerase, yielding linear dsDNA products of heterogeneous length. The linear partial dimer and multimer dsDNA products having two or more origins of replication can potentially serve as templates for Rep-dependent RCR, yielding circular ssDNA monomers, and for Pol II transcription of viral mRNAs (Figure 8B; Pooggin 2013). RCR and RDR mechanisms operate simultaneously during geminivirus infection (Jeske *et al.*, 2001; Preiss & Jeske, 2003). So far, RDR was not described for nanovirids.

Even though BBTV is a single virus species causing the bunchy top disease, its sequenced isolates are classified in two phylogenetic groups with distinct geographical delineation: the **Pacific and Indian Oceans** (**PIO**) group (including Hawaii, Tonga, Samoa, Fiji, Australia, Egypt, Cameroon, Gabon, Republic of Congo, Democratic Republic of Congo, Rwanda, Burundi, Malawi, India, Sri Lanka, Pakistan and Myanmar) and the **South East Asia** (**SEA**) group (including Japan, China, Vietnam, Philippines, Indonesia, Thailand and Taiwan) (Figure 9 - Karan *et al.*, 1994; Stainton *et al.*, 2015). Initially, this delineation was based on the pairwise sequence identity of DNA-R (DNA-1), because it is the least variable BBTV component, and therefore the easiest to amplify with generic primers and subsequently to sequence. The first sequence analysis of DNA-R isolates showed ca. 1-3% intra-group variability and ca. 10 % inter-group variability (Karan *et al.*, 1994). Based on this analysis, others studies used DNA-1 for demarcation of the phylogenetic groups, but it was also extended to DNA-3, and Karan et al. (1994) also proposed a demarcation threshold based on comparison of CR-M sequences (Wanitchakorn *et al.*, 2000b; Zhuang Jun & Zhi-Xin, 2005; Yu *et al.*, 2012; Banerjee *et al.*, 2014; Rao *et al.*, 2017). Although more and more studies classified BBTV isolates into the two groups (see Figure 5), additional efforts on phylogenetic analysis should be done to further understand the evolution of BBTV across space and time (Stainton *et al.*, 2015).

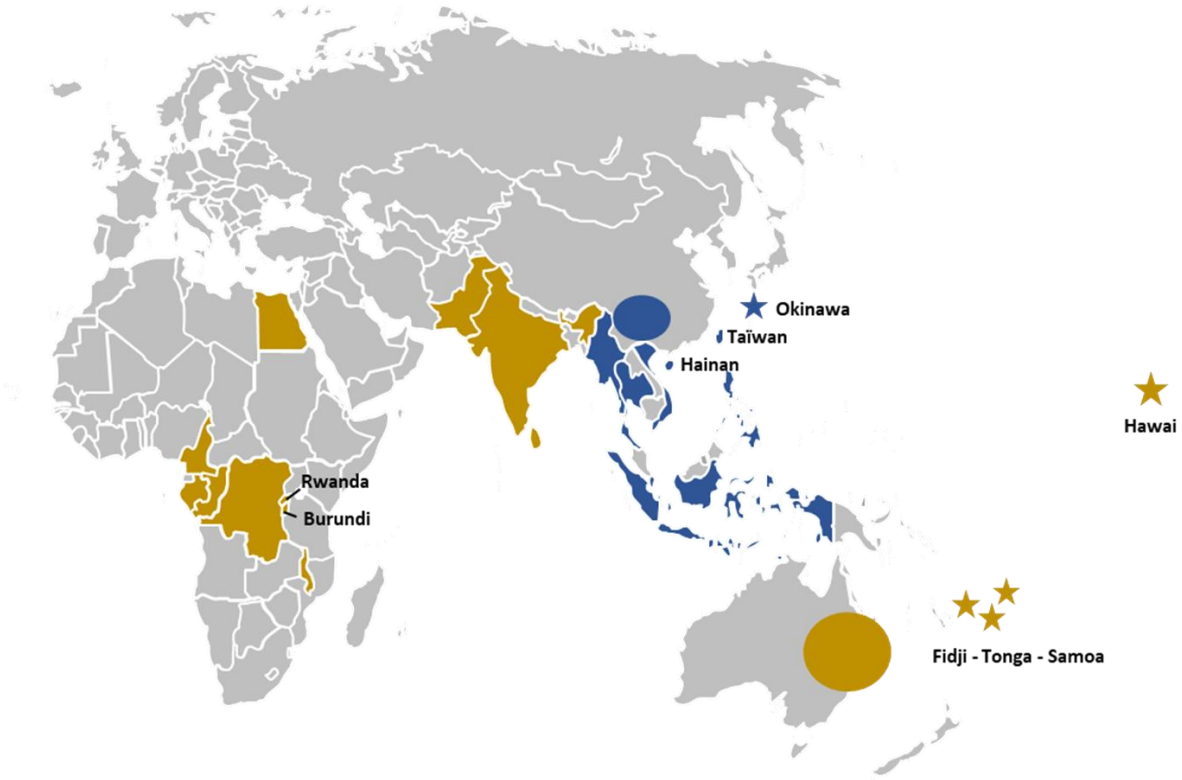


Figure 9. Geographic delineation of the SEA and PIO phylogenetic groups of BBTV.

- In gold: the Pacific and Indian Oceans (PIO) phylogenetic group, including Hawaii, Tonga, Samoa, Fiji, Australia, Egypt, Cameroon, Gabon, Republic of Congo, Democratic Republic of Congo, Rwanda, Burundi, Malawi, India, Sri Lanka, Pakistan and Myanmar.

- In blue: the South-East Asia (SEA) phylogenetic group, including Japan, China, Vietnam, Philippines, Indonesia, Thailand and Taiwan.

Free editable world map was taken from https://free-editable-worldmap-for-powerpoint-slides.fr.softonic.com/

Multipartite nature of BBTV

According to the Baltimore classification, there are seven distinct types of viral genomes: dsDNA, ssDNA, +ssRNA, -ssRNA, dsRNA, ssRNA-RT and dsDNA-RT (King *et al.*, 2011; Koonin *et al.*, 2021). Furthermore, viral genomes can be classified as monopartite (a single nucleic acid sequence encapsidated inside one virion), segmented (all genome segments encapsidated together inside one virion), and multipartite or multicomponent (genome components encapsidated separately, in group or not, but never all together in the same virion). Monopartite and segmented viruses are found in all kingdoms of cellular life, but multipartite viruses were so far found only in plants and fungi (although there exists a special case in mosquitoes) (Sicard *et al.*, 2016). Until now, a clear advantage for a virus to take the multipartite form during evolution was not found. However, it is not an unusual evolutionary event, thus so far this type of replicative form was adopted by ca. 30-40 % of plant viruses, and multipartite can be a virus with +ssRNA, -ssRNA, dsRNA and ssDNA genome. Some advantages of dividing viral genetic information in two or more separate components can be: i) adaptation of viral replication and transcription to the host, environment, and virulent stage, ii) facilitation of viral replication rate, iii) promotion of genetic exchange and reassortment. But all these putative benefits could be applied to both segmented and multipartite viruses. However, the cost is a clear consensus among scientific community: dividing its genetic information makes it more difficult for a virus i) to be acquired by a vector, ii) to be transmitted to a new host and iii) to invade a new cell, which are consistently reduced with an increasing number of segments or components (Sicard *et al.*, 2016). But recently, the last point, which was a paradigm among scientists, was broken by the demonstration that faba bean necrotic stunt virus (FBNSV, genus *Nanovirus*, family *Nanoviridae*), a multipartite virus with eight components, does not have all its components infecting the same cell, but instead the viral components are separated between cells, and probably exchange of viral proteins or mRNAs between cells to assure functional viral replication cycle (Sicard *et al.*, 2019). Like FBSNV, BBTV can potentially have a similar multicellular life, with viral components separated between companion cells of the phloem tissue. Nonetheless, both reassortment and intra-recombination events occur between BBTV components, increasing the viral genome sequence variability (Stainton *et al.*, 2012, 2015).

In addition to the six components of BBTV genome, circular ssDNA alphasatellites (family *Alphasatellitidae*) were found to be frequently associated with BBTV isolates in SEA. Alphasatellites encode a Rep-like protein that enables autonomous replication (Horser *et al.*, 2001). However, alphasatellite Rep is unable to mediate trans-replication of the *bona fide* genome components of its helper virus. Like other plant ssDNA virus-associated satellites, alphasatellites depend on their helper viruses for movement, encapsidation and plant-to-plant transmission (Briddon & Stanley, 2006). Interestingly, in the order Mulpavirales, there are two families of multipartite ssDNA viruses: *Nanoviridae* and *Metaxyviridae*. The latter family has currently only one genus (*Cofodevirus*) represented with a single species: *Coconut foliar decay virus* whose multipartite genome contains, instead of DNA-R, an alphasatellite coding for a master Rep (Gronenborn *et al.*, 2018). The role of alphasatellites in BBTV replication, transcription, antiviral defense evasion and transmission by the aphid vector is the subject of this PhD work.

Spoiler alert: Alphasatellitidae (review)

This draft review manuscript placed below as the last chapter of the thesis describes the history, diversity and evolution of alphasatellites associated with plant ssDNA viruses as well as their interactions with helper viruses in the host plant and the insect vector and the outcome of these interactions.

RNA interference-based antiviral defences

Viruses and other mobile genetic elements drove evolution of all cellular organisms since the Last Universal Common Ancestor (LUCA), obliging them to continuously evolve their defence systems (tenOever, 2016). To counter viruses and other invasive nucleic acids such as transposons or transgenes, plants and other eukaryotes have evolved the RNA interference (RNAi) and gene silencing system that also regulates endogenous expression and chromatin states (Ding & Voinnet, 2007; Borges & Martienssen, 2015; Pooggin, 2018). The plant RNAi and gene silencing pathways are mediated by small RNAs (sRNAs) classified into micro RNAs (miRNAs) and small interfering RNAs (siRNAs) which have similar chemical structures but differ in function and mode of biogenesis. Both miRNAs and siRNAs are generated by RNase III-like Dicer-like (DCL) family enzymes from double-stranded (ds)RNA precursors. In the model plant *Arabidopsis thaliana*, four distinct DCLs generate 21-nt (DCL4 and DCL1), 22-nt (DCL2) and 24-nt (DCL3) siRNAs that mediate defences against transposons, transgenes and DNA viruses via transcriptional and post-transcriptional gene silencing (TGS and PTGS, directed by 21-22 and 24-nt siRNAs, respectively) (Akbergenov *et al.*, 2006; Blevins *et al.*, 2006, 2011; Aregger *et al.*, 2012; Pooggin, 2013, 2018 - Figure 10). Both miRNAs and siRNAs are produced by DCLs in the form of duplexes with 2-nucleotide 3´-terminal overhangs and are methylated at 2´-OH groups of the 3´-terminal nucleotides by the methyltransferase HUA ENHANCER 1 (HEN1) to be protected from degradation. One of the two strands of sRNA duplex is loaded on an Argonaute (AGO) family protein based on sRNA size, 5'-terminal nucleotide and other sequence features to form an RNA induced silencing complex (RISC). RISCs are guided by sRNAs to complementary sequences of target genes and mediate their TGS and/or PTGS in a sequence-specific manner. TGS of transposons, transgenes and DNA viruses directed by 24-nt siRNA-RISCs involves RNA-directed DNA methylation (RdDM) at cytosines in all (CG, CHG and CHH) contexts (Pooggin, 2013; Borges & Martienssen, 2015). To establish successful infection, plant DNA viruses of the families *Geminiviridae* (ssDNA) and *Caulimoviridae* (dsDNA-RT) can suppress or evade RNAi and RdDM (reviewed in Pooggin, 2013, 2018). Little is known about interactions of ssDNA nanovirids (family *Nanoviridae*) and alphasatellites (family *Alphasatellitidae*) with the plant RNAi and RdDM machinery. For BBTV, it was shown that Clink, MP and CP encoded by DNA-C, DNA-M and DNA-S, respectively, are able to suppress transgene-induced PTGS in the model plant *Nicotiana benthamiana* (Niu *et al.*, 2009; Amin *et al.*, 2011). Nawaz-ul-Rehman et al. (2010) showed that alphasatellite Rep proteins encoded by two different

alphasatellites are able to suppress PTGS. Kumar et al. (2014) observed, by blot hybridization analysis, that accumulation of helper geminivirus-derived siRNAs was reduced in the presence of alphasatellite. Abbas et al. (2019) showed that alphasatellite Rep could suppress TGS in *N. benthamiana*. The mechanisms of silencing evasion by nanovirids and alphasatellites are likely to be similar to that proposed for geminivirids due to resemblance of their DNA replication mechanisms (Pooggin, 2013).

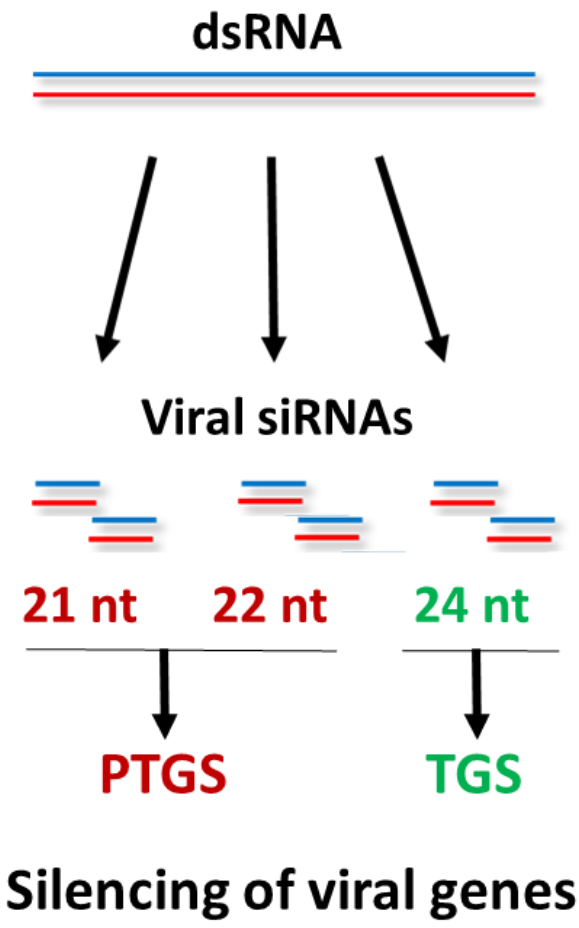


Figure 10. Mechanism of viral siRNA biogenesis and function in plants

Context and main objectives of the PhD thesis work

Illumina sequencing and bioinformatics analysis of BBTV isolates from Democratic Republic of the Congo (DRC) performed in the host institute team in 2017, before this PhD work had started, allowed to identify a novel alphasatellite. This is the first alphasatellite found to be associated with BBTV isolates from the PIO phylogenetic group. BLASTn analysis indicated that this alphasatellite shares the highest pairwise identity with alphasatellites from the genus *Fabenesatellite* associated with faba bean necrotic yellows virus (FBNYV, genus *Nanovirus*, family *Nanoviridae*) and only distantly related to alphasatellites associated with BBTV isolates from the SEA phylogenetic group, which belong to the *Babusatellite* and *Muscarsatellite* genera. During my Master II internship in the host institute team in 2018, this novel alphasatellite was shown to share less than 67 % pairwise identity with any fabenesatellite (the highest nucleotide sequence identity is 66.9 %) and therefore, based on genus demarcation criteria (Briddon *et al.*, 2018), it was classified in a new genus of the subfamily *Nanoalphasatellitinae*. This alphasatellite was maintained in BBTV-infected banana plants at the host institute greenhouse and we routinely performed its transmission from plant to plant by two clonal populations of *P. nigronervosa* originating from Congo-DRC and Gabon. We also found using immune-capture (IC)-PCR that the alphasatellite DNA is encapsidated by its helper virus coat protein, and our bioinformatics analysis of the Illumina sequencing data revealed that it is the most abundant viral DNA present in the majority of infected plants and viruliferous aphids.

Because the role of alphasatellites in helper virus infection and disease transmission is not well understood, particularly for the genus *Babuvirus* of *Nanoviridae*, and based on our preliminary results, the main objective of this PhD project was to explore an impact of the newly discovered alphasatellite on BBTV replication, transcription, suppression or evasion of antiviral RNAi and transmission by the aphid vector.

The discovery of the first BBTV alphasatellite in Africa (and PIO) raised the question on the prevalence and diversity of this and other alphasatellites associated with BBTV in PIO and SEA. To address this question, we took advantage of the field surveys conducted by the host team members and their collaborators in Sub-Saharan Africa (Congo-DRC, Gabon, Nigeria, Benin and Togo) and South-East Asia (Vietnam, Laos and China) and, using Illumina sequencing and bioinformatics, reconstructed the virome components of cultivated and wild

banana plants as well as non-banana plants grown in close vicinity of BBTV-infected banana plants.

The results of these studies are presented below in the form of three research articles – one just published in the peer-reviewed international open-access journal PLoS Pathogens with a high impact factor (Chapter 1/doi: 10.1371/journal.ppat.1010448) and two draft manuscripts to be finalized and submitted for publication as soon as possible (Chapters 2 and 3) – as well as a draft manuscript of the review article describing the history, diversity and evolution of alphasatellites (family *Alphasatellitidae*) and their impact on helper viruses (Chapter 4), which we plan to finalize after submission of the two research articles.

Chapter 1: Role of alphasatellites in helper virus replication, gene expression, antiviral RNAi evasion and transmission by insect vectors

Article 1 (published in PLoS Pathogens; doi: 10.1371/journal.ppat.1010448):

A newly emerging alphasatellite affects banana bunchy top virus replication, transcription, siRNA production and transmission by aphids

Valentin Guyot[1], **Rajendran Rajeswaran**[1,#], **Huong Cam Chu**[1], **Chockalingam Karthikeyan**[1], **Nathalie Laboureau**[1], **Serge Galzi**[1], **Lyna F.T. Mukwa**[2], **Mart Krupovic**[3], **P. Lava Kumar**[4], **Marie-Line Iskra-Caruana**[1,5], **Mikhail M. Pooggin**[1,*]

[1] PHIM Plant Health Institute, University of Montpellier, INRAE, CIRAD, IRD, Institute Agro, 34398 Montpellier, France

[2] Faculté des Sciences Agronomiques, Université Pédagogique Nationale, B.P. 141 Kinshasa, Democratic Republic of the Congo

[3] Institut Pasteur, Université de Paris, Archaeal Virology Unit, F-75015 Paris, France

[4] International Institute of Tropical Agriculture (IITA), Oyo Road, PMB 5320, Ibadan, Nigeria

[5] CIRAD, DGD-RS, 34398 Montpellier, France

*Correspondence: Dr. Mikhail M. Pooggin; mikhail.pooggin@inrae.fr

present address: Swiss Federal Institute of Technology, Zurich (ETH-Zurich), Department of Biology, Zurich, Switzerland

Co-author e-mail addresses: valentin.guyot@inrae.fr; r.rajeswar@gmail.com; chucamhuong224@gmail.com; karthikbiomdu@gmail.com; nathalie.laboureau@cirad.fr; lynamukwa@yahoo.fr; serge.galzi@cirad.fr; mart.krupovic@pasteur.fr; L.kumar@cgiar.org; Marie-Line.Caruana@Cirad.Fr

Abstract

Banana bunchy top virus (BBTV) is a six-component ssDNA virus (genus *Babuvirus*, family *Nanoviridae*) transmitted by aphids, infecting monocots (mainly species in the family Musaceae) and likely originating from South-East Asia where it is frequently associated with self-replicating alphasatellites. Illumina sequencing analysis of banana aphids and leaf samples from Africa revealed an alphasatellite that should be classified in a new genus, phylogenetically related to alphasatellites of nanoviruses infecting dicots. Alphasatellite DNA was encapsidated by BBTV coat protein and accumulated at high levels in plants and aphids, thereby reducing helper virus loads, altering relative abundance (formula) of viral genome components and interfering with virus transmission by aphids. BBTV and alphasatellite clones infected dicotyledonous *Nicotiana benthamiana*, followed by recovery and symptomless persistence of alphasatellite, and BBTV replication protein (Rep), but not alphasatellite Rep, induced leaf chlorosis. Transcriptome sequencing revealed 21, 22 and 24 nucleotide small interfering (si)RNAs covering both strands of the entire viral genome, monodirectional Pol II transcription units of viral mRNAs and pervasive transcription of each component and alphasatellite in both directions, likely generating double-stranded precursors of viral siRNAs. Consistent with the latter hypothesis, viral DNA formulas with and without alphasatellite resembled viral siRNA formulas but not mRNA formulas. Alphasatellite decreased transcription efficiency of DNA-N encoding a putative aphid transmission factor and increased relative siRNA production rates from Rep- and movement protein-encoding components. Alphasatellite itself spawned the most abundant siRNAs and had the lowest mRNA transcription rate. Collectively, following African invasion, BBTV got associated with an alphasatellite likely originating from a dicot plant and interfering with BBTV replication and transmission. Molecular analysis of virus-infected banana plants revealed new features of viral DNA transcription and siRNA biogenesis, both affected by alphasatellite. Costs and benefits of alphasatellite association with helper viruses are discussed.

Author Summary

Self-replicating alphasatellites are frequently associated plant ssDNA viruses. Their origin and costs versus benefits for helper virus replication, antiviral defense evasion and transmission by insect vectors are poorly understood. Here we describe identification in Africa and in depth molecular and biological characterization of a newly emerging alphasatellite of BBTV, a multicomponent ssDNA babuvirus causing one of the most economically-important diseases of monocotyledonous bananas and plantains. Phylogenetically, this alphasatellite represents a novel genus and is more related to alphasatellites of nanoviruses infecting dicot hosts than to other BBTV alphasatellites previously identified only in Asia. Consistent with its hypothetical dicot origin, cloned alphasatellite and BBTV can establish systemic infection in a model dicot plant, followed by recovery and symptomless alphasatellite persistence. In banana plants, alphasatellite competes for the host replication and transcription machinery and accumulates at high levels, thereby reducing loads of the helper virus, modifying relative abundance of its components and interfering with its acquisition and transmission by aphids. On the other hand, plant antiviral defences silence alphasatellite gene expression at both transcriptional and posttranscriptional levels, generating highly-abundant 21, 22 and 24 nucleotide small interfering RNAs, suggesting that alphasatellite may serve as a decoy protecting its helper virus from gene silencing.

Introduction

Bananas and plantains (genus *Musa*, family *Musaceae*) are monocots susceptible to many diseases and pests. Banana bunchy top disease (BBTD) is one of the most economically important, causing up to 100% yield losses (Dale 1987; Kumar et al. 2015). BBTD symptoms include discontinuous dark-green streaks on the leaf veins, midrib, petioles and pseudostem, chlorosis of leaf margins and strong reduction in size of leaves and petioles, resulting in a foliar bunch at the top of stunted plant (Dale 1987; Hooks et al. 2008). BBTD is caused by banana bunchy top virus (BBTV) which is restricted to phloem tissues and transmitted by the banana aphid *Pentalonia nigronervosa* in a persistent circulative non-propagative manner (Magee 1940; Hafner et al. 1995; Hu et al. 1996; Watanabe et al. 2013). BBTV host range was so far found to be restricted to monocots from the families *Musaceae* and *Zingiberaceae* (Rahayuniati et al. 2021).

BBTV has a six-component circular single-stranded (ss)DNA genome and is classified within the genus *Babuvirus* of family *Nanoviridae* (Burns et al. 1994; Burns et al. 1995; Gronenborn 2004; Thomas et al. 2021). Each circular ssDNA component of the BBTV genome has a size of ca. 1.0 to 1.1 Kb and is individually encapsidated within isometric 18-20 nm virions (Wu and Su 1990). BBTV replicates in the nucleus via a rolling circle mechanism mediated by viral replication protein (Rep) and the host replication machinery. Each BBTV component, namely DNAs C, M, N, R, S and U3, possesses a single ORF preceded by the host RNA polymerase II (Pol II) promoter region with a TATA or TATA-like box and followed by the Pol II terminator region with a polyadenylation signal (Beetham et al. 1997; Beetham et al. 1999; Herrera-Valencia et al. 2007). DNA-R encodes a master Rep protein mediating replication of DNA-R itself and trans-replication of other components (Beetham et al. 1997; Horser et al. 2001a). DNA-C encodes a cell-cycle link protein (Clink) enhancing viral replication (Aronson et al. 2000; Lageix et al. 2007). DNA-M encodes a movement protein (MP) (Wanitchakorn et al. 2000). DNA-S encodes a coat protein (CP) encapsidating all viral ssDNA components (Wanitchakorn et al. 1997). DNA-N encodes a nuclear shuttle protein (NSP) (Wanitchakorn et al. 2000), whose homologue is required for aphid transmission of eight-component ssDNA viruses from the genus *Nanovirus* of family *Nanoviridae* (Grigoras et al. 2018; Di Mattia et al. 2020). DNA-U3 encodes a small protein of unknown function (Beetham et al. 1999). All six components contain two regions of high sequence similarity: a common region stem-loop (CR-SL) and a common region major (CR-M) (Burns et al. 1995). The CR-SL is an origin of replication containing a highly conserved nonanucleotide sequence TATTATTAC and three 5-

bp-long repeats (iterons) that are likely involved in specific recognition and binding of the master Rep (Burns et al. 1995; Herrera-Valencia et al. 2006). The CR-M is thought to be involved in regulation of transcription (Burns et al. 1995) and also contains primer-binding sites for complementary strand DNA synthesis (Hafner et al. 1997).

BBTD was first recorded in 1889 in Fiji and then found throughout Asia, Oceania and Africa (Dale 1987; Stainton et al. 2015). Sequenced BBTV isolates are classified into two phylogenetic groups with a distinct geographical delineation: the Pacific and Indian Oceans (PIO) group (Australia, Africa, India, Sri Lanka, Tonga, Hawaii) and the South-East Asia (SEA) group (Vietnam, Philippines, Indonesia, Taiwan, China, Japan) (Karan et al. 1994; Kumar et al. 2011; Stainton et al. 2015). SEA is a diversity hotspot and most likely origin of BBTV (Stainton et al. 2015) from where the virus was exported to other regions of the world, following banana domestication between 7,000 and 10,000 years ago (Denham et al. 2003; Perrier et al. 2011), via long-distance movements of virus-infected propagules. The first of such movements was from SEA to India approximately 1,000 years ago and the Indian subcontinent was in turn an origin of two movement events to Sub-Saharan Africa between 1825 and 1934 (Stainton et al. 2015). The SEA isolates of BBTV are frequently associated with alphasatellites (Wu 1994; Wu et al. 1994; Yeh et al. 1994; Bell et al. 2002; Horser et al. 2001b; Tsao 2008; Fu et al. 2009; Yu et al. 2012; Yu et al. 2019). Alphasatellites (family *Alphasatellitidae*) have a circular ssDNA genome encoding a Rep protein that enables autonomous replication of alphasatellite DNA but cannot trans-replicate BBTV components (Horser et al. 2001a). Similar to other plant ssDNA virus-associated satellites (betasatellites and deltasatellites; family *Tolecusatellitidae*), alphasatellites depend on their helper virus for movement, encapsidation and plant-to-plant transmission (Briddon and Stanley 2006). Alphasatellites are associated with helper viruses of the *Babuvirus* and *Nanovirus* genera of family *Nanoviridae*, the *Cofodevirus* genus of *Metaxyviridae*, and the *Begomovirus* and *Mastrevirus* genera of *Geminiviridae* and are classified into subfamilies *Nanoalphasatellitinae*, *Petromoalphasatellitinae* and *Geminialphasatellitinae* (Briddon et al. 2018; Varsani et al. 2021). *Nanoalphasatellitinae* and *Petromoalphasatellitinae* alphasatellites have a size of ca. 1.1 Kb, similar to a component size of their helper nano-, cofode- and babuviruses, whereas *Geminialphasatellitinae* ones have a size of ca. 1.3 Kb, approximately a half-size of their helper geminivirids. Notably, *Petromoalphasatellitinae* alphasatellites are associated with BBTV, cardamom bushy stunt virus (another member of the genus *Babuvirus*) and coconut foliar decay virus (genus

Cofodevirus) which all infect monocots, while *Nanoalphasatellitinae* alphasatellites are associated with helper viruses of the genus *Nanovirus* which infect dicots.

Impacts of alphasatellites on BBTV infection and disease transmission by aphids have not been studied so far. In the case of related nanovirids, an alphasatellite associated with faba bean necrotic yellows virus (genus *Nanovirus*) had a negative impact on helper virus infectivity (reducing the number of infected plants following agroinoculation with viral infectious clones) and reduced disease severity (Timchenko et al. 2006). Another study in the same pathosystem, however, revealed no effect of the alphasatellite on disease severity and demonstrated that the transmission rate of the helper virus by aphids was increased in the presence of alphasatellite, despite a substantial reduction of helper virus DNA loads in aphids (Mansourpour et al. 2021). Alphasatellites associated with geminivirids were also reported to affect helper virus accumulation and transmission as well as disease symptom development (Wu and Zhou 2005; Nawaz-ul-Rehman et al. 2010; Idris et al. 2011; Kumar et al. 2014; Mar et al. 2017; Nogueira et al. 2021). Thus, alphasatellites associated with bipartite begomoviruses (genus *Begomovirus*) increased symptom severity, affected helper virus accumulation and interfered with virus transmission by the whitefly *Bemisia tabaci* (Mar et al. 2017; Nogueira et al. 2021). Interestingly, an alphasatellite associated with a monopartite geminivirus of the genus *Mastrevirus* enhanced disease severity in monocot wheat plants, which coincided with increased accumulation of helper virus DNA and decreased accumulation of helper virus-derived small interfering (si)RNAs (Kumar et al. 2014). However, an alphasatellite of a monopartite geminivirus of the genus *Begomovirus* did not affect siRNA production from its helper virus or betasatellite in *N. benthamiana* (Amin et al. 2011a). In the latter study, alphasatellite-derived siRNAs were also detected by blot hybridization, indicating that both helper virus and alphasatellite are targeted by the plant antiviral defense based on RNA interference (RNAi) (Amin et al. 2011a).

The RNAi machinery has evolved in most eukaryotes to regulate endogenous gene expression and to defend against invasive nucleic acids such as viruses, satellites, viroids, transposons and transgenes (Pooggin 2018). The plant RNAi pathways are mediated by small (s)RNAs classified into siRNAs and micro (mi)RNAs, which are generated by Dicer-like (DCL) family proteins from double-stranded (ds)RNA precursors and are predominantly 21, 22 and 24 nucleotide (nt) long. All sRNA types and size classes bind Argonaute (AGO) family proteins and guide the resulting silencing complexes to repress target gene expression post-transcriptionally and/or transcriptionally in a sequence-specific manner. RNA-directed DNA

methylation (RdDM) is one of the plant RNAi pathways operating in the nucleus to establish and maintain transcriptionally-silent chromatin via *de novo* cytosine methylation directed by 24 nt siRNAs (Matzke et al. 2015; Borges and Martienssen 2015). DNA virus-infected plants accumulate 24 nt viral siRNAs that can potentially silence viral DNA transcription via RdDM as well as 21 and 22 nt viral siRNAs that can potentially silence viral mRNAs post-transcriptionally through their cleavage and degradation or translational repression (Akbergenov et al. 2006; Blevins et al. 2006; Blevins et al. 2011; Yang et al. 2011; Aregger et al. 2012). Yet, DNA viruses of the families *Geminiviridae* and *Caulimoviridae* (reverse-transcribing dsDNA viruses) can evade and/or suppress RNAi (reviewed in Pooggin 2013; Pooggin 2018). Little is known about interactions of nanovirids with the plant RNAi machinery. For BBTV, MP and CP of a SEA isolate and Clink and MP of a PIO isolate were reported to suppress transgene silencing in *N. benthamiana* when expressed from an RNA virus vector (Niu et al. 2009; Amin et al. 2011b). Mechanisms of RNAi evasion by nanovirids can be similar to those proposed for geminivirids (Pooggin 2013) due to resemblance in their replication mechanisms (Gronenborn 2004). Interestingly, Rep proteins of begomoviral alphasatellites suppressed post-transcriptional transgene silencing in *N. benthamiana* (Nawaz-ul-Rehman et al. 2010) and, when expressed from an RNA virus vector, could restore expression of a transcriptionally silenced transgene in *N. benthamiana* (Abbas et al. 2019).

Here we describe the identification and molecular and biological characterization of a first alphasatellite associated with BBTV isolates from the PIO phylogenetic group. It was identified in Democratic Republic of the Congo (DRC), one of the first countries in Sub-Saharan Africa where BBTD was introduced and first reported in Kisangani region of Tshopo province in 1958 (Kumar et al. 2011; Mukwa et al. 2014). This alphasatellite represents a new genus of the family *Alphasatellitidae*, related to several genera of the subfamily *Nanoalphasatellitinae*. We demonstrate that BBTV and alphasatellite clones can infect the dicot *N. benthamiana*, followed by recovery and symptomless persistence of the alphasatellite. We also show that the alphasatellite reduces BBTV accumulation in banana plants and aphids and interferes with virus transmission by aphids. By transcriptome and sRNA-ome sequencing we uncovered new features of Pol II transcription of viral mRNAs and biogenesis of viral siRNAs, both affected by the alphasatellite.

Results and Discussion

Discovery and molecular characterization of the first alphasatellite associated with BBTV in Africa

Banana aphids (*P. nigronervosa*) collected in December 2016 on a bunchy top diseased banana plantain (genotype AAB) in the Boko village of Congo-Central province of DRC were placed on healthy seedlings of *Musa acuminata* Cavendish (genotype AAA) for disease transmission under laboratory conditions (see Materials and Methods). All the recipient plants eventually developed characteristic BBTD symptoms (**S1A Fig**), while no infection was obtained with aphids collected on a symptomless banana plant in the same field at the same time. PCR and immuno-capture (IC) PCR analyses with DNA-R specific primers confirmed the presence of BBTV in both the viruliferous aphids and the infected plants, respectively. One of the infected plants (numbered p4.3) which had been the first to exhibit disease symptoms was then used as a source plant for BBTD transmission tests. Both the viruliferous DRC aphids born on the source plant and the virus-free aphids from DRC and Gabon (GAB aphids) fed for 24 hrs on a detached leaf of the source plant transmitted the disease to recipient plants (**S1B-C Fig**) with a latency period of 3 to 9 weeks. A single DRC aphid placed on a recipient plant leaf was also able to transmit the virus (**S1D-E Fig**).

We then undertook *de novo* reconstruction of a complete viral genome from the infected plants and aphids using rolling circle amplification (RCA) of viral DNA followed by Illumina sequencing. RCA products were verified by digestion with AvaI (**S1G-H Fig**), the enzyme with a conserved restriction site in the BBTV CR-SL of previously characterized isolates from DRC (Stainton et al. 2015; Mukwa et al. 2016). Based on AvaI restriction analysis, the following 11 samples were selected for Illumina sequencing of undigested RCA products: (i) two leaf samples from the source plant p4.3 taken at two time points (one month apart), along with two pools of DRC aphids (8-10 adults) collected from this plant at the respective time points, (ii) two leaf samples taken from the recipient plants following disease transmission by DRC and GAB aphids, along with two pools of respective aphids (8-10 adults), and (iii) single DRC aphids from three recipient plants (**S2 Fig**). The Illumina reads were *de novo* assembled into contigs and their consensus sequences were verified as described in Material and Methods. BLASTn analysis of the resulting contigs revealed that they represent terminally-redundant sequences of the six circular components of the BBTV genome in all 11 samples. The reconstructed BBTV components from the source plant p4.3 shared high sequence similarity to

respective components of other BBTV isolates from DRC, such as Mbk-24 collected in 2012 (Mukwa et al. 2016): 99.0 % pairwise identity in DNA-C (KU759879), 99.4 % in DNA-M (KU759878); 99.5 % in DNA-N (KU759880), 99.1 % in DNA-R (KU687050), 99.5 % in DNA-S (KU759877) and 98.9 % in DNA-U3 (KU759869). Single nucleotide polymorphism (SNP) analysis revealed that at both time points the source plant p4.3 and all the aphid samples from this plant (including the single aphid) contained a single variant for DNAs C, N and R and two variants (v1 and v2) for DNAs M, S and U3, which differed by 3, 6 and 2 SNPs, respectively (**S1A Dataset**). In all cases the less abundant variant v2 was supported by substantial and comparable proportions of reads in all the plant and aphid samples (**S3 Fig**), indicating the stability of BBTV quasispecies population in the plant and its viruliferous aphids' colony over one month period between the sampling time-points. However, BBTV quasispecies had slightly evolved upon transmission to new plants. Following DRC aphid transmission, the recipient plant and the aphids' progeny contained only one of the two variants of DNAs M and S (v1 in both cases), while the variant v2 of DNA-U3 became more abundant than the variant v1 (**S3 Fig**). Following GAB aphid transmission, both variants of DNAs M and S were present in the recipient plant and the aphids' progeny, albeit at altered ratios, while the two variants of DNA-U3 were replaced by a single variant (v3) derived from v2 (8 new SNPs) (**S3 Fig**; **S1A Dataset**).

Besides BBTV, 7 of the 11 sequenced samples (JGF-1, JGF-2, JGF-4, JGF-5, JGF-6, JGF-7, JGF-10) contained terminally-redundant contigs representing a 1.1 Kb genome of an alphasatellite, most closely related to four alphasatellites (AJ005966, AJ132187, MF510474, MF510475) associated with faba bean necrotic yellows virus (genus *Nanovirus*) (BLASTn: 67.4-68.7 % pairwise identity within 69-75 % query coverage; **S1D Dataset**). All the 7 samples including the leaf and aphid samples from the source plant p4.3 and two samples from recipient plants contained a single variant of this alphasatellite, indicating its genetic stability before and after transmission. Mapping of Illumina reads to the alphasatellite sequence revealed its presence in one of the remaining 4 samples (JGF-8), while 3 others contained very low numbers of alphasatellite reads (**S4 Fig**) that may represent cross-contamination during Illumina sequencing of multiplexed libraries in one flowcell.

To validate the Illumina sequencing results, we amplified by PCR and cloned the alphasatellite and six BBTV components from the aphid pool of the source plant p4.3 at the first sampling time-point (**S2 Fig**, sample JGF-5). Sanger sequencing of the resulting full-length clones confirmed the *de novo* assembled viral sequences and the presence of two variants (v1 and v2) for DNAs M, S and U3. We then designed specific diagnostic primers for each BBTV

component and alphasatellite and performed single, duplex and multiplex PCR analysis of all the 11 samples, which confirmed the presence of respective components in each sample (**Fig S5A-B**). Consistent with Illumina read counts (**S4 Fig**), semi-quantitative multiplex PCR confirmed that alphasatellite is one of the most abundant virome components in the source plant p4.3 and its aphid colony and that relative abundance of BBTV components is altered when alphasatellite loads are reduced (**S5B Fig**).

To evaluate prevalence of the newly discovered alphasatellite in DRC and other countries of Sub-Saharan Africa, we performed PCR screening of BBTV-infected banana leaf samples collected in DRC in 2012 (n = 401) (Mukwa et al. 2014; Mukwa et al. 2016), Gabon in 2016 (n = 6), Nigeria in 2019 (n = 16), Togo in 2019 (n = 5) and Benin in 2018-2020 (n = 33). One of the samples from DRC (AAB plantain, Kimpoko village, Congo-Central province) was found to be PCR-positive for alphasatellite, whereas all other samples scored either PCR-negative, or yielded very low abundance or diffuse products (**S6A-D Fig**). Furthermore, BBTV-infected banana plants originating from Malawi and New Caledonia (maintained in CIRAD greenhouse, Montpellier) were also found to be alphasatellite-negative. IC-PCR analysis of the leaf tissues from selected DRC-2012 samples confirmed the presence of alphasatellite in the PCR-positive sample and demonstrated encapsidation of its DNA by BBTV coat protein, similar to DNA of DRC-2016 alphasatellite in BBTV-infected Cavendish (**S6A Fig**). Illumina sequencing of RCA-amplified viral DNA of selected samples from DRC (n = 5), Gabon (n = 3), Benin (n = 1), Malawi (n = 1) and New Caledonia (n = 1) confirmed the presence of alphasatellite in the PCR-positive DRC-2012 sample, while other samples did not contain any alphasatellite. A *de novo* assembled sequence of the DRC-2012 alphasatellite shared 98.1 % identity with the sequence of DRC-2016 alphasatellite: the differences include 1-, 6- and 10-nt indels in the non-coding sequence and 18 SNPs, 3 of which resulting in 3 amino-acid substitutions in the Rep protein (**S1C Dataset**). The BBTV components of DRC-2012 and DRC-2016 isolates (**S1A-B Dataset**) were also found to differ: 98.8 % identity in DNA-C, 99.3 % in DNA-M, 99.3 % in DNA-N, 99.5 % in DNA-R, 99.5 % in DNA-S, and 97.9 % in DNA-U3. These findings indicate that the alphasatellite got associated with BBTV in DRC before its sampling in 2012 and then persisted in this country and evolved together with the helper virus in the following years, most likely via aphid transmission as we demonstrated under the laboratory conditions. The two villages in DRC where the alphasatellite was found are located ca. 100 km away from each other. So far, we obtained no evidence for this alphasatellite to spread in other countries, suggesting its recent emergence, which is also supported by its extremely low prevalence in the

samples collected in DRC in 2012. Further surveys in DRC and elsewhere in the BBTV-affected areas are required to assess alphasatellite incidence and prevalence in BBTV-infected bananas in Sub-Saharan Africa.

Sequence comparison and phylogenetic analysis of the DRC alphasatellite

Pairwise sequence comparisons (**S7 Fig**) and phylogenetic analysis (**Fig 1**) of all alphasatellite sequences available in GenBank in September 2021 plus the DRC-2012 and DRC-2016 isolates revealed that, based on current species and genus demarcation criteria (81 % and 68 % sequence identity, respectively) the DRC alphasatellites represent a novel species that belongs to a new genus falling deeply within the subfamily *Nanoalphasatellitinae* rather than *Petromoalphasatellitinae* (**Fig 1**). The DRC alphasatellite isolates share only 56.7 to 62.8 % identity with BBTV SEA-associated alphasatellites and other alphasatellites of the subfamily *Petromoalphasatellitinae* (**S7A Fig**), and share higher identities (64.0 to 67.7 %) with isolates of faba bean necrotic yellows alphasatellite 2 (FBNYA2) of the subfamily *Nanoalphasatellitinae* (**S7B Fig**). FBNYA2 is the only member of genus *Fabenesatellite* and is associated with four different viruses of the genus *Nanovirus* which infect dicots (Timchenko et al. 1999; Heydarnejad et al. 2017; Kraberger et al. 2018; Vetten et al. 2019; Hassan-Sheikhi et al. 2020; Hasanvand et al. 2021).

Sequence similarity-based clustering analysis of alphasatellite Rep proteins showed (**S8 Fig**) that the DRC alphasatellite is connected to the alphasatellites from *Fabenesatellite*, *Mivedwarsatellite*, *Subclovsatellite* and *Clostunsatellite* genera of the subfamily *Nanoalphasatellitinae*, all associated with the nanoviruses infecting dicot Fabaceae hosts. It is also connected to the alphasatellites from the genus *Gosmusatellite* of *Geminialphasatellitinae,* which are associated with the geminivirids infecting dicot Malvaceae and Asteraceae hosts. In contrast, the DRC alphasatellite is not directly connected to any previously identified alphasatellites of the *Babusatellite* and *Muscarsatellite* genera infecting monocot Musaceae and Zingiberaceae hosts in complexes with their helper babuviruses BBTV [banana bunchy top alphasatellite 1 (BBTA1), BBTA2 and BBTA3; Wu 1994; Wu et al. 1994; Yeh et al. 1994; Horser et al. 2001b; Bell et al. 2002; Tsao 2008; Fu et al. 2009; Fu et al. 2009; Yu et al. 2012;Yu et al. 2019] and cardamom bushy dwarf virus (Mandal et al. 2013) (**S8 Fig**).

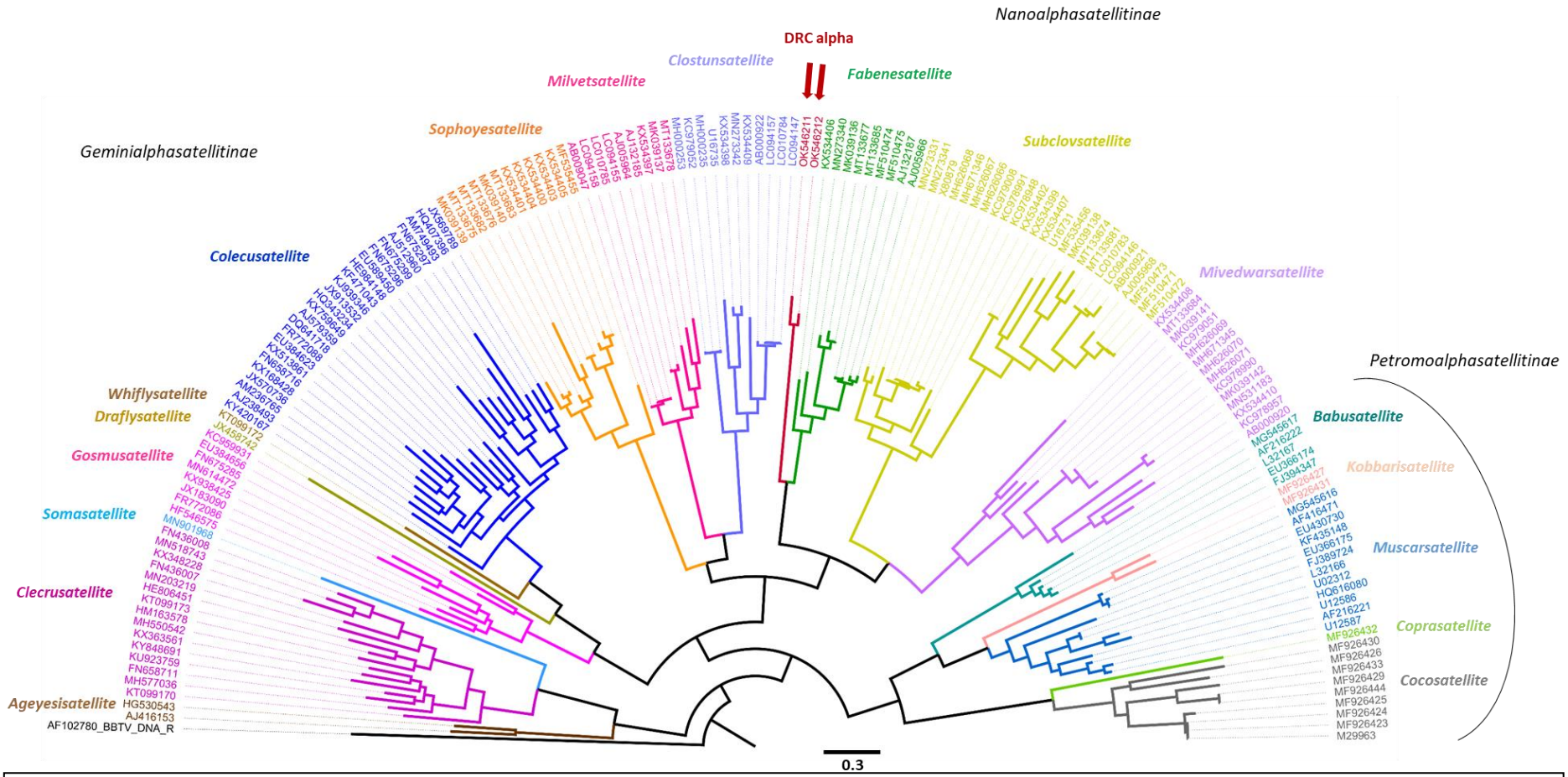


Fig 1. Phylogenetic analysis of DRC alphasatellite.

A maximum likelihood phylogenetic tree of complete nucleotide sequences of the two isolates of DRC alphasatellite and all isolates of other alphasatellites associated with helper viruses of the families *Nanoviridae*, *Metaxyviridae* and *Geminiviridae* is rooted with BBTV DNA-R. The Genbank accession number is given for each alphasatellite. Alphasatellites' genera are color-coded and their subfamilies delineated. Position of the DCR alphasatellite's genus (*Banaphisatellite*) is indicated with red arrow. Note that the subfamily *Petromoalphasatellitinae* comprises all previously-identified BBTV alphasatellites representing 3 species, namely, BBTA1 (L32167, Wu 1994; AF216222, Horser et al. 2001b; FJ394347, Fu et al. 2009; MG545617, Yu et al. 2019; EU366174, unpublished), BBTA2 (AF416471, Bell et al. 2002; EU430730, Tsao 2008; MG545616, Yu et al. 2019) and BBTA3 (L32166, Wu 1994; U02312, Yeh et al. 1994; AF216221, Horser et al. 2001b; HQ616080, Yu et al. 2012; FJ389724, Fu et al. 2009; EU366175, U12586 and U12587, unpublished), and alphasatellites of cardamom bushy dwarf virus (CBDA: KF435148; Mandal et al. 2013) and coconut foliar decay virus (CFDA1: M29963, MF926423, MF926424, MF926425; CFDA2: MF926426; CFDA3: MF926427, MF926431; CFDA4: MF926429, MF926444, MF926433; CFDA5: MF926430; CFDA7: MF926432; Gronenborn et al. 2018), which belong to the genera *Babusatellite* (BBTA1), *Muscarsatellite* (BBTA2, BBTA3, CBDA), *Cocosatellite* (CFDA1, CFDA2, CFDA4, CFDA5), *Coprasatellite* (CFDA7) and *Kobbarisatellite* (CFDA3). The genus *Fabenesatellite* of subfamily *Nanoalphasatellitinae*, the closest to BBTV DRC alphasatellite's genus, comprises variants of faba been necrotic yellows alphasatellite 2 associated with faba bean necrotic yellows virus (AJ005966, AJ132187, MF510474, MF510475; Timchenko et al. 1999; Kraberger et al. 2018), sophora alopecuroides yellow stunt virus (KX534406; Heydarnejad et al. 2017), milk vetch chlorotic dwarf virus (MN273330=MN273340; Hassan-Sheikhi et al. 2020) and parsley severe stunt-associated virus (MK039136, MT133677, MT133685; Vetten et al. 2019; Hasanvand et al. 2021).

Taken together, two DRC alphasatellite isolates, which we name here banana bunchy top alphasatellite 4 (BBTA4) isolate DRC-2016 and BBTA4 isolate DRC-2012, belong to a novel species representing a new genus in the subfamily *Nanoalphasatellitinae* and likely originating from a dicot plant. We propose to name this new genus "*Banaphisatellite*" (banana aphid satellite), which would highlight discovery of BBTA4 in both banana and banana aphids and its ability to be transmitted by the aphids as well as to distinguish it from the genus *Babusatellite* of subfamily *Petromoalphasatellitinae* which includes the first alphasatellite species (BBTA1) found to be associated with BBTV in SEA (Wu 1994; Wu et al. 1994). To conform a binominal nomenclature recently adopted by the ICTV (Walker et al. 2021) we propose to name this new alphasatellite species *Banaphisatellite musae1*.

BBTV and DRC alphasatellite can establish systemic infection in the dicot plant *Nicotiana benthamiana*

To test the hypothesis on the dicot plant origin of DRC alphasatellite we constructed infectious clones of the alphasatellite and its helper BBTV. To this end the monomeric clones of six BBTV components (DNA-C, OK546213; DNA-M_v1, OK546214; DNA-N, OK546215; DNA-R, OK546216; DNA-S_v1, OK546217; DNA-U3_v1, OK546218) and alphasatellite (OK546211) from the DRC-2016 aphids were used for construction of partial or, in the case of DNA-R, complete dimer clones (**S9A-B Fig**). The resulting constructs were mobilized to *Agrobacterium tumefaciens* for agro-inoculation (via leaf infiltration) of *N. benthamiana* seedlings. Six days post-inoculation (dpi), plants agroinoculated with the BBTV +/- alphasatellite constructs started to develop disease symptoms including chlorosis and downward curling of systemic leaves (**Fig 2C**). Semi-quantitative duplex PCR analysis of total DNA extracted from the symptomatic leaves at 8 dpi confirmed accumulation of BBTV DNA-R and also alphasatellite DNA in the case of its co-inoculation. Alphasatellite DNA accumulated at higher levels than DNA-R (**Fig 2A**), which is consistent with our findings in banana plants (**S5B Fig**), suggesting that alphasatellite is capable of more efficient self-replication than BBTV DNA-R. In line with this hypothesis, an alphasatellite Rep was ~10 times more active than a master Rep of its helper nanovirus in a replication origin cleavage and nucleotidyl-transfer reaction *in vitro* (Timchenko et al. 1999). Comparative analysis of systemically-infected leaves of agroinoculated *N. benthamiana* (at 21 dpi) and aphid-inoculated *M. acuminata* Cavendish using IC-PCR (immuno-capture of viral particles, followed by DNase treatment and duplex PCR analysis of

remaining encapsidated DNA) revealed that both DNA-R and alphasatellite DNA are encapsidated by BBTV coat protein in both plant species (**Fig 2B**).

Two months post-inoculation all the infected *N. benthamiana* plants appeared to recover from infection as newly growing leaves and eventually flowering organs did not exhibit any clear disease symptoms. These findings of systemic infection followed by recovery were reproduced in two independent agroinfiltration experiments with BBTV +/- alphasatellite constructs. PCR analysis of RCA-amplified total DNA extracted from upper leaves of 9 plants co-inoculated with BBTV and alphasatellite and recovered from disease symptoms, revealed that, in 6 plants, all BBTV components were below detection, while in the remaining 3 plants, five, four or two of the six components were readily detectable (**Fig 2D**; plants p3, p7 and p4, respectively). Thus, the recovery coincided with diminished accumulation of one or more components of BBTV genome. Remarkably, alphasatellite DNA was detected in four of the nine plants, three of which being PCR negative for any component of the helper virus (**Fig 2D**; plants p2, p6 and p8). This indicates that DRC alphasatellite can persist in asymptomatic leaf tissues even in the absence of helper virus, likely owing to its ability to self-replicate.

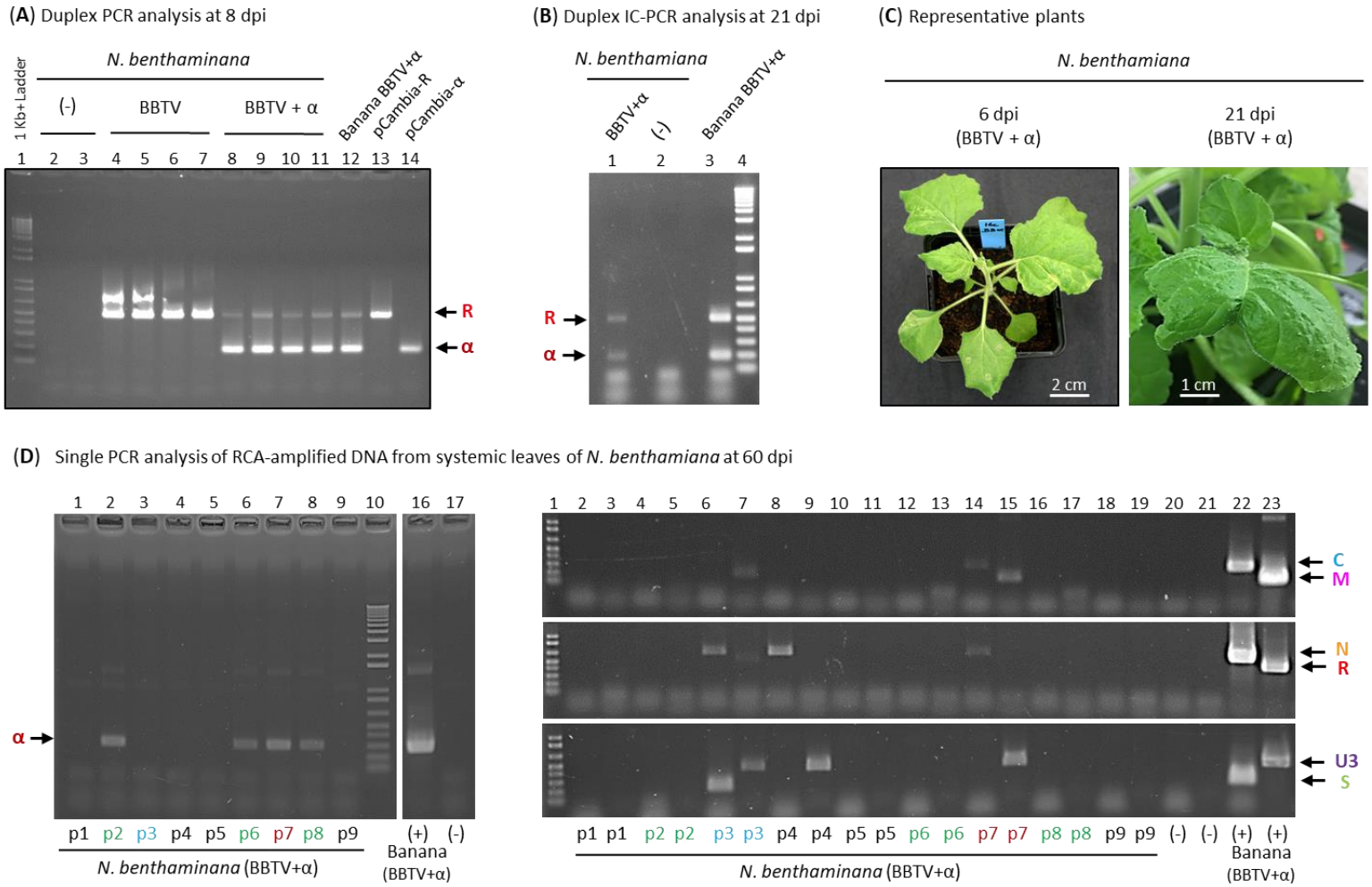


Fig 2. PCR and IC-PCR analysis of *Nicotiana benthamiana* plants agroinoculated with BBTV and alphasatellite clones.

(**A**) Duplex PCR analysis with BBTV DNA-R and DRC alphasatellite (α) specific primers of total DNA from symptomatic leaves of *N. benthamiana* agroinoculated with six BBTV components (lanes 3-6), six BBTV components and alphasatellite (lanes 7-10) or empty vector (lanes 1-2) at 8 days post-inoculation (dpi). Total DNA from Cavendish banana co-infected with BBTV and DRC alphasatellite and plasmid DNA of BBTV DNA-R and alphasatellite constructs (pCambia-R and pCambia-α) were used as positive controls. Positions of DNA-R and alphasatellite PCR products are indicated by arrows. (**B**) Duplex IC-PCR analysis of systemic leaf tissues of *N. benthamiana* agroinoculated with six BBTV components and alphasatellite at 21 dpi (lane 1), mock inoculated *N. benthamiana* (lane 2) and Cavendish banana infected with BBTV and alphasatellite following aphid inoculation (lane 3). (**C**) Representative *N. benthamiana* plants agroinoculated with 6 BBTV components and alphasatellite at 6 and 21 dpi. Note that wild type and 16c lines of *N. benthamiana* exhibited similar and reproducible symptoms in two independent experiments. (**D**) Single PCR analysis with BBTV component (C, M, N, R, S, U3)- and alphasatellite-specific primers of RCA-amplified total DNA from systemic leaves of nine *N. benthamiana* plants (p1, p2, p3, p4, p5, p6, p7, p6, p9) agroinoculated with six BBTV components and alphasatellite at 60 dpi. Positions of PCR products for each BBTV component and alphasatellite are indicated by arrows. Note that the original gel image (available on request) was spliced to remove five lanes between lanes 10 and 16.

Transient expression of BBTV Rep, but not alphasatellite Rep, induces chlorosis in *N. benthamiana*

To test whether any of the viral proteins can induce and/or suppress antiviral responses in *N. benthamiana*, we subcloned all the viral ORFs from the DRC-2016 alphasatellite and BBTV genome sequences under the control of CaMV 35S promoter and terminator for transient expression and silencing suppression assays in *N. benthamiana* 16c line transgenic for green fluorescence protein (GFP) (**S10A-B Fig**). The resulting constructs were co-expressed together with the GFP silencing inducer construct in *N. benthamiana* 16c leaves via agroinfiltration. Potyviral protein HC-Pro and bacterial β-glucuronidase (GUS) expression constructs were used as respectively positive and negative controls for silencing suppression. At 8 dpi, only HC-Pro strongly suppressed silencing, resulting in prominent GFP fluorescence under UV light, whereas none of the six BBTV protein or alphasatellite Rep expression constructs exhibited any strong silencing suppression activity (**S10C Fig**). Inspection of infiltrated leaves under day light revealed that the BBTV Rep construct, but not any other construct, induced strong chlorosis of the infiltrated tissues (**S10D Fig**). We hypothesize that, similar to other viruses, BBTV infection may induce both RNA silencing- and innate immunity-based defense responses (reviewed by Zvereva and Pooggin 2012) resulting in plant recovery from viral infection, while DRC alphasatellite persistence following recovery can be explained by its ability to evade defence responses in *N. benthamiana*. The mechanism of defense evasion remains to be investigated.

DRC alphasatellite interferes with BBTV transmission by banana aphids

To determine if DRC-2016 alphasatellite has any impact on BBTV transmission by aphids and/or development of BBTD symptoms, Cavendish banana plants infected with BBTV alone or co-infected with alphasatellite were taken for two consecutive transmission experiments designated C (4 BBTV vs 3 BBTV+alphasatellite plants) and D (4 BBTV vs 4 BBTV+alphasatellite plants). In each experiment, virus-free DRC aphids were placed on these source plants for three weeks to build colonies. Adult aphids were then taken from each source plant to inoculate recipient plants (5 aphids per recipient plant). After 4 days of inoculation, all aphids were collected from each recipient plant in pools for PCR diagnostics of DNA-R and alphasatellite and quantitative (q)PCR analysis of viral loads. All the aphid pools analysed by PCR contained the virus and the virus with alphasatellite when it was present in the source plant

(**S2 Dataset**). Symptom development was monitored during 80 days to record the dates of appearance of dark-green streaks (Morse code) on leaf midrib and veins (designated as first symptoms) and emergence of next leaf with systemic symptoms (a smaller leaf with shorter petiole, yellow margins and Morse code).

After 80 days, upper leaf samples from symptomatic recipient plants were taken for PCR diagnostics of DNA-R and alphasatellite and qPCR analysis of viral loads. In the case of source plants infected with BBTV alone, all the examined symptomatic recipient plants (n = 55) were found to be PCR-positive for BBTV. In the case of BBTV- and alphasatellite-coinfected source plants, all the examined symptomatic recipient plants (n = 28) were found to be PCR-positive for BBTV and most of them (n = 20) were also PCR positive for alphasatellite. Since all the 28 aphid pools contained alphasatellite DNA together with six BBTV components (**S2 Dataset**), alphasatellite could be lost during aphid inoculation or development of systemic infection in some of the recipient plants.

The results of two transmission experiments showed a negative impact of DRC alphasatellite on the BBTV transmission rate (percentage of virus-infected plants among inoculated recipient plants) which was reduced from 63 % to 37 % (**Fig 3A**). Using parametric Student t and linear model tests, the impact of the alphasatellite was found to be statistically significant (P = 0.041 and P = 0.048, respectively). However, a non-parametric Kruskal-Wallis test yielded a P-value (P = 0.072) exceeding an arbitrary threshold ($P \leq 0.05$) of statistical significance. This is due to variation in BBTV transmission rates between the individual, alphasatellite-containing source plants, two of which giving bigger proportions of virus-infected per inoculated recipient plants (10/12 and 8/12) than the other five (2/12, 0/11, 3/12, 3/8, 3/12). In contrast, in the absence of alphasatellite the variation between the individual source plants was smaller (5/12, 7/12, 8/11, 8/12, 7/11, 7/12, 6/12 and 12/12) (**S2 Dataset**; **Fig 3A**). Thus, alphasatellite strongly interfered with BBTV transmission from 5 out of 7 source plants. When comparing individual recipient plants, the success of viral infection did not depend on viral DNA loads in the aphid pools after 4-day inoculation (**S2 Dataset**). Nonetheless, the median load of helper virus DNA in the aphids was found to be lower in the presence of alphasatellite than in its absence (Kruskal-Wallis P = 0.045) (**Fig 4B**). Likewise, alphasatellite reduced median loads of helper virus DNA in both source and recipient plants, although only in the recipient plants the reduction was found to be statistically significant (Kruskal-Wallis P = 0.007 vs P = 0.086) (**Fig 4A** and **Fig 4C**).

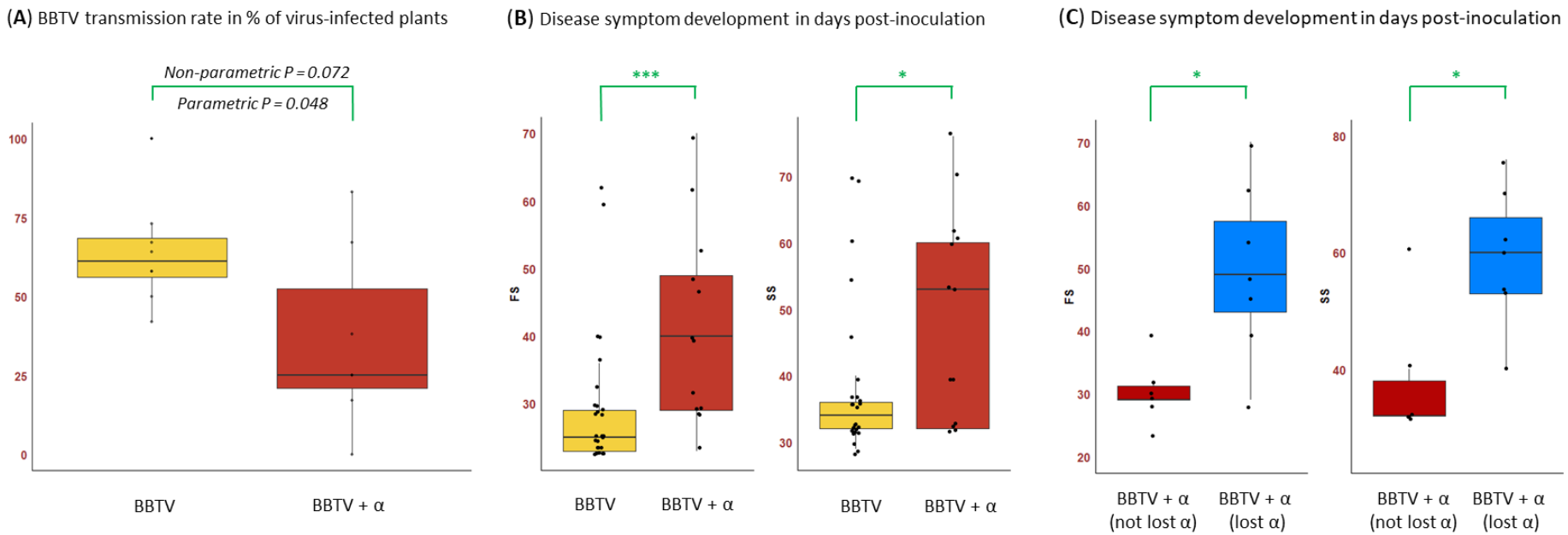


Fig 3. Impact of DRC alphasatellite on BBTV transmission by banana aphids and disease symptom development in recipient plants.

(A) Transmission rates in percentage of virus-infected recipient plants per source plant infected with BBTV alone (n = 8) and co-infected with BBTV and alphasatellite (n = 7) are shown for two transmission experiments (C and D). Statistical significance of differences in the transmission rates with and without alphasatellite was calculated using in a parametric linear model test and a non-parametric Kruskal-Wallis and the resulting P-values are indicated. **(B)** Delay in development of first (FS) and systemic (SS) symptoms in recipient plants in days post-inoculation with BBTV alone (n = 26) and BBTV+alphasatellite (n = 13) in the transmission experiment D. **(C)** Delay in development of first (FS) and systemic (SS) symptoms in the recipient plants of experiment D in days post-inoculation with BBTV+alphasatellite plotted separately for the plants that contained alphasatellite (n = 6) and those that lost alphasatellite (n = 7). * Kruskal-Wallis $P < 0.05$. *** Kruskal-Wallis $P < 0.005$.

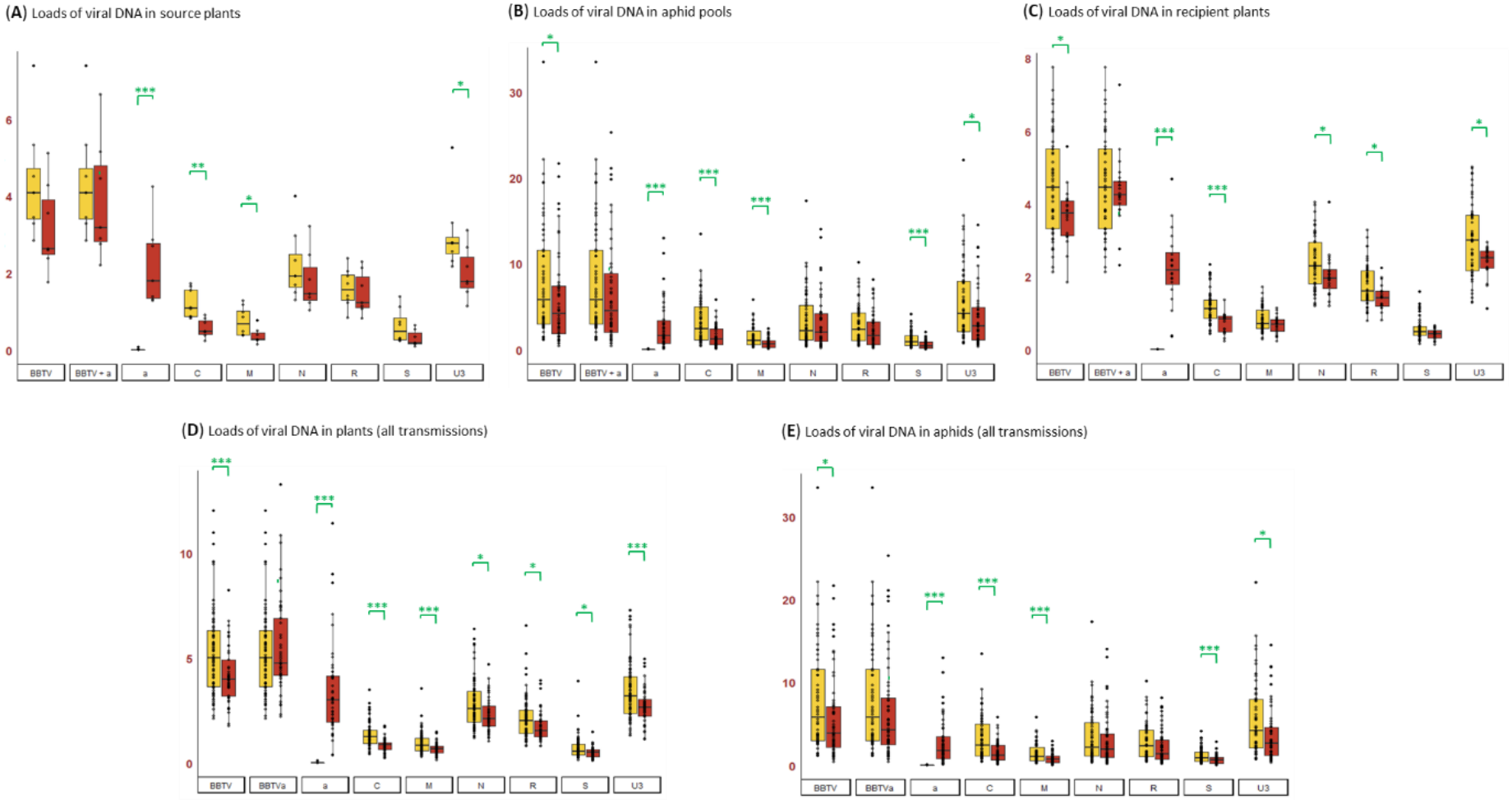


Fig 4. Impact of DRC alphasatellite on viral DNA loads in source plants, aphid pools and recipient plants.

Viral DNA loads for each BBTV component (C, M, N, R, S, U3) and alphasatellite (a) as well as for total helper virus without (BBTV) or with (BBTVa) alphasatellite measured by quantitative PCR in (**A**) source plants without (n = 8) and with (n = 7) alphasatellite (transmission experiments C and D), (**B**) aphid pools without (n = 58) and with (n = 46) alphasatellite (experiments C and D), (**C**) recipient plants without (n = 52) and with (n = 20) alphasatellite (experiments C and D), (**D**) all plants without (n = 77) and with (n = 47) alphasatellite (all experiments) and (**E**) all aphid samples without (n = 58) and with (n = 56) alphasatellite (all experiments). * Kruskal-Wallis $P < 0.05$. *** Kruskal-Wallis $P < 0.005$.

It should be noted that all the examined pools of viruliferous aphids (independent of the transmission success) contained all the six DNA components of BBTV genome. Likewise, all the examined symptomatic recipient plants contained the complete BBTV genome, with an exception for two plants where DNA-N was below detection (**S2 Dataset**). Since these two plants developed disease symptoms, the nuclear shuttle protein (NSP) encoded by DNA-N is not required for systemic infection and symptom development. When the plant lacking DNA-N was taken in a follow-up aphid transmission experiment, none of 12 recipient plants developed any disease symptoms or contained any detectable amounts of viral DNA. This suggests that BBTV NSP may serve as an aphid transmission factor, similar to nanoviral NSP (Grigoras et al. 2018; Di Mattia et al. 2020). In further support of this hypothesis, a banana plant naturally infected with a mild strain of BBTV in Taiwan lacked DNA-N and banana aphids were not able to transmit this strain to new plants (Fu et al. 2009).

The effect of DRC alphasatellite on disease symptom development could be properly evaluated only in the experiment D (because the experiment C was disturbed by thrips infestation) and found to be negative: the first symptoms (FS) were delayed for ca. 15 days (Kruskal-Wallis P = 0.005), while the systemic symptoms (SS) for ca. 20 days (Kruskal-Wallis P = 0.032) (**Fig 3B**; **S2 Dataset**). Remarkably, both first and systemic symptoms were delayed much more pronouncedly in the recipient plants that had lost alphasatellite upon inoculation (n = 7) than in those containing the alphasatellite (n = 6) (Kruskal-Wallis P = 0.014 and 0.013, respectively) (**Fig 3C**). Note that in the experiment C, one of the 13 examined recipient plants lacked alphasatellite, but, in that plant, DNA-N was also below detection (**S2 Dataset**). We speculate that the loss of alphasatellite could be due to its competition with helper virus components for viral coat protein within a single co-infected cell and, additionally, a higher probability for six BBTV components to create the disease complex than for the six components to be associated with alphasatellite in single or adjacent cells. It is puzzling, however, why the loss of alphasatellite correlated with the pronounced delay in BBTV symptom development.

DRC alphasatellite interferes with BBTV replication in plants and reduces BBTV loads both in plants and aphids

To evaluate the impact of DRC alphasatellite on BBTV replication in banana plants and on BBTV acquisition and persistence in aphids we performed qPCR analysis of viral DNA in BBTV-infected Cavendish plants (n = 122) and viruliferous DRC aphids (n = 114) sampled

during all our transmission experiments from 2017 to 2021. The loads of each BBTV component (C, M, N, R, S, U3) and alphasatellite (alpha) as well as total viral DNA (C+M+N+R+S+U3+alpha) were measured using a host housekeeping gene (banana RPS2 and aphid EF1α) as an internal control for normalization (**Fig 4D-E**; **S11 Fig**; note that in these Figures "BBTV" indicates the helper virus DNA, while "BBTVa" - the total viral DNA).

In banana plants, median loads of all BBTV components were reduced in the presence of alphasatellite, thereby resulting in a ~25 % decrease of the median load of total helper virus DNA (Kruskal-Wallis P = 0.004). The median loads of total viral DNA (BBTV+alpha) were found to be comparable in the presence and absence of alphasatellite (Kruskal-Wallis P = 0.70) (**Fig 4D**; **S11A Fig**). This suggests that self-replicating alphasatellite competes with the helper virus for the host replication machinery, thereby reducing accumulation of helper virus DNA. Notably, the median load of alphasatellite is the highest among viral components, followed by DNAs U3, N and R, which also accumulate at relatively high levels, compared to DNAs C, M and S. In the absence of alphasatellite, DNA-U3 is the most abundant component followed by DNAs N, R, C, M and S (in that order) (**Fig 4D**). DNA-S encoding the coat protein is the least abundant component at both conditions, which might become a limiting factor for virus encapsidation in the presence of alphasatellite that reduces its accumulation.

In aphids, median loads of all BBTV components were also reduced in the presence of alphasatellite, thereby resulting in a ~40 % decrease in the median load of total helper virus DNA (Kruskal-Wallis P = 0.025). The difference in median loads of total viral DNA in the presence and absence of alphasatellite was not statistically significant (Kruskal-Wallis P = 0.12) (**Fig 4E**; **S11B Fig**). While alphasatellite is the most abundant viral component in virus-infected plants, in aphids it becomes the third most abundant after DNAs U3 and N and accumulates at levels comparable to those of DNA-R (**Fig 4**). Since the plant phloem sap ingested by aphids contains encapsidated viral DNA, a large proportion of alphasatellite may not be encapsidated *in planta* due to a limiting amount of the viral coat protein. In support of this hypothesis, our semi-quantitative duplex IC-PCR vs duplex PCR analyses revealed that high-abundance alphasatellite DNA is encapsidated less efficiently than low-abundance DNA-R in both banana and *N. benthamiana* plants (**Fig 2B** versus **Fig 2A** and **S5 Fig**).

To further evaluate the impact of DRC alphasatellite on helper virus replication, we analysed its effect on relative abundance (formula) of BBTV genome components in banana plants. Being the most abundant virome component (on average 35 % of total viral DNA; **Fig 5E**), alphasatellite leads to a substantial and statistically-significant decrease in relative abundance

of DNA-C (Kruskal-Wallis P = 1.97E-05) but not other components, although median values are increased for DNAs N and R (**Fig 5A**) which together compensate for the more pronounced decrease in relative abundance of DNA-C. Therefore, alphasatellite might interfere more substantially with master Rep-mediated trans-replication of DNA-C than other components, which in turn might lead to reduced efficiency of viral replication. Indeed, Clink encoded by DNA-C is a replication enhancer interacting with retinoblastoma-related protein and modulating cell cycle in favor of viral replication (Aronson et al. 2000; Lageix et al. 2007).

In aphids, alphasatellite becomes relatively less abundant (15% of total viral DNA; **Fig 5E**) and has a more pronounced impact on the helper virus formula, leading to a bigger decrease in relative abundance of DNA-C (Kruskal-Wallis P = 2.20E-16) and a more substantial (and statistically-significant) increase in relative abundances of DNAs N and R (Kruskal-Wallis P = 2.20E-16 and P = 0.0002, respectively) (**Fig 5B**). Thus, alphasatellite-mediated alterations in relative abundances of BBTV components *in planta* become more pronounced in aphids.

Further comparison of BBTV genome formulas between plants and aphids revealed that, in the absence of alphasatellite, DNA-N is relatively more abundant in plants, while DNA-C is relatively more abundant in aphids (**Fig 5C**). In the presence of alphasatellite, DNA-N is still relatively more abundant in plants, while DNA-C is still relatively more abundant in aphids, but these differences are less pronounced than in alphasatellite´s absence (**Fig 5D**).

In the presence of alphasatellite, the overall formula of viral DNA components differs between plants and aphids: in plants, relative abundance of alphasatellite DNA is comparable to that of DNA-U3 which is the most abundant component of helper virus, while alphasatellite becomes relatively less abundant than DNA-U3 in aphids and accumulates at levels comparable of those of DNA-N (which is the second most abundant component of the helper virus after DNA-U3 in plants and the third most abundant in aphids) (**Fig 5E**). This indicates that both alphasatellite and DNA-N are less efficiently acquired by aphids compared to other components, possibly due to differences in relative encapsidation of viral DNA. No difference in the relative abundance of DNAs N and M was found between plants and aphids, while relative abundances of DNAs C, R, S and U3 are slightly (and statistically significantly) increased in aphids compared to plants (**Fig 5E**)

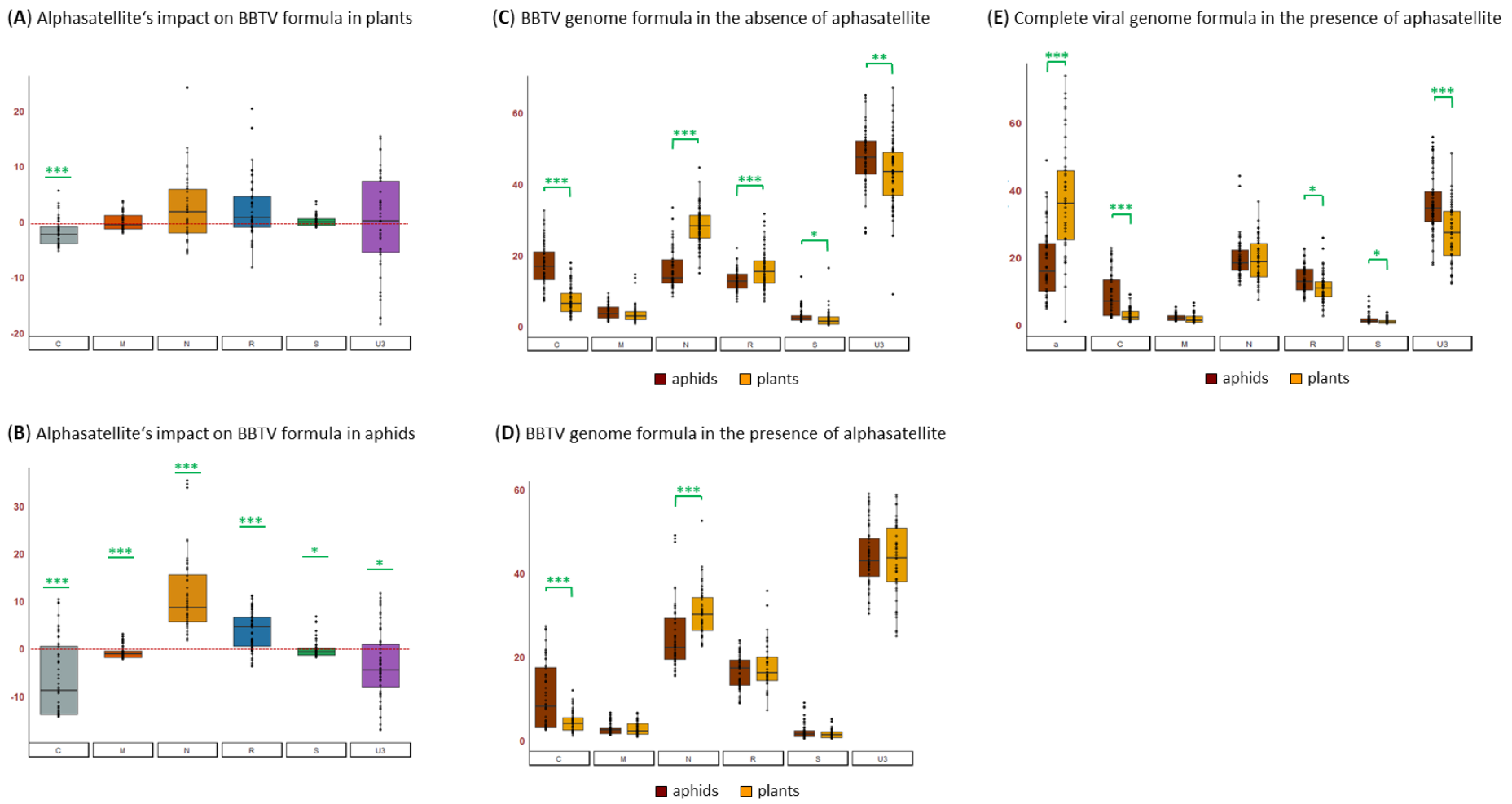


Fig 5. Impact of DRC alphasatellite on BBTV genome formulas in banana plants and aphids.

Relative abundance (formula) of BBTV genome components in the presence and absence of alphasatellite was calculated in infected plants and aphids from all transmission experiments (see Fig 5D-E). (**A-B**) Alphasatellite-mediated alterations in BBTV genome formula in plants (**A**) and aphids (**B**) plotted for each component as difference in its percentage (above zero when increased and below zero when decreased). (**C-D**) Comparisons between BBTV genome formulas in plants and aphids in the absence (**C**) and presence (**D**) of alphasatellite plotted in percentage of total BBTV DNA. (**E**) Comparison between virome (BBTV + alphasatellite) formulas in the plants and aphids containing alphasatellite plotted in percentage of total viral DNA. * Kruskal-Wallis P < 0.05. *** Kruskal-Wallis P < 0.005.

Taken together our findings suggest that DRC alphasatellite being the most abundant viral component *in planta* interferes with encapsidation of helper virus components and their acquisition by aphids and/or with their persistence in aphids. In addition to apparent competition with helper virus for the plant replication machinery leading to reduced accumulation of all viral components, DRC alphasatellite has a more pronounced negative impact on accumulation of replication enhancer-encoding DNA-C, which might further reduce helper virus replication efficiency.

DRC alphasatellite affects BBTV transcription

To further understand molecular mechanisms underlying alphasatellite-mediated alterations in helper virus accumulation, genome formula and transmission rate, we examined the impact of DRC alphasatellite on viral gene expression and antiviral RNAi responses by Illumina sequencing analysis of long and small RNAs from Cavendish plants infected with BBTV alone (n = 3) and co-infected with alphasatellite (n = 3) (**S3 Dataset** and **S4 Dataset**). The plants were selected based on qPCR analysis of viral DNA loads and formulas to be representative of the analysed infected plants with respect of both between-sample variation and differences in median values in the presence and absence of alphasatellite. In the selected plants, alphasatellite constituted on average ca. 40% of total viral DNA and the average load of total helper virus DNA was ca. 50% higher in the absence of alphasatellite than in its presence: the negative impact of alphasatellite on helper virus accumulation was distributed between all six DNA components (**S12 Fig**). Such selection enabled us to pool together RNA-seq reads for three plants per each condition (+/– alphasatellite) in order to account for between-plant variation and to quantify an average impact of alphasatellite on helper virus transcription and viral siRNA production.

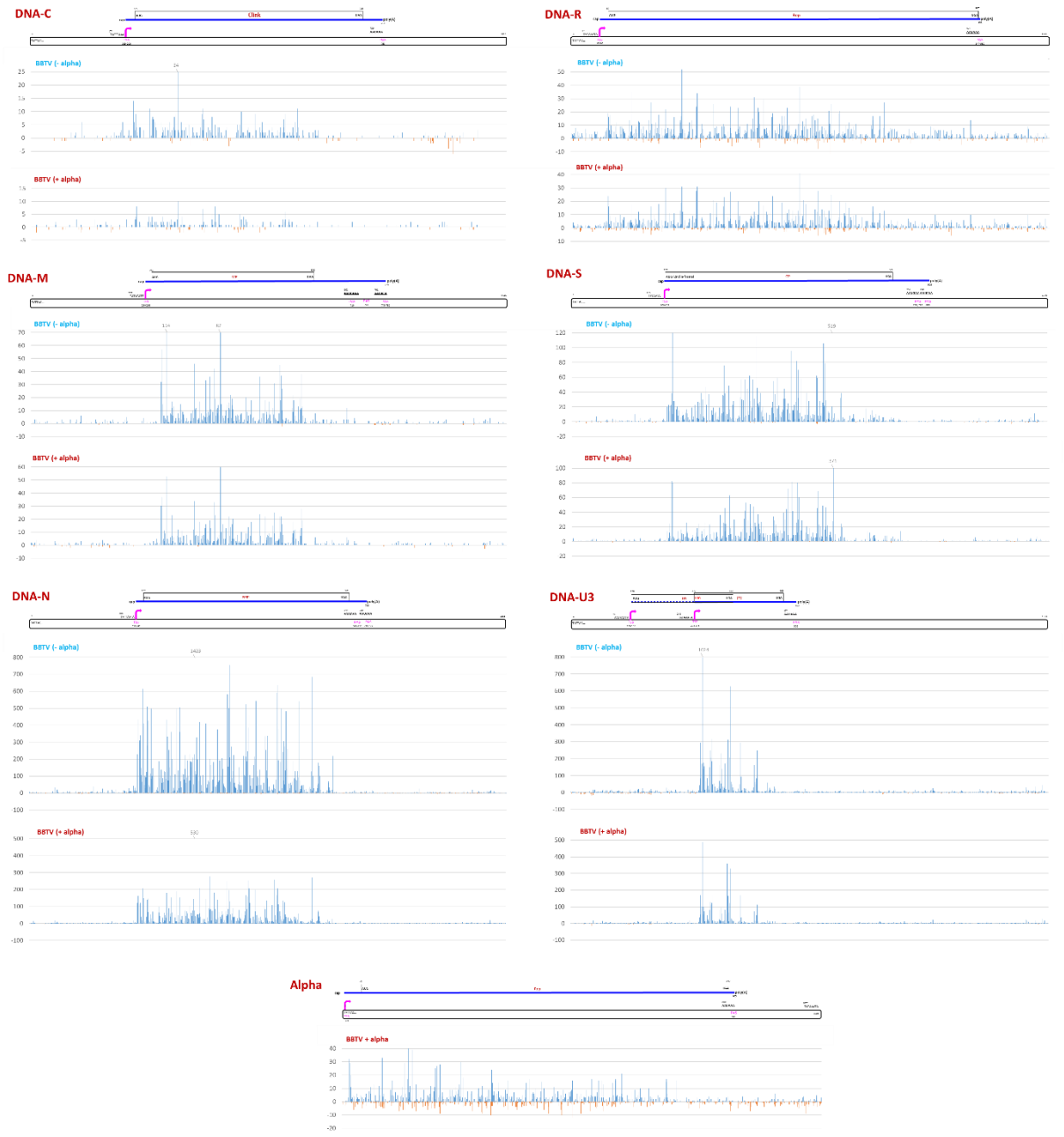


Fig 6. Single nucleotide resolution maps of Illumina mRNA-seq reads representing poly(A)-enriched viral transcripts from BBTV-infected Cavendish banana plants with or without DRC alphasatellite.

For each of the two conditions, i.e. without (BBTV-alpha) and with (BBTV+alpha) alphasatellite, Illumina 75 nt paired-end reads of the three biological replicates (leaf tissues of three plants) were combined and mapped simultaneously onto the reference sequences of six BBTV components (-/+ alphasatellite). Histograms plot the numbers of viral 75 nt reads at each nucleotide position of the 1018-to-1111 nt BBTV genome components (DNAs C, M, N, R, S, U3) and 1105 nt alphasatellite (Alpha): blue bars above the axis represent sense reads starting at each respective position, while red bars below represent antisense reads ending at the respective position. The genome organizations of BBTV components and alphasatellite are shown schematically above the respective histograms, with the Pol II promoter (TATA-box and transcription start site, TSS) and terminator (polyA signal, PAS) elements indicated in pink, capped and polyadenylated mRNA shown as solid blue lines, viral protein-coding ORFs boxed and their nucleotide positions given.

Mapping of Illumina 75 nt paired-end mRNA-seq reads onto the viral reference sequences revealed monodirectional Pol II transcription units of viral mRNAs for each BBTV component and alphasatellite (**Fig 6**; **S13 Fig**; **S5 Dataset**). With exceptions for DNAs M and U3, these units corresponded to those predicted *in silico* based on positions of (i) TATA or TATA-like boxes at a sufficient distance (>31 nts) upstream of the AUG start codons of viral ORFs and (ii) poly(A) signals (AAUAAA or its variants with one nucleotide substitution) driving pre-mRNA cleavage and polyadenylation downstream of the stop codons. Abundance of the reads representing viral mRNAs was much lower at the predicted 5'-termini and 3'-regions, possibly due to biases in the Illumina mRNA-seq protocol that included poly(A) enrichment, RNA fragmentation, random primer cDNA synthesis and adapter ligation steps. In the case of DNA-S, the second in-frame AUG (A at nt 213) is likely used as an authentic translation initiation codon, as the first AUG (nt 198) is too close to the conserved TATA box (first T at nt 175) and too far from the region of high-abundance reads (starting from nt 216; **S13 Fig**). This finding is consistent with transcription start mapping by 5'-RACE at nt 208 of DNA-S of an Australian BBTV isolate (Herrera-Valencia et al. 2007) and N-terminal sequencing of its coat protein (Wanitchakorn et al. 1997). In the case of DNA-M, high-abundance reads begin at nt 287, downstream of the AUG start codon located at nt 267 and the 5'-end of mRNA-M previously mapped by 5'-RACE at nt 254 (Wanitchakorn et al. 1997). Thus, the conserved TATA-box at nt 247 appears to drive transcription of the shortened mRNA-M lacking the first AUG, while a TATA-like box at nt 223 may drive transcription of its longer version mapped previously. In the case of DNA-U3, a 5'-half of the previously mapped mRNA, whose transcription could be driven by a TATA-like box at nt 97 (Wanitchakorn et al. 1997), was not represented with high-abundance reads (**Fig 6**; **S13 Fig**). Sequence inspection upstream of the high-abundance read region (from nt 283) revealed a conserved TATA box at nt 243. However, the only downstream ORF would begin with a non-AUG start codon UUG (nt 272) located too close to the TATA box and in a poor initiation context (UAA UUG CUC) lacking purines at -3 and +4 positions. Analysis of Illumina sRNA-seq reads revealed a highly-abundant 21 nt sRNA belonging to the conserved miR156 family (Zhu et al. 2019) that maps with 2 mismatches to the DNA-U3 positions 272-292 in reverse orientation. This banana miRNA could potentially direct cleavage of mRNA-U3 between nts 281 and 282. For other BBTV mRNAs, high-abundance reads begin upstream of the predicted AUG start codons and at a short distance downstream of the 5´-ends mapped previously (Herrera-Valencia et al. 2007): for DNA-C, at nt 220 (vs nt 209); for DNA-N, at nt 246 (vs nt 239); and for DNA-R at nt 83 (vs nt 69 and nt 72) (**Fig 6**; **S13 Fig**).

The poly(A) sites that we mapped for our African isolate (**Fig 6**; **S13 Fig**) correspond for most but not all viral mRNAs to those previously mapped for the Australian isolate (Beetham et al. 1999). In DNA-C, a single poly(A) site is located at the same position (nt 753) in both isolates. In DNA-M, poly(A) sites are at nt 710, 738 and 779, with the latter being 2 nts upstream of the previously mapped site. In DNA-N, a major poly(A) site is at nt 749, which corresponds to the previously mapped site, while minor sites are located just upstream (nt 746) and downstream (nt 768 and 772). In DNA-R, poly(A) sites are at nt 947 and 950, with the former corresponding to the previously mapped site. In DNA-S, poly(A) sites are at nt 778, 805 and 808, one of which (nt 778) corresponding to the previously mapped site. In DNA-U3, a single poly(A) site is at nt 502, which is too far upstream of the previously mapped site, making the 3´-UTR much shorter (35 vs 110 nts). It should be noted that, unlike other components, DNA-U3 sequences of the African and Australian isolates differ substantially and the predicted ORFs encode 77 vs 88 amino acid proteins, respectively.

In the case of the alphasatellite transcription unit, high-abundance mRNA-seq reads begin at nt 17, that is 45 nts downstream of a conserved TATA box (nt 1076) and upstream of the AUG start codon of Rep-encoding ORF (nt 41). The alphasatellite poly(A) site is at nt 903, that is at an optimal distance downstream of a conserved poly(A) signal (nt 880) and downstream of the Rep ORF stop codon (nt 892) (**Fig 6**; **S13 Fig**).

The presence of alphasatellite did not result in any substantial alteration in positions and relative abundances of the reads representing viral mRNAs (**Fig 6**; **S13 Fig**) but had a negative impact on their average accumulation levels (calculated as mRNA unit forward reads normalized per million of total plant + viral reads), resulting in a 2.4-fold decrease in accumulation of total helper virus mRNA. This is more pronounced than a 2.1-fold decrease in accumulation of total helper virus DNA (**Fig 7A** vs **Fig 7D**). For each BBTV component, the average mRNA load was reduced, albeit at a different degree. For DNA-N, the reduction is most pronounced (about 3-fold) and notable, because mRNA-N is the most abundant in the absence of alphasatellite: it does remain the most abundant in the presence of alphasatellite, but its relative abundance becomes substantially lower in the BBTV mRNA formula (55 % vs 67 %), while mRNAs of other components become either relatively more abundant (M, R, S) or remain almost unchanged (C, U3) (**Fig 7A**). Remarkably, neither loads nor formulas of viral mRNAs correlated with those of viral DNAs (**Fig 7A** vs **Fig 7D**).



Fig 7. Impact of alphasatellite on accumulation and relative abundance of viral mRNAs, viral antisense transcripts and viral siRNAs.

For each of the two conditions, i.e. without (BBTV) and with (BBTVa) alphasatellite, Illumina mRNA-seq 75 nt reads or sRNA-seq 19-25 nt reads of each of the three biological replicates (leaf tissues of three plants) were mapped with zero mismatches onto the reference sequences of all viral mRNA units and all complete viral genome components. Mapped mRNA unit sense reads, complete genome antisense reads, or complete genome sRNA sense+antisense reads were counted in reads per million (RPM) of total reads and their relative abundance (mRNA, antisense transcript, or sRNA formula) was calculated for helper virus (BBTV) and virome (BBTV + alphasatellite) in % of total. (A-C) Loads and formulas of viral mRNAs (A), viral antisense transcripts (B) and viral sRNAs (C). (D) Loads and formulas of viral DNA calculated by qPCR for each of the three plants (biological replicates) per condition (+/- alphasatellite). Error bars represent standard deviations.

This suggests differences in their transcription rates likely due to unequal strength of their Pol II promoters. Normalization of mRNA counts (in reads per million) by viral DNA loads and by transcription unit sizes to estimate mRNA transcription rates revealed that the least abundant DNA-S (1.1 % of total viral DNA) is the most transcriptionally efficient component (61.1 % of total), followed by DNAs N (20.0 %), M (10.7 %), U3 (5.7 %), R (1.7 %) and C (0.8 %) (**Fig 8A**). In the presence of alphasatellite, the transcription rate was slightly increased for DNAs S (66.9 %), M (13.0 %), R (2.3 %) and C (1.2 %) and reduced for DNAs U3 (4.3 %) and N (12.2 %), with the latter being the most strongly affected, accounting for the most pronounced negative impact of alphasatellite on the load of mRNA-N. Notably, alphasatellite DNA being the most abundant (42.5 % of total viral DNA) exhibits the lowest transcription rate (0.8 % of total) (**Fig 8**). Previous analysis of viral DNA and RNA loads in two Cavendish plants sampled in China and found to be co-infected with BBTV and one or two alphasatellites, respectively, revealed substantial differences between viral DNA and RNA formulas (Yu et al. 2019). The differences are generally comparable to those observed in our study: alphasatellite DNA was the most abundant, followed by DNAs U3 and N, while alphasatellite RNA was one of the least abundant. Furthermore, mRNA-N was the most abundant in both plants, followed by mRNAs S and U3 and the DNA-N promoter had the highest activity in driving GFP transgene expression in tobacco protoplasts, followed by the promoters of DNAs S and M (Yu et al. 2019). In our case, however, the transcription rate of DNA-S was the highest, followed by DNAs N and M (**Fig 8A**).

Taking together, DRC alphasatellite competes with helper virus for both replication and transcription machinery of the host plant, thereby reducing accumulation of helper virus DNA and mRNA, and has the most pronounced negative impact on DNA-N transcription by Pol II. Together these effects of alphasatellite may synergistically reduce transmission efficiency of the helper virus, since mRNA-N encodes a putative aphid transmission factor (as described above).

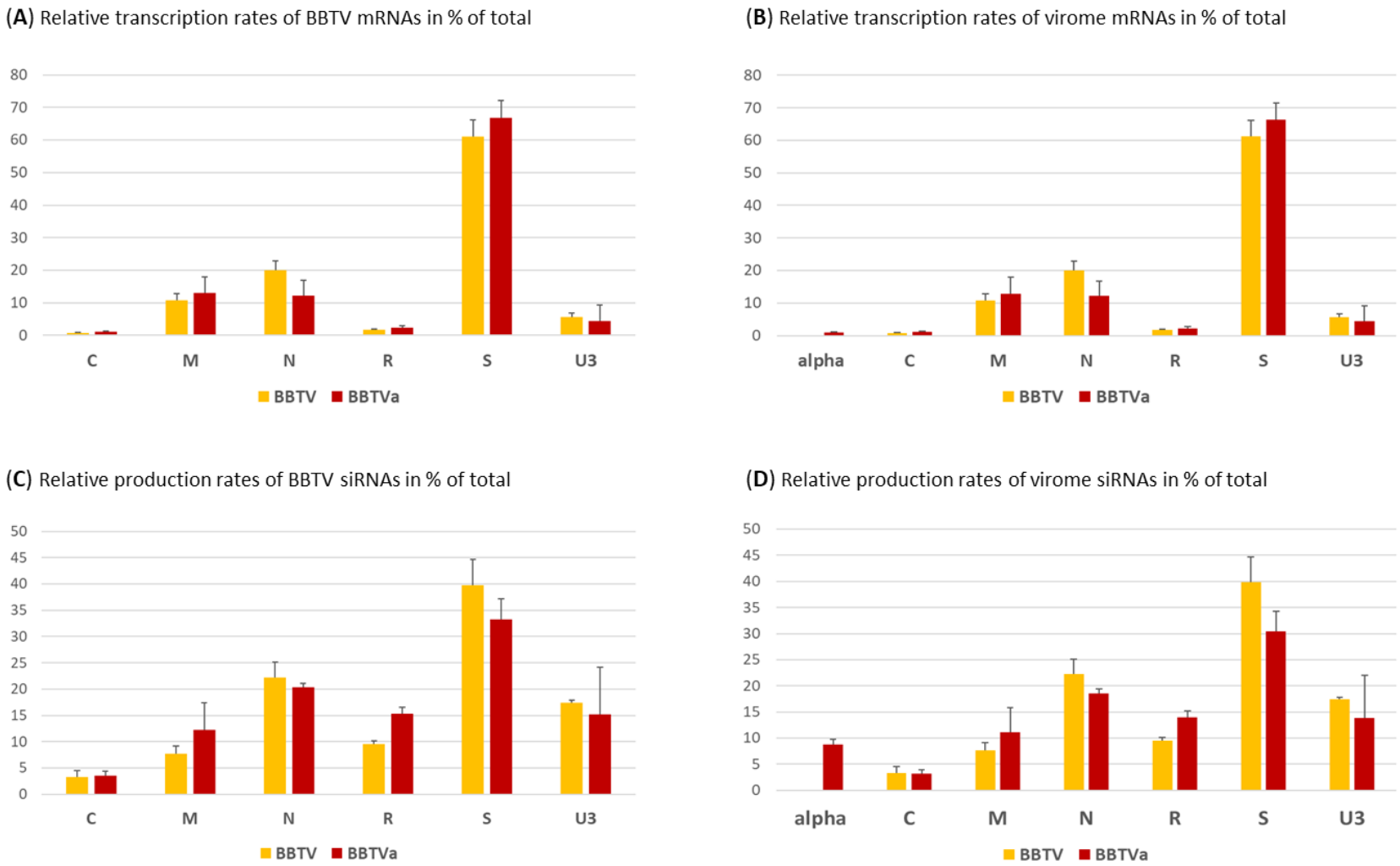


Fig 8. Impact of alphasatellite on relative transcription rates of viral mRNAs and relative production rates of viral siRNAs.

Transcription rates of viral mRNAs were calculated by dividing mRNA counts in reads per million by viral DNA loads (measured by qPCR) and by mRNA unit sizes in nucleotides (alphasatellite = 902 nts; C = 549 nts; M = 526 nts; N = 511 nts; R = 883 nts; S = 603 nts; U3 = 231 nts) and the relative transcription rates of viral genome components in the presence (red) and absence (yellow) of alphasatellite (alpha) was plotted in percentage (%) of total BBTV (**A**) and virome (**B**) transcription rates. Production rates of viral 19-25 nt sRNAs were calculated by dividing sRNA counts in reads per million (RPM) by viral DNA loads (measured by qPCR) and the relative sRNA production rates from viral genome components in the presence (red) and absence (yellow) of alphasatellite (alpha) were plotted in % of total BBTV (**C**) and virome (**D**) production rates. Error bars represent standard deviations for three biological replicates.

Besides higher-abundance reads representing the monodirectional mRNA units, lower-abundance reads were found to cover complete (or near-complete) forward and reverse strands of each BBTV component and alphasatellite (**Fig 6**; **S14 Fig**; **S5 Dataset**). This is indicative of pervasive transcription of the entire viral DNA in both sense and antisense orientations by Pol II or another RNA polymerase(s). In contrast to mRNA, viral antisense transcript loads and formulas in the presence or absence of alphasatellite generally well reflect viral DNA loads and formulas with exception for DNAs R and S showing higher antisense transcription levels on the expense of DNAs C, M and N (**Fig 7B** vs **Fig 7D**). These effects of alphasatellite on antisense transcription may in turn modulate expression of respective viral genes at transcriptional and/or posttranscriptional levels. Indeed, antisense and other aberrant viral transcripts would induce the antiviral RNAi machinery generating viral siRNAs that potentially direct both transcriptional silencing of viral DNA and posttranscriptional silencing of viral mRNAs.

DRC alphasatellite affects viral siRNA production

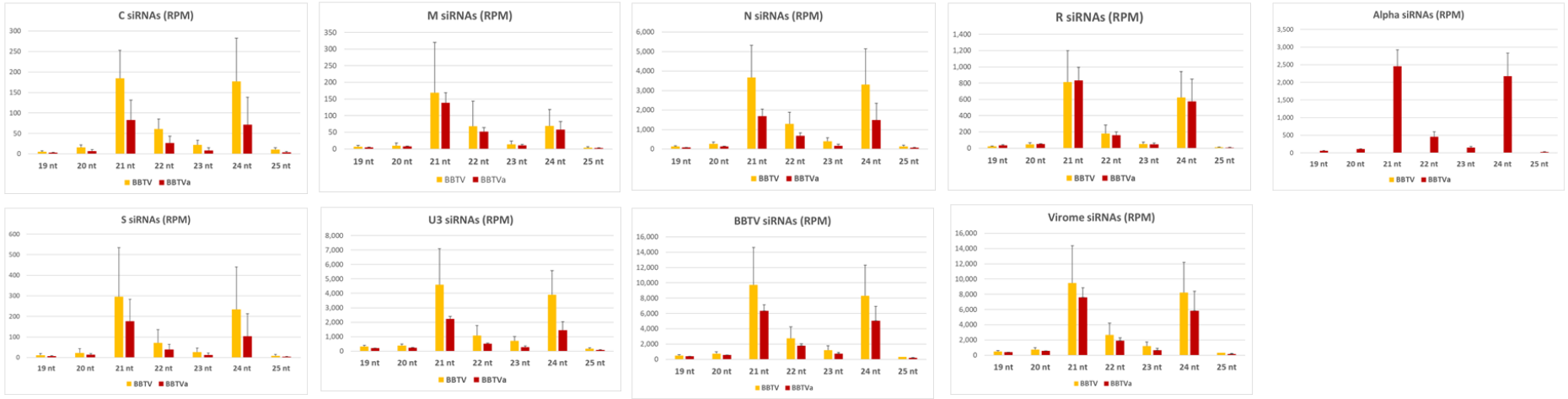
Mapping Illumina sRNA-seq reads (15-34 nt range) onto viral reference sequences with zero mismatches revealed that viral siRNAs predominantly belong to 21, 22 and 24 nt classes and cover both strands of the entire BBTV genome and alphasatellite, with their hotspots being unequally distributed between viral components (**Fig 9** and **Fig 10**; **S6 Dataset**). In the presence of alphasatellite, the average load of helper virus-derived siRNAs (normalized per million of total plant+viral sRNA reads) was reduced by 50 %, which is proportional to 2.1-fold reduction in its DNA load (**Fig 7C** vs **Fig 7D**). This overall reduction affected all the BBTV genome components except R whose siRNA loads are comparable in the presence and absence of alphasatellite despite a 1.7-fold reduction in the average load of DNA-R. Remarkably, relative abundance of helper virus components-derived siRNAs (siRNA formula) resembles its DNA formula in that DNAs U3 and N spawn the first and second most abundant siRNAs (47.4 and 39.1 % of total), followed by DNA-R (7.5 %) (**Fig 7C** vs **Fig 7D**). However, the least abundant DNA-S spawns siRNAs at a higher rate (calculated as an siRNA load divided by a DNA load) than other components (~2, 3, 4, 5 and 8 times higher than DNAs N, U3, R, M and C, respectively; **Fig 8C**) and as a result the relative abundance of DNA-S-derived siRNAs (2.6 %) is higher than that of siRNAs derived from higher-abundance DNAs M (1.3 %) and C (2.1 %) (**Fig 7C** vs **Fig 7D**).



Fig 9. Single nucleotide resolution maps of Illumina sRNA-seq reads representing viral siRNAs from BBTV-infected Cavendish banana plants with or without DRC alphasatellite.

For each of the two conditions, i.e. without (BBTV-alpha) and with (BBTV+alpha) alphasatellite, Illumina 20-25 nt reads of the three biological replicates (leaf tissues of three plants) were combined and mapped simultaneously onto the reference sequences of six BBTV components (-/+ alphasatellite). Histograms plot the numbers of viral 20-25 nt reads at each nucleotide position of the 1018-to-1111 nt BBTV genome components (DNAs C, M, N, R, S, U3) and 1105 nt alphasatellite (Alpha): blue bars above the axis represent sense reads starting at each respective position, while red bars below represent antisense reads ending at the respective position. The genome organizations of BBTV components and alphasatellite are shown schematically above the respective histograms, with the Pol II promoter (TATA-box and transcription start site, TSS) and terminator (polyA signal, PAS) elements indicated in pink, capped and polyadenylated mRNA shown as solid blue lines, viral protein-coding ORFs boxed and their nucleotide positions given.

(**A**) viral siRNA size-class loads



(**B**) viral siRNA size-class profiles



Fig 10. Impact of alphasatellite on accumulation and relative abundance of viral siRNA size-classes.

For each of the two conditions, i.e. without (BBTV) and with (BBTVa) alphasatellite, Illumina sRNA-seq 19-25 nt reads of each of the three biological replicates (leaf tissues of three plants) were mapped with zero mismatches onto the reference sequences of complete viral genome components (DNAs C, M, N, R, S, U3, alpha). For each component the loads of each viral sRNA size class were counted in reads per million (RPM) of total plant+viral reads and plotted in panel (**A**), and the relative abundance (formula) of viral sRNA size classes was then calculated in percentage of total and plotted in panel (**B**). Error bars represent standard deviations.

In the presence of alphasatellite, siRNA production rates were increased for DNAs R and M but not other components (**Fig 8C**), which resulted in a pronounced increase in relative abundance of siRNAs derived from both DNA-R (from 7.5 to 14.5 %) and DNA-M (1.3 to 2.3 %) and a concomitant but much less pronounced decrease in relative abundances of siRNAs derived from other components except DNA-S (showing a slight increase from 2.6 to 2.8 %) (**Fig 7C**). The increased siRNA production may in turn downregulate expression of the respective viral genes at both transcriptional and/or posttranscriptional levels and thereby interfere with viral replication (Rep gene) and movement (MP gene). Notably, the siRNA production rate for alphasatellite is somewhat lower than for any BBTV component except DNA-C (**Fig 7C** vs **Fig 7D**), but because its DNA is highly abundant, the load and relative abundance of alphasatellite siRNAs is comparable to and even slightly exceeding those of DNAs N and U3 (**Fig 7C** vs **Fig 7D**).

Alphasatellite-mediated increase in the production rates of siRNAs from DNAs M and R is proportional for all three major siRNA classes (21, 22 and 24 nt) whose average loads remain unaltered despite the reduction in loads of DNAs M and R. In contrast, average loads of 21, 22 and 24 nt siRNAs derived from other components are strongly and proportionally reduced in the presence of alphasatellite (**Fig 10A**). For each component, relative abundance of viral siRNA size classes (size profile) is comparable in the presence and absence of alphasatellite (**Fig 10B**; **S4D Dataset**). Thus, DRC alphasatellite does not appear to suppress or induce any specific plant DCL activity previously shown to generate 21 nt (DCL4), 22 nt (DCL2) or 24 nt (DCL3) viral siRNAs in ssDNA begomovirus-infected plants (Akbergenov et al. 2006; Blevins et al. 2006; Aregger at al. 2012; reviewed in Pooggin 2013, Pooggin 2018). Likewise, 5'-nucleotide identity profiles of the three major viral siRNA classes, dominated by 5'U (>50 % in the 21 nt and 22 nt classes) and 5'A (>60 % in the 24 nt class), were not significantly affected by alphasatellite (**S4D Dataset**), indicating unaltered activities of the AGO proteins recognizing respective 5'-nucleotides as previously established in other plants (Borges and Martienssen 2015). These findings are consistent with our above-described finding that transient expression of alphasatellite Rep did not interfere with transgene silencing in *N. benthamiana*.

Despite drastic differences in the hotspot profiles of sense vs antisense siRNAs (**Fig 9**), their overall accumulation levels are comparable for both alphasatellite and helper virus, although DNAs C and S spawn respectively 2- and 1.6-times more abundant antisense siRNAs, while DNAs R and U3 spawn respectively 1.7- and 1.3-times less abundant antisense siRNAs. Alphasatellite did not have any substantial impact on the hotspot profiles or sense/antisense

ratios of siRNAs derived from any component or complete BBTV genome **(Fig 9**; **S4 Dataset** and **S6 Dataset**). The siRNA hotspots are mostly concentrated within the mRNA transcription units on both sense and antisense strands but they also occur outside of them and on both strands, most notably in DNAs C and U3 (**Fig 9**). This suggests that dsRNA precursors of viral siRNAs are derived from the entire BBTV genome and alphasatellite and that the antisense transcripts may represent antisense strands of the dsRNA precursors that were not fully processed by DCLs into siRNAs. Consistent with this hypothesis, the loads and formulas of antisense transcripts roughly resemble those of siRNAs, with exception for DNAs R and N generating respectively higher and lower abundance antisense transcripts as could be expected from their relative DNA loads (**Fig 7**). Nonetheless, the most abundant virome components - alphasatellite and DNA-U3 (in the absence of alphasatellite) - produce both most abundant antisense RNAs and most abundant siRNAs.

Conclusions and open questions

In this work, we report a new alphasatellite, representative of a new genus in the family *Alphasatellitidae*, which got associated with BBTV in DRC, the first country in Sub-Saharan Africa where BBTD was recorded in 1958. Recent emergence of this alphasatellite is emphasized by the fact that it was found only in one of 401 banana samples collected in DRC in 2012 and that it was absent in samples from Gabon, Nigeria, Benin and Togo collected in the following years. Tight association of BBTV with this alphasatellite is manifested by its presence in viruliferous aphids collected in DRC in 2016, 4 years after the first sampling (at a distance of ca. 100 km), as well as its ability to be transmitted by aphids under laboratory conditions in the following 5 years. This is despite its negative impact on helper virus accumulation in plants and aphids and on virus transmission by aphids. Nonetheless, we found that this alphasatellite can be lost during or after aphid transmission in a variable number of recipient plants depending on source plants and/or experimental conditions. Remarkably, the loss of alphasatellite correlated with a delay in BBTV symptom development, a phenomenon that remains to be further investigated.

Our molecular analyses revealed that DRC alphasatellite competes with helper virus for components of the host DNA replication and transcription machinery, thereby reducing viral DNA and mRNA accumulation and altering relative abundances and transcription rates of viral genome components. Most notable are its pronounced negative effects on accumulation of

DNA-C encoding the replication enhancer protein and on transcription rate of DNA-N encoding a nuclear shuttle protein implicated in aphid transmission. The former effect would further contribute to reducing viral DNA replication efficiency, while the latter would further interfere with virus acquisition and/or transmission by aphids. Indeed, we found that aphids were not able to transmit the virus lacking DNA-N. Although alphasatellite DNA accumulates in plants at higher levels than any component of the helper virus, it appears to be less efficiently encapsidated by viral coat protein and, as a result, becomes only the third most abundant component in aphids. Thus, in the presence of alphasatellite, which reduces loads of all viral DNAs including DNA-S, the coat protein might also become a limiting factor for encapsidation of helper virus components and hence their acquisition by aphids. All these findings at the molecular level explain the observed negative impact of DRC alphasatellite on virus transmission by aphids, highlighting apparent costs for the helper virus to be in a complex with a self-replicating competitor for the host cell resources. What would then be the benefits (if any) for the helper virus to compensate for those costs, which could explain persistence of the DRC alphasatellite that we observed under the laboratory and field conditions? Since BBTV components engage the host nuclear and cytoplasmic machinery in virus multiplication and movement and thereby induce defense responses and disease symptoms, the alphasatellite-mediated reduction in loads of all viral components may prolong life of diseased plants under field conditions and in turn increase chances for both helper virus and alphasatellite to be transmitted to new plants. Another potential benefit would be that highly abundant alphasatellite DNA could serve as a decoy diverting antiviral RNAi machinery from the helper virus. Indeed, we found that DRC alphasatellite spawns 21, 22 and 24 nt siRNAs and antisense transcripts whose levels exceed those spawned by any component of the helper virus and that the transcription rate of alphasatellite mRNA is one of the lowest, indicating potent RNAi targeting alphasatellite DNA in the nucleus. In further support of the decoy hypothesis, the production rate of siRNAs from the least abundant DNA-S was decreased in the presence of alphasatellite, concomitant with slightly increased transcription rate of mRNA-S. This would compensate for the reduced load of DNA-S, which might become a limiting factor for helper virus encapsidation as argued above. On the other hand, production of 21, 22 and 24 nt siRNAs and antisense transcripts from DNA-R was not decreased in the presence of alphasatellite as their loads remain unaltered and their relative abundance becomes substantially higher. The same was observed for DNA-M, although at much lower levels of siRNA and antisense RNA production. In fact, the loads and transcription rates of DNA-R and DNA-M are least affected by alphasatellite, which explains unaltered RNAi activity targeting these components. Keeping

in line with the decoy hypothesis, DRC alphasatellite does not interfere with DCL or AGO activities, as neither size-class nor 5'-nucleotide profiles of viral siRNAs are altered in its presence. Consistent with these findings, alphasatellite Rep protein does not suppress transgene silencing in *N. benthamiana*.

Our phylogenetic analysis of DRC alphasatellite, placing its genus deeply within the subfamily *Nanoalphasatellitinae* in close vicinity of the genus *Fabenesatellite* comprising alphasatellites infecting legumes and other dicot Fabaceae hosts, suggests its provenance from a dicot plant. In fact, all members of *Nanoalphasatellitinae* infect dicots. Moreover, our comparison of alphasatellite Rep proteins revealed that the DRC alphasatellite is also related to alphasatellites of the genus *Gosmusatellite* (subfamily *Geminialphasatellitinae*) which infect dicot Malvaceae and Asteraceae hosts. In further support of the dicot origin hypothesis, DRC alphasatellite and BBTV clones could co-infect *N. benthamiana*, a dicot plant of the family Solanaceae, and the alphasatellite DNA alone could persist in symptomless *N. benthamiana* following recovery from BBTV disease. Such persistence of alphasatellite in the absence of helper virus can be explained by its ability to self-replicate and to evade plant defense responses.

So far BBTV was not reported to infect any dicot plant, consistent with the fact that its specialized vector *P. nigronervosa* (banana aphid) is restricted to monocots of the families Musaceae, Araceae and Zingiberaceae, two of which (Musaceae and Zingiberaceae) contain known BBTV hosts. Our findings raise a possibility that BBTV can potentially be transmitted by banana aphids to some dicot plant(s), where it could have encountered a DRC alphasatellite progenitor being associated with a nanovirus. Even a short-term feeding of viruliferous banana aphids on such a nanovirus and alphasatellite co-infected plant would be sufficient to acquire an encapsidated alphasatellite DNA and transmit it back to a banana or other BBTV host plant. Alternatively, some other aphid species feeding on a nanovirus and alphasatellite host plant could have transmitted a DRC alphasatellite progenitor to a BBTV-infected banana or other plant. It should be noted that the origin of alphasatellites associated with BBTV in South-East Asia is not known, although phylogenetically they are all classified into the subfamily *Petromoalphasatellitinae* comprising alphasatellites of perennial tropical monocots. Rep proteins of these alphasatellites are only distantly related to those of DRC alphasatellite and other members of *Nanoalphasatellitinae.* It remains to be investigated by more comprehensive surveys of cultivated and non-cultivated plants how frequent in their evolution alphasatellites could swap genera (and perhaps even families) of helper viruses and thereby expand their host plant range. Likewise, further investigation of molecular mechanisms underlying alphasatellite-

helper virus-host plant interactions and insect vector transmission of alphasatellite-helper virus complexes should shed more light on costs and benefits of alphasatellites for 6-to-8-component viruses of the family *Nanoviridae*, in comparison with 1-to-2-component viruses of the family *Geminiviridae*.

Materials and Methods

Banana aphids and plants

The clonal populations of *Pentalonia nigronervosa* aphids established and maintained at PHIM originate from Democratic Republic of the Congo (DRC aphids) and Gabon (GAB aphids). The DRC aphids were collected separately from symptomless and BBTD-infected banana plantain plants (10 adult insects per plant) in the DRC province Congo Central (village Boko) in December 2016. The GAB aphids were kindly provided by Dr. Sebastien Massart in 2017, after their rearing on healthy banana plants for >15 years at the University of Liege (Gembloux, Belgium).

In vitro seedlings of *Musa acuminata* triploid (AAA) banana cultivar Cavendish obtained from Vitropic (Saint Mathieu de Treviers, France) were grown under the tropical conditions (26°C, 70-80 % humidity) and then taken for aphid rearing and transmission tests.

Aphid transmission tests

The DRC aphids collected from the BBTD-infected and symptomless banana plants were placed onto Cavendish seedlings (at a 4-leaf stage) in two separate insect-proof plastic boxes (box 4 for the infected plant aphids and box 5 for the symptomless plant aphids). The boxes were placed in a phytochamber at 23°C/19°C day/night temperature and 16/8 hours day/night length and after three months moved into another phytochamber at 24°C (day and night), 70% humidity, 12/12 hours day/night length. After about two-three weeks, one of the three plants in the box 4 (plant p4.3) started to develop the characteristic BBTD symptoms and later got severely infected (**S1A Fig**). This plant was used as a source plant for the transmission tests depicted in **S2 Fig**.

To perform BBTV transmission using viruliferous aphids grown on the source plant p4.3, about 15 adult insects from this plant were placed on healthy *M. acuminata* seedlings. The recipient

plants were maintained in insect-proof plastic boxes in a phytochamber at 24°C (day and night), 70 % humidity, 12/12 hrs day/night length.

To carry out BBTD transmission using virus-free DRC and GAB aphids reared on healthy Cavendish plants, about 30-40 adult insects were taken from those plants and placed on a detached leaf from the source plant p4.3 for virus acquisition. After 24 hours, the insects were transferred to healthy Cavendish seedlings for virus transmission as described above. The leaf tissues and aphids were sampled at different time points for (IC-)PCR and RCA analyses (**S1B-D Fig**).

To test if a single aphid can transmit BBTV, adult aphids were selected from BBTD-infected plants (plant p4.4 of the box 4 for DRC aphids and plant 1 of the box 2 for GAB aphids). Each aphid was placed on a banana leaf surface in a micro-container (aphid placed under the leaf) allowing it to feed (but not escape). After one day, each aphid was collected for PCR analysis of viral DNA-R. The recipient plants were taken from the 24ºC chamber and, after chemical treatment to kill the aphids, placed in the greenhouse to monitor the development of BBTV symptoms. One month post-inoculation, upper leaf samples were taken for IC-PCR analysis of viral DNA-R (**S1D-E Fig**).

To study the impact of DRC alphasatellite on BBTV transmission, two consecutive large-scale transmission experiments (C and D) were performed. The BBTV-infected source plants with and without alphasatellite were selected using duplex IC-PCR with DNA-R and alphasatellite specific primers (**S1 Table**) from recipient plants of the above-described transmission experiments with DRC aphids. The transmission experiments were performed in a humid phytochamber under the tropical conditions as described above. Virus-free DRC aphids were placed on each source plant to feed for three weeks and build a colony of viruliferous aphids. Adult aphids born on the source plant were then taken from each plant to inoculate 12 recipient Cavendish seedlings (5 aphids per recipient plant) grown inside an insect-proof cage (one cage per source plant). After 4 days of feeding, all aphids were collected from each recipient plant in pools and frozen. The follow-up monitoring of the recipient plants as well as molecular analyses of the recipient plants and the viruliferous aphids are described in the Results section and detailed in **S2 Dataset**. To analyse statistical significance of alphasatellite impact on BBTV transmission, we used the R package lsmeans (Lenth 2016) to calculate least-square means using a linear model where the success of viral transmission was a function of the transmission experiment and the presence of alphasatellite in source plants. The impact of alphasatellite on

delay in appearance of the first and systemic symptoms was evaluated using a non-parametric Kruskal-Wallis test (Kruskal and Wallis 1952).

Immuno-capture (IC)-PCR analysis of encapsidated viral DNA

To detect viral DNA encapsidated in BBTV virions, an IC-PCR test was performed using the commercial coating antibody for BBTV (Agdia; http://www.agdia.com). Sterile polypropylene 0.2 ml microfuge tubes were coated with 25 µl of the BBTV antibody diluted at 1:100 in carbonate buffer [15 mM Na_2CO_3, 34 mM $NaHCO_3$, pH 9.7] for 4 hrs at 37°C (or overnight at 4°C), followed by washing three times with 200 µl of 1X PBS-T buffer [8 mM Na_2HPO_4, 150 mM NaCl, 2 mM KH_2PO_4, 3 mM KCl, 0.05 % Tween 20, pH 7.4]. Plant extracts were prepared by grinding 0.5 g fresh leaf tissue in 5 ml grinding buffer [2 % polyvinylpyrrolidone (PVP), 0.2 % sodium sulfite, 0.2 % bovine serum albumin in 1X PBS-T buffer] in plastic grinding bags (Agdia). Aliquots of the plant extracts (25 µl) were centrifuged at 7000 rpm for 3 min at room temperature and the supernatant was loaded into the BBTV-antibody coated tubes. The tubes were incubated at 37°C for 4 hrs or at 4°C overnight and then washed three times with 200 µl of 1X PBS-T buffer and one time with 200 µl of RNase-free water, followed by brief air-drying. Then, 30 µl DNase solution [3 µl of 10X buffer (400 mM Tris-HCl pH 8.0, 100 mM $MgSO_4$, 10 mM $CaCl_2$), 3 µl of RQ1 RNase-Free DNase (1U/µl, Promega), and 24 µl water] was added to remove potential contaminant DNA, followed by incubation for 30 min to 1 hr at 37°C. The DNase was inactivated at 95°C 10 min and the solution was removed by pipetting. The PCR reactions were carried out directly in the tube in a total volume of 25 µl [5 µl of 5X Green GoTaq Reaction Buffer (Promega), 10 mM forward and reverse primers (or 2 pairs of primers in the case of duplex PCR; **S1 Table**), 10 mM dNTPs (2.5 mM each), 0.2 µl (5 U/µl) GoTaq DNA Polymerase (Promega), and mili-Q water to adjust the volume]. The reaction mixture was denatured at 98°C for 5 min, followed by 30 cycles of 30 sec at 95°C, 30 sec at 58°C and 30 sec at 72°C, and finally one cycle at 72°C for 10 min.

Total DNA extraction from banana leaves

Fresh leaf tissue (100 mg) was ground in liquid nitrogen and 500 µl extraction buffer [100 mM Tris-HCl pH 8.0, 1.4 M NaCl, 20 mM EDTA, 2 % alkyltrimethylamonium bromide (MATAB), 1 % polyethyleneglycol 6000, 0.5 % sodium sulfite] pre-heated at 74°C and supplemented with

0.4 µl RNase (100 mg/ml) was added to the frozen powder. The mixture was vortexed for 20 sec, 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 13000 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 the appearance of a hank and then spun at 13000 rpm and 4°C for 30 min. The pellet was washed twice with 500 µl of 70 % ethanol, air dried and dissolved in 100 µl of mili-Q water.

Total DNA extraction from aphids

Total DNA was extracted from alive aphids or aphids stored at -80°C. Aphids were grounded in 150 µl TNES buffer (50 mM Tris-HCl pH 7.5, 400 mM NaCl, 20 mM EDTA, 0.5 % SDS) with a pestle. The resulting crude extract was mixed with 45 µl of cold 5 M NaCl, followed by centrifugation at 6500 rpm for 10 min at 4°C. The supernatant was mixed with 500 µl of cold absolute ethanol and incubated for 20 min at -80°C, followed by centrifugation at 14000 rpm for 20 min at 4°C. The pellet was washed with 250 µl of 70 % ethanol, followed by centrifugation at 14000 rpm for 10 min at 4°C. The supernatant was discarded, and the pellet was dried and resuspended in 50 µl H_2O.

Rolling circle amplification (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 µl Sample buffer and 1 µl total DNA extracted from aphids or banana leaf tissues were mixed and heated at 95°C for 3 min. The samples were cooled and 5 µl Reaction buffer and 0.2 µl enzyme mix were added, followed by incubation at 30°C for 18 hours. The enzyme was inactivated by heating at 65°C for 10 min.

Restriction analysis and purification of RCA products

RCA products were digested with the restriction enzyme (RE) AvaI or DraI as follows: 4 µl RCA product were mixed with 1.5 µl of RE buffer and 1 µl (1 U/ µl) of RE in a final volume of 15 µl and incubated at 37°C for 1 hr. The digested products were analysed by agarose gel

electrophoresis (**S1G-H Fig**). Undigested RCA products were purified using NucleoSpin Gel and PCR Clean-up kit (Macherey-Nagel, USA) following the manufacturer protocol. DNA concentration was measure by Qubit fluorimeter using Qubit dsDNA HS Assay Kit (Thermo Fischer Scientific).

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

Five ng of the cleaned undigested RCA products were taken for Illumina sequencing at Fasteris AG (www.fasteris.com). Libraries were prepared using Nextera XT standard DNA protocol and 11 libraries (JGF1-11, **S3 Fig**) were multiplexed and sequenced in one flowcell of HiSeq2500 with a 2x 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, https://www.ebi.ac.uk/~zerbino/velvet/) 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 BWA (Burrows and Wheeler 1994) and visualisation using MISIS-2 (Seguin et al. 2016).

PCR and qPCR analyses of viral DNA in plants and aphids

BBTV diagnostics in banana plants and aphids was performed by PCR using a pair of primers (5'-GGCGCGATATGTGGTATGCTGG and 5'-CCAAACTCGAAGGGACCTTCG) specific for a DNA-R region conserved in BBTV isolates from DRC (Mukwa et al. 2016) yielding a 285 bp product. The PCR reaction was performed in a volume of 25 µl containing 1 µl total DNA (ca. 40 ng plant DNA or ca. 10 ng of aphid DNA), 2 µl primer mix (10 µM each primer), 1 µl of 2.5 mM dNTPs, 5 µl of 5x GoTaq buffer and 1U GoTaq DNA polymerase (Promega). After denaturation at 94°C for 5 min, DNA was amplified for 30 cycles of 30 sec at 94°C, 30 sec at 60°C, 30 sec at 72°C, followed by a final extension at 72°C for 10 min.

Quality of total DNA extracted from banana samples, was verified by PCR with a pair of primers (5'-TCCTTTCGCTCTATGCCAGT and 5'-GCCCATCGGGAAGTTCATAG) specific the actin-101-like gene (NCBI Genbank accession XM_009410731).

For duplex and multiplex PCR analysis of the *de novo* reconstructed BBTV components and alphasatellite from the DRC-2016 isolate, the primers (**S1 Table**) were designed using the web-tool Multalin (http://multalin.toulouse.inra.fr/), considering amplicon sizes to be differentiated on 2% agarose gel, and verified using the NCBI Primer-Blast. Degenerate primers for alphasatellite detection (**S6 Fig**; **S1 Table**) were designed by multiple sequence alignment of DRC-2016 alphasatellite to all alphasatellites of the genus *Fabenesatellite* (Briddon et al. 2018) using Seaview (Gouy et al. 2010) to identify conserved regions for primer annealing sites with minimal SNPs.

Quantitative (q)PCR with specific primers designed with the web-tool Primers3 v.0.4.0 (https://bioinfo.ut.ee/primer3-0.4.0/) for each BBTV core component and DRC-alphasatellite as well as banana RPS2 or aphid EF1a housekeeping gene (**S1 Table**) was carried out in 384-well plates using a fluorescent label SYBR Green (Roche) and LightCycler 480 thermal cycler (Roche). The reaction mixture was denatured at 95°C for 5 min, followed by 40 cycles of 10 sec at 95°C, 10 sec at 56°C and 10 sec at 72°C. Each sample was replicated twice on the same plate. The fluorescence was converted into ng of DNA by referring to a standard curve carried out on each plate (one curve per amplicon). Each standard curve consisted of five dilutions of a mixture of the PCR products obtained for each BBTV component, DRC alphasatellite and housekeeping gene separately with their qPCR primer pairs. The qPCR data were analysed using LinRegPCR 12.11.0.0 (Rutledge and Stewart 2008) to calculate DNA concentrations. The concentrations were normalized by internal housekeeping genes using Excel to calculate the loads of each BBTV component and alphasatellite. Statistical analysis of the data was performed with Rstudio (Team 2015) using a non-parametric Kruskal-Wallis test (Kruskal and Wallis 1952), because DNA concentrations and normalized loads of viral genome components in plant and aphid samples did not follow a normal distribution. Figures were prepared using the R packages ggplot2 (Wickham 2008).

Construction of infectious clones of DRC-2016 alphasatellite and BBTV components

The BBTV components and alphasatellite were cloned from the total DNA extracted from viruliferous DRC-2016 aphids (sample JGF-5, **S2 Fig**) using inverse overlapping PCR primers designed on a unique restriction site in each viral component (**S1 Table**). The PCR products were digested with the respective restriction enzymes and cloned into pMiniT 2.0 vector using NEB PCR cloning kit. A bitmer fragment of each viral component (except DNA-R) containing

the common-region stem-loop (CR-SL) was initially cloned from the pMiniT2.0 constructs into pBluescript II SK(-) and the full-length onemer fragment was subsequently cloned in the bitmer clone to construct the partial tandem dimer clones. In the case of DNA-R, the full-length insert was directly cloned from the pMiniT2.0 backbone into pBluescript II SK(-) and a clone with two copies of DNA-R in tandem repeat was chosen as the dimer clone. The partial dimer and the dimer inserts were subcloned from the pBluescript II SK(-) constructs into the binary vector pCAMBIA2300 using convenient restriction sites (**S9 Fig**). All the clones were verified by both restriction analysis and Sanger sequencing after each step of cloning. *Agrobacterium tumefaciens* strain EHA105 was transformed with these binary plasmids and the intactness of the viral inserts in the resulting agrobacterial strains was again confirmed by both restriction analysis and Sanger sequencing.

Agroinoculation of *N. benthamiana* plants

N. benthamiana wild type and 16c line seeds were sown in a nursery pot and maintained in a phytochamber at 20-21°C, 16 hrs daylight and 8 hrs darkness. Then, seedlings were transplanted in individual pots and 3 weeks-old seedlings were used for agroinoculation. One fresh colony from each strain of EHA105 *A. tumefaciens* carrying the binary construct was inoculated into pre-culture of 2 ml Luria-broth (LB) medium supplied with appropriate antibiotics and grown for 8 hrs at 28°C. Then, all the pre-cultures were added in 50 ml of LB medium supplied with appropriate antibiotics and grown overnight at 28°C. Bacterial cells were pelleted and resuspended in agro-inoculation buffer [10 mM MES, 10 mM MgCl2, 100 µM acetosyringone] at optical density at 600 nm (OD600) = 1. Mixture of agrobacteria at OD600 = 1 containing all six BBTV components +/- alphasatellite in equal proportions was then prepared and placed in darkness at room temperature for 3 hrs and then 1 ml of the mixture was used for agro-infiltration of younger leaves of *N. benthamiana* with a syringe. The inoculated plants were kept in a phytochamber at the above-described conditions and leaf tissue samples were harvested at 8, 21 and 60 days post inoculation (dpi) for molecular analysis.

Transient expression analysis of BBTV and alphasatellite proteins in *N. benthamiana*

The ORFs of the DRC-2016 alphasatellite and BBTV components were PCR amplified from either the pMiniT2.0 or the pBluescript II SK(-) constructs and cloned under the control of

CaMV 35S promoter and terminator in the binary plasmid pB7WG2 as detailed in **S10A-B Fig**. These plasmids were transformed into *Agrobacterium tumefaciens* strain GV3101 and their intactness in the resulting agrobacterial strains was confirmed by restriction analysis and Sanger sequencing of the viral ORF inserts. Transient expression analysis of the viral ORF constructs was performed using a classical silencing suppression assay in leaves of *N. benthamiana* 16c line plants as we described in detail previously (Rajeswaran et al. 2014b).

Total RNA extraction from banana leaves

RNA extraction was performed as described previously (Rajeswaran et al. 2014a). Briefly, 0.5 g fresh leaf tissue was ground in liquid nitrogen and mixed with 2.5 ml extraction buffer [100 mM Tris-HCl (pH 7.5), 500 mM NaCl, 25 mM EDTA (pH 8.0), 1.5 % SDS, 2 % polyvinylpyrrolidone, 0.7 % beta-mercaptoethanol (added just before use)]. The mixture was incubated for 10 min at room temperature and then centrifuged at 3700 rpm for 15 min at 4°C. The supernatant was taken and mixed with 1/3 volume of 5 M sodium acetate (pH 6.0) and kept on ice for 30 min, followed by centrifugation at 13000 rpm for 15 min at 4°C. The supernatant was taken and mixed vigorously with an equal volume of phenol/chloroform/isoamyl alcohol (25:24:1), followed by centrifugation at 13000 rpm for 10 min at 4°C. The aqueous phase was taken and mixed vigorously with one volume of chloroform/isoamyl alcohol (24:1), followed by centrifugation as described above. The aqueous phase was collected and mixed with one volume of isopropanol, followed by incubation for 5 min at room temperature and for 30 min at -80°C. RNA was pelleted by centrifugation for 15 min at a maximum speed at 4°C. The pellet was washed with 500 µl of 75 % ethanol, air-dried and dissolved in 50 µl RNase-free water.

Illumina sequencing and bioinformatics analysis of viral transcriptome and small RNA-ome

Integrity of high and low molecular weight 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 same total RNA extracts for library preparations with (i) Illumina stranded mRNA protocol (DNase treatment, poly-A selection, transcripts fragmentation, 1st strand cDNA synthesis with random primers, 2nd strand cDNA synthesis, ligation of indexed adapters, amplification of the library) and (ii) 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 mRNA libraries were multiplexed and sequenced in one flowcell of NovaSeq 6000, yielding 28'568'938 to 42'820'006 75 nt paired-end reads with Q30 = 92.6 to 93.5. The sRNA libraries were multiplexed and sequenced in one flowcell of NovaSeq 6000 yielding 33'698'415 to 50'148'596 reads with Q30 = 95.5 to 95.9. In both cases, the libraries were de-multiplexed, followed by adapter trimming. The resulting reads were mapped using Burrow-Wheeler Aligner (BWA) 0.7.12 (Burrow and Wheeler 1994) onto the reference sequences of DRC-2016 alphasatellite and six BBTV components with and without mismatches. Mapped viral reads were sorted by polarity (forward, reverse) and, in the case of sRNAs, by size (15 to 34 nts) and 5´-terminal nucleotide identity (5'U, 5'A, 5'G, 5'C) and counted, followed by normalization in reads per million (RPM) of total (viral + non-viral) reads (**S3 Dataset** and **S4 Dataset**). Additionally, mRNA-seq reads were mapped with BWA onto the reference sequences of the BBTV and alphasatellite mRNA units (spanning from the transcription start site and the polyadenylation site indicated in **Fig 6** and **S13 Fig**; note that in the case of U3 the shorter unit was taken). The mapped forward reads representing viral mRNAs, viral antisense transcripts (mapped on full-length references) and total viral siRNAs were counted in RPM and the resulting values were divided by viral DNA load values (qPCR concentrations/RPS2) for the respective components and alphasatellite to calculate transcription rates of viral mRNAs by Pol II and antisense transcription rates. Likewise, siRNA production rates were calculated as counts (in RPM) of viral sense+antisense siRNAs (mapped to full-length references) divided by viral DNA loads. Mean values with standard deviations of the mRNA, antisense transcript, total siRNA and siRNA size-class read counts (in RPM) and the mRNA transcription rates and siRNA production rates in three biological replicates per condition (+/- alphasatellite) were calculated and plotted in **Fig 7** and **Fig 10**. Relative abundances of viral mRNAs, viral antisense transcripts, total siRNAs and size-classes of siRNAs from six BBTV components and alphasatellite were calculated as percentage of reads representing each component in total BBTV and virome (BBTV + alphasatellite) and the mean values of three biological replicates with standard deviations plotted in "BBTV formula" and "Virome formula" panels of **Fig 7**, **Fig 8** and **Fig 10**.

Single-nucleotide resolution maps of viral mRNA reads (75 nt) and viral siRNAs (a 20-25 nt range) shown in **Fig 6**, **Fig 9**, **S13 Fig** and **S14 Fig** were generated using MISIS-2 (Seguin et al. 2016) that analyses BAM files of the mapped reads yielding table files (**S5 Dataset** and **S6**

Dataset) and visualizes mapped reads and SNPs at each position of the reference sequence (see **S13 Fig**).

Phylogenetic analysis of alphasatellites

Reference sequences of all isolates of alphasatellites available in the NCBI Genbank in September 2021 were downloaded, adjusted to begin with the conserved nonanucleotide sequence TAGTATTAC and aligned to the DRC alphasatellite sequences (OK546211 and OK546212) using Sequence Demarcation Tool (SDT) v1.2 with Muscles (Muhire et al. 2014). A maximum likelihood phylogenetic tree was constructed using Seaview (Gouy et al. 2010) with Muscles 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).

Data availability

Viral genome sequences obtained in this study were deposited in the NCBI Genbank with the accession numbers OK546211 (DRC-2016 alphasatellite), OK546212 (DRC-2012 alphasatellite), OK546213-OK546218 (DRC-2016 BBTV: DNAs C, M, N, R, S, U3) and OK546219-OK546224 (DRC-2012 BBTV: DNAs C, M, N, R, S, U3).

Acknowledgements

We thank Myrianne Andeime (Institut de Recherches Agronomiques et Forestières, Gabon) and Aman Omondi (Bioversity International, Benin) for providing samples from Gabon and Benin, respectively, and Loic Baerlocher and Sebastien Ravel for bioinformatics assistance. The work was supported by grants from the CGIAR Research Program on Roots, Tubers and Bananas and the CGIAR Trust Fund contributors to L.K., M-L.I-C. and M.M.P. (members of the Alliance for BBTV Control in Africa - BA3.4) and by the Institute Agro (Montpellier) PhD scholarship to V.G.

Author contributions

M.M.P. and M-L.I-C. conceived and designed the study; V.G., R.R., H.C.C., C.K., N.L., S.G., L.F.T.M. and M-L.I-C. performed the experiments; V.G., M.K. and M.M.P. analysed bioinformatically and interpreted the sequencing data; L.K. provided biological material and financial support for R.R. and K.K.; M.M.P. and V.G. wrote the draft manuscript and prepared the Figures and the Supplementary Data files; R.R. and M-L.I-C corrected and extended the manuscript; all the authors approved the final manuscript for submission.

Competing Interests Statement

The authors declare no competing interests.

References

Abbas Q, Amin I, Mansoor S, Shafiq M, Wassenegger M, Briddon RW. 2019. The Rep proteins encoded by alphasatellites restore expression of a transcriptionally silenced green fluorescent protein transgene in Nicotiana benthamiana. Virusdisease 30: 101-105. doi: 10.1007/s13337-017-0413-5.

Akbergenov R, Si-Ammour A, Blevins T, Amin I, Kutter C, Vanderschuren H, Zhang P, Gruissem W, Meins F Jr, Hohn T, Pooggin MM. 2006. Molecular characterization of geminivirus-derived small RNAs in different plant species. Nucleic Acids Res 34: 462–471. doi: 10.1093/nar/gkj447.

Aronson MN, Meyer AD, Györgyey J, Katul L, Vetten HJ, Gronenborn B, Timchenko T. 2000. Clink, a nanovirus-encoded protein, binds both pRB and SKP1. J Virol. 74: 2967-72. doi: 10.1128/jvi.74.7.2967-2972.2000.

Amin I, Hussain K, Akbergenov R, Yadav JS, Qazi J, Mansoor S, Hohn T, Fauquet CM, Briddon RW. 2011a. Suppressors of RNA silencing encoded by the components of the cotton leaf curl begomovirus-betasatellite complex. Mol Plant Microbe Interact 24: 973–983. doi: 10.1094/MPMI-01-11-0001.

Amin I, Ilyas M, Qazi J, Bashir R, Yadav JS, Mansoor S, Fauquet CM, Briddon RW. 2011b. Identification of a major pathogenicity determinant and suppressors of RNA silencing

encoded by a South Pacific isolate of Banana bunchy top virus originating from Pakistan. Virus Genes 42: 272–281. doi: 10.1007/s11262-010-0559-3.

Aregger M, Borah BK, Seguin J, Rajeswaran R, Gubaeva EG, Zvereva AS, Windels D, Vazquez F, Blevins T, Farinelli L, Pooggin MM. 2012. Primary and Secondary siRNAs in Geminivirus-induced Gene Silencing. PLOS Pathogens 8: e1002941. doi: 10.1371/journal.ppat.1002941.

Bell KE, Dale JL, Ha CV, Vu MT, Revill PA. 2002. Characterisation of Rep-encoding components associated with banana bunchy top nanovirus in Vietnam. Arch Virol. 147: 695-707. doi: 10.1007/s007050200019.

Beetham PR, Harding RM, Dale JL. 1999. Banana bunchy top virus DNA-2 to 6 are monocistronic. Arch Virol. 144: 89-105. doi: 10.1007/s007050050487.

Beetham PR, Hafner GJ, Harding RM, Dale JL. 1997. Two mRNAs are transcribed from banana bunchy top virus DNA-1. J Gen Virol. 78: 229-36. doi: 10.1099/0022-1317-78-1-229. doi: 10.1099/0022-1317-78-1-229

Blevins T, Rajeswaran R, Aregger M, Borah BK, Schepetilnikov M, Baerlocher L, Farinelli L, Meins F, Hohn T, Pooggin MM. 2011. Massive production of small RNAs from a non-coding region of Cauliflower mosaic virus in plant defense and viral counter-defense. Nucleic Acids Res 39: 5003–5014. doi: 10.1093/nar/gkr119.

Blevins T, Rajeswaran R, Shivaprasad PV, Beknazariants D, Si-Ammour A, Park HS, Vazquez F, Robertson D, Meins F Jr, Hohn T, Pooggin MM. 2006. Four plant Dicers mediate viral small RNA biogenesis and DNA virus induced silencing. Nucleic Acids Res 34: 6233–6246. doi: 10.1093/nar/gkl886.

Borges F, Martienssen RA. 2015. The expanding world of small RNAs in plants. Nat Rev Mol Cell Biol 16: 727–741. doi: 10.1038/nrm4085

Briddon RW, Martin DP, Roumagnac P, Navas-Castillo J, Fiallo-Olivé E, Moriones E, Lett JM, Zerbini FM, Varsani A. 2018. Alphasatellitidae: a new family with two subfamilies for the classification of geminivirus- and nanovirus-associated alphasatellites. Arch Virol. 163:2587-2600. doi: 10.1007/s00705-018-3854-2.

Briddon RW, Stanley. 2006. Subviral agents associated with plant single-stranded DNA viruses. Virology 344: 198–210. doi: 10.1016/j.virol.2005.09.042.

Burns TM, Harding RM, Dale JL. 1994. Evidence that banana bunchy top virus has a multiple component genome. Arch Virol 137: 371–380. doi: 10.1007/BF01309482.

Burns TM, Harding RM, Dale JL. 1995. The genome organization of banana bunchy top virus: Analysis of six ssDNA components. J Gen Virol 76: 1471–1482. doi: 10.1099/0022-1317-76-6-1471.

Burrows M, Wheeler DJ. 1994. A Block-Sorting Lossless Data Compression Algorithm. Technical Report 124, Digital Equipment Corporation, Digital Equipment Corporation Press, Palo Alto, CA, U.S.A. Pages 1-24.

Dale JL. 1987. Banana bunchy top: An economically important tropical plant virus disease. Adv Virus Res 33: 301–325. doi: 10.1016/s0065-3527(08)60321-8.

Denham TP, Haberle SG, Lentfer C, Fullagar R, Field J, Therin M, Porch N, Winsborough B. 2003. Origins of agriculture at Kuk Swamp in the highlands of New Guinea. Science 301: 189-93. doi: 10.1126/science.1085255.

Di Mattia J, Vernerey M-S, Yvon M, Pirolles E, Villegas M, Gaafar Y, Ziebell H, Michalakis Y, Zeddam J-L, Blanc S. 2020. Route of a Multipartite Nanovirus across the Body of Its Aphid Vector. J Virol. 94: e01998-19. doi: 10.1128/JVI.01998-19.

Frickey T, Lupas A. 2004. CLANS: A Java application for visualizing protein families based on pairwise similarity. Bioinformatics 20: 3702–3704. https://doi.org/10.1093/bioinformatics/bth444

Fu HC, Hu JM, Hung TH, Su HJ, Yeh HH. 2009. Unusual events involved in Banana bunchy top virus strain evolution. Phytopathology 99: 812-22. doi: 10.1094/PHYTO-99-7-0812.

Gouy M, Guindon S, Gascuel O. 2010. Sea view version 4: A multiplatform graphical user interface for sequence alignment and phylogenetic tree building. Mol Biol Evol. 27: 221–224. https://doi.org/10.1093/molbev/msp259.

Grigoras I, Vetten HJ, Commandeur U, Ziebell H, Gronenborn B, Timchenko T. 2018. Nanovirus DNA-N encodes a protein mandatory for aphid transmission. Virology 522: 281–291. doi: 10.1016/j.virol.2018.07.001.

Gronenborn B. 2004. Nanoviruses: genome organisation and protein function. Vet Microbiol. 98: 103-9. doi: 10.1016/j.vetmic.2003.10.015.

Gronenborn B, Randles JW, Knierim D, Barriere Q, Vetten HJ, Warthmann N, Cornu D, Sileye T, Winter S, Timchenko T. 2018. Analysis of DNAs associated with coconut foliar decay disease implicates a unique single-stranded DNA virus representing a new taxon. Sci Rep 8: 5698. doi: 10.1038/s41598-018-23739-y.

Hafner GJ, Harding RM, Dale JL. 1995. Movement and transmission of banana bunchy top virus DNA component one in bananas. J Gen Virol 76: 2279–2285. doi: 10.1099/0022-1317-76-9-2279.

Hafner GJ, Harding RM, Dale JL. 1997. A DNA primer associated with banana bunchy top virus. J Gen Virol 78: 479–486. doi: 10.1099/0022-1317-78-2-479.

Hasanvand V, Heydanejad J, Massumi H, Kleinow T, Jeske H, Fontenele RS, Kraberger S, Varsani A. 2021. Genome characterization of parsley severe stunt-associated virus in Iran. Virus Genes 57: 293–301. https://doi.org/10.1007/s11262-021-01835-y.

Hassan-Sheikhi P, Heydarnejad J, Massumi H, Kraberger S, Varsani A. 2020. Novel nanovirus and associated alphasatellites identified in milk vetch plants with chlorotic dwarf disease in Iran. Virus Res. 276: 197830. doi: 10.1016/j.virusres.2019.197830.

Herrera-Valencia VA, Dugdale B, Harding RM, Dale JL. 2006. An iterated sequence in the genome of Banana bunchy top virus is essential for efficient replication. J Gen Virol. 87: 3409-3412. doi: 10.1099/vir.0.82166-0

Herrera-Valencia VA, Dugdale B, Harding RM, Dale JL. 2007. Mapping the 5' ends of banana bunchy top virus gene transcripts. Arch Virol. 152: 615-20. doi: 10.1007/s00705-006-0889-6.

Heydarnejad J, Kamali M, Massumi H, Kvarnheden A, Male MF, Kraberger S, Stainton D, Martin DP, Varsani A. 2017. Identification of a Nanovirus-Alphasatellite Complex in Sophora alopecuroides. Virus Res. 235: 24-32. doi: 10.1016/j.virusres.2017.03.023.

Hooks CRR, Wright MG, Kabasawa DS, Manandhar R, Almeida. 2008. Effect of banana bunchy top virus infection on morphology and growth characteristics of banana. Ann Appl Biol 153: 1–9. https://doi.org/10.1111/j.1744-7348.2008.00233.x

Horser CL, Harding RM, Dale JL. 2001a. Banana bunchy top nanovirus DNA-1 encodes the 'master' replication initiation protein. J Gen Virol. 82: 459-464. doi: 10.1099/0022-1317-82-2-459.

Horser CL, Karan M, Harding RM, Dale JL. 2001b. Additional rep-encoding DNAs associated with banana bunchy top virus. Arch Virol. 146: 71-86. doi: 10.1007/s007050170192.

Hu JS, Wang M, Sether D, Xie W, Leonhardt KW. 1996. Use of polymerase chain reaction (PCR) to study transmission of banana bunchy top virus by the banana aphid (Pentalonia nigronervosa). Ann Appl Biol 128: 55–64. https://doi.org/10.1111/j.1744-7348.1996.tb07089.x

Idris AM, Shahid MS, Briddon RW, Khan AJ, Zhu JK, Brown JK. 2011. An unusual alphasatellite associated with monopartite begomoviruses attenuates symptoms and reduces betasatellite accumulation. J Gen Virol. 92: 706-17. doi: 10.1099/vir.0.025288-0.

Kraberger S, Kumari SG, Najar A, Stainton D, Martin DP, Varsani A. 2018. Molecular characterization of faba bean necrotic yellows viruses in Tunisia. Arch Virol. 163: 687-694. doi: 10.1007/s00705-017-3651-3.

Kruskal WH, Wallis WA. 1952. Use of Ranks in One-Criterion Variance Analysis. Journal of the American Statistical Association 47: 583–621. https://doi.org/10.1080/01621459.1952.10483441

Karan M, Harding RM, Dale JL. 1994. Evidence for two groups of banana bunchy top virus isolates. J Gen Virol 75: 3541–3546. doi: 10.1099/0022-1317-75-12-3541.

Kumar PL, Hanna R, Alabi OJ, Soko MM, Oben TT, Vangu GHP, Naidu RA. 2011. Banana bunchy top virus in sub-Saharan Africa: Investigations on virus distribution and diversity. Virus Res 159: 171–182. doi: 10.1016/j.virusres.2011.04.021.

Kumar J, Kumar J, Singh SP, Tuli R. 2014. Association of satellites with a mastrevirus in natural infection: complexity of Wheat dwarf India virus disease. J Virol. 88: 7093-104. doi: 10.1128/JVI.02911-13.

Kumar PL, Selvarajan R, Iskra-Caruana ML, Chabannes M, Hanna R. 2015. Biology, etiology, and control of virus diseases of banana and plantain. Adv Virus Res. 91: 229-69. doi: 10.1016/bs.aivir.2014.10.006.

Lageix S, Catrice O, Deragon JM, Gronenborn B, Pélissier T, Ramírez BC. 2007. The nanovirus-encoded Clink protein affects plant cell cycle regulation through interaction with the retinoblastoma-related protein. J Virol. 81: 4177-85. doi: 10.1128/JVI.02103-06.

Lenth RV. 2016. Least-Squares Means: The R Package lsmeans. Journal of Statistical Software 69: 1-33. doi:10.18637/jss.v069.i01

Magee CJP. 1940. Transmission Studies on the Banana Bunchy-top Virus. J Aust Inst Agric Sci 6: 109-110. https://www.cabi.org/isc/abstract/19410500254

Mandal B, Shilpi S, Barman AR, Mandal S, Varma A. 2013. Nine novel DNA components associated with the foorkey disease of large cardamom: evidence of a distinct babuvirus species in Nanoviridae. Virus Res. 178: 297-305. doi: 10.1016/j.virusres.2013.09.027.

Mansourpour M, Gallet R, Abbasi A, Blanc S, Dizadji A, Zeddam JL. 2021. Effects of an alphasatellite on life cycle of the nanovirus Faba bean necrotic yellows virus. J Virol. Nov 24:JVI0138821. doi: 10.1128/JVI.01388-21.

Mar TB, Mendes IR, Lau D, Fiallo-Olivé E, Navas-Castillo J, Alves MS, Murilo Zerbini F. 2017. Interaction between the New World begomovirus Euphorbia yellow mosaic virus and its associated alphasatellite: effects on infection and transmission by the whitefly Bemisia tabaci. J Gen Virol. 98: 1552-1562. doi: 10.1099/jgv.0.000814.

Matzke MA, Kanno T, Matzke AJ. 2015. RNA-Directed DNA Methylation: The Evolution of a Complex Epigenetic Pathway in Flowering Plants. Annu Rev Plant Biol. 66: 243-67. doi: 10.1146/annurev-arplant-043014-114633.

Muhire BM, Varsani A, Martin DP. 2014. SDT: A virus classification tool based on pairwise sequence alignment and identity calculation. PLoS One 9: e108277. doi: 10.1371/journal.pone.0108277.

Mukwa L, Muengula M, Zinga I, Kalonji A, Iskra-Caruana ML, Bragard C 2014. Occurrence and Distribution of Banana bunchy top virus Related Agro-Ecosystem in South Western, Democratic Republic of Congo. American Journal of Plant Sciences 5: 647-658. doi: 10.4236/ajps.2014.55079.

Mukwa LF, Gillis A, Vanhese V, Romay G, Galzi S, Laboureau N, Kalonji-Mbuyi A, Iskra-Caruana ML, Bragard C. 2016. Low genetic diversity of Banana bunchy top virus, with a sub-regional pattern of variation, in Democratic Republic of Congo. Virus Genes 52: 900-905. doi: 10.1007/s11262-016-1383-1.

Nawaz-ul-Rehman MS, Nahid N, Mansoor S, Briddon RW, Fauquet CM. 2010. Post-transcriptional gene silencing suppressor activity of two non-pathogenic alphasatellites associated with a begomovirus. Virology 405: 300–308. doi: 10.1016/j.virol.2010.06.024

Niu S, Wang B, Guo X, Yu J, Wang X, Xu K, Zhai Y, Wang J, Liu Z. 2009. Identification of two RNA silencing suppressors from banana bunchy top virus. Arch Virol. 154: 1775-83. doi: 10.1007/s00705-009-0515-5.

Nogueira AM, Nascimento MB, Barbosa TMC, Quadros AFF, Gomes JPA, Orílio AF, Barros DR, Zerbini FM. 2021. The Association between New World Alphasatellites and Bipartite Begomoviruses: Effects on Infection and Vector Transmission. Pathogens 10: 1244. https://doi.org/10.3390/pathogens10101244

Perrier X, De Langhe E, Donohue M, Lentfer C, Vrydaghs L, Bakry F, Carreel F, Hippolyte I, Horry JP, Jenny C, Lebot V, Risterucci AM, Tomekpe K, Doutrelepont H, Ball T, Manwaring J, de Maret P, Denham T. 2011. Multidisciplinary perspectives on banana (Musa spp.) domestication. Proc Natl Acad Sci U S A 108: 11311-8. doi: 10.1073/pnas.1102001108.

Pooggin MM. 2018. Small RNA-omics for plant virus identification, virome reconstruction, and antiviral defense characterization. Front Microbiol. 9: 2779. doi: 10.3389/fmicb.2018.02779.

Pooggin MM. 2013. How can plant DNA viruses evade siRNA-directed DNA methylation and silencing? Int J Mol Sci. 14: 15233-59. doi: 10.3390/ijms140815233.

Rahayuniati RF, Subandiyah S, Hartono S, Somowiyarjo S, Kurniawan REK, Prakoso AB, Crew K, Vance ME, Ray JD, Thomas JE. 2021. Recent distribution and diversity analysis on banana bunchy top virus of banana and alternative host in Indonesia. Trop Plant Pathol. 46: 506-51. https://doi.org/10.1007/s40858-021-00443-3

Rajeswaran R, Seguin J, Chabannes M, Duroy P-O, Laboureau N, Farinelli L, Iskra-Caruana M-L, Pooggin MM. 2014a. Evasion of short interfering RNA-directed antiviral silencing in Musa acuminata persistently infected with six distinct banana streak pararetroviruses. J Virol 88: 11516-11528. doi: 10.1128/JVI.01496-14.

Rajeswaran R, Golyaev V, Seguin J, Zvereva AS, Farinelli L, Pooggin MM. 2014b. Interactions of Rice tungro bacilliform pararetrovirus and its protein P4 with plant RNA-silencing machinery. Mol Plant Microbe Interact. 27: 1370-8. doi: 10.1094/MPMI-07-14-0201-R.

Rutledge RG, Stewart D. 2008. Critical evaluation of methods used to determine amplification efficiency refutes the exponential character of real-time PCR. BMC Mol Biol 9: 1–12. https://doi.org/10.1186/1471-2199-9-96

Seguin J, Otten P, Baerlocher L, Farinelli L, Pooggin MM. 2016. MISIS-2: A bioinformatics tool for in-depth analysis of small RNAs and representation of consensus master genome in viral quasispecies. J Virol Methods 233: 37-40. doi: 10.1016/j.jviromet.2016.03.005.

S**tainton D, Martin DP, Muhire BM, Lolohea S, Halafihi M, Lepoint P, Blomme G, Crew KS, Sharman M, Kraberger S, Dayaram A, Walters M, Collings DA, Mabvakure B, Lemey P, Harkins GW, Thomas JE, Varsani A. 2015**. The global distribution of Banana bunchy top virus reveals little evidence for frequent recent, human-mediated long distance dispersal events. Virus Evol. 1: vev009. doi: 10.1093/ve/vev009.

Thomas JE, Gronenborn B, Harding RM, Mandal B, Grigoras I, Randles JW, Sano Y, Timchenko T, Vetten HJ, Yeh HH, Ziebell H, Ictv Report Consortium. 2021. ICTV Virus Taxonomy Profile: Nanoviridae. J Gen Virol. 102(3). doi: 10.1099/jgv.0.001544.

Timchenko T, Katul L, Aronson M, Vega-Arreguín JC, Ramirez BC, Vetten HJ, Gronenborn B. 2006. Infectivity of nanovirus DNAs: induction of disease by cloned genome components of Faba bean necrotic yellows virus. Journal of General Virology 87: 1735–1743. doi: 10.1099/vir.0.81753-0.

Timchenko T, de Kouchkovsky F, Katul L, David C, Vetten HJ, Gronenborn B. 1999. A single rep protein initiates replication of multiple genome components of faba bean necrotic yellows virus, a single-stranded DNA virus of plants. J Virol. 73: 10173-82. doi: 10.1128/JVI.73.12.10173-10182.1999.

Tsao TTH. 2008. Towards the development of transgenic banana bunchy top virus(bbtv)-resistant banana plants: interference with replication. Bachelor thesis. Queensland University of Technology, Australia. https://core.ac.uk/download/pdf/10885632.pdf

Varsani A, Martin DP, Randles JW, Vetten HJ, Thomas JE, Fiallo-Olivé E, Navas-Castillo J, Lett JM, Zerbini FM, Roumagnac P, Gronenborn B. 2021. Taxonomy update for the family Alphasatellitidae: new subfamily, genera, and species. Arch Virol. Sep 22. doi: 10.1007/s00705-021-05232-6.

Vetten HJ, Knierim D, Rakoski MS, Menzel W, Maiss E, Gronenborn B, Winter S, Krenz B. 2019. Identification of a novel nanovirus in parsley. Arch Virol. 164: 1883-1887. doi: 10.1007/s00705-019-04280-3.

Walker PJ, Siddell SG, Lefkowitz EJ, Mushegian AR, Adriaenssens EM, Alfenas-Zerbini P, Davison AJ, Dempsey DM, Dutilh BE, García ML, Harrach B, Harrison RL, Hendrickson RC, Junglen S, Knowles NJ, Krupovic M, Kuhn JH, Lambert AJ, Łobocka M, Nibert ML, Oksanen HM, Orton RJ, Robertson DL, Rubino L, Sabanadzovic S, Simmonds P, Smith DB, Suzuki N, Van Dooerslaer K, Vandamme AM, Varsani A, Zerbini FM. 2021. Changes to virus taxonomy and to the International Code of Virus Classification and Nomenclature ratified by the International Committee on Taxonomy of Viruses (2021). Arch Virol. 166: 2633-2648. doi: 10.1007/s00705-021-05156-1.

Wanitchakorn R, Hafner GJ, Harding RM, Dale JL. 2000. Functional analysis of proteins encoded by banana bunchy top virus DNA-4 to -6. J Gen Virol. 81: 299–306. doi: 10.1099/0022-1317-81-1-299.

Wanitchakorn R, Harding RM, Dale JL. 1997. Banana bunchy top virus DNA-3 encodes the viral coat protein. Arch Virol.;142(8):1673-80. doi: 10.1007/s007050050188.

Watanabe S, Greenwell AM, Bressan A. 2013. Localization, concentration, and transmission efficiency of banana bunchy top virus in four asexual lineages of Pentalonia aphids. Viruses 5: 758–776. doi: 10.3390/v5020758.

Wu RY. 1994. Two Circular Single-stranded DNAs Associated with Banana Bunchy Top Virus. Journal of Phytopathology 142: 292-300. https://doi.org/10.1111/j.1439-0434.1994.tb04542.x

Wu RY, Su HJ. 1990. Purification and Characterization of Banana Bunchy Top Virus. J Phytopathol. 128: 153–160. doi: 10.1111/j.1439-0434.1990.tb04261.x.

Wu RY, You LR, Soong TS. 1994. Nucleotide sequences of two circular single-stranded DNAs associated with banana bunchy top virus. Phytopathology 84: 952–958. DOI: 10.1094/Phyto-84-952.

Wu PJ, Zhou XP. 2005. Interaction between a nanovirus-like component and the Tobacco curly shoot virus/satellite complex. Acta Biochim Biophys Sin (Shanghai). 37: 25-31. doi: 10.1093/abbs/37.1.25.

Yang X, Wang Y, Guo W, Xie Y, Xie Q, Fan L, Zhou X. 2011. Characterization of small interfering RNAs derived from the geminivirus/betasatellite complex using deep sequencing. PLoS One 6: e16928. doi: 10.1371/journal.pone.0016928.

Yeh HH, Su HJ, Chao YC. 1994. Genome characterization and identification of viral-associated dsDNA component of banana bunchy top virus. Virology 198: 645-52. doi: 10.1006/viro.1994.1076.

Yu NT, Xie HM, Zhang YL, Wang JH, Xiong Z, Liu ZX. 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: 573. doi: 10.1186/s12864-019-5901-0.

Yu NT, Zhang YL, Feng TC, Wang JH, Kulye M, Yang WJ, Lin ZS, Xiong Z, Liu ZX. 2012. Cloning and sequence analysis of two banana bunchy top virus genomes in Hainan. Virus Genes 44: 488-94. doi: 10.1007/s11262-012-0718-9.

Zhu H, Zhang Y, Tang R, Qu H, Duan X, Jiang Y. 2019. Banana sRNAome and degradome identify microRNAs functioning in differential responses to temperature stress. BMC Genomics 20: 33. doi:10.1186/s12864-018-5395-1

Zerbino DR, Birney E. 2008. Velvet: algorithms for de novo short read assembly using de Bruijn graphs. Genome Res. 18: 821-829. doi: 10.1101/gr.074492.107.

Zvereva AS, Pooggin MM. 2012. Silencing and innate immunity in plant defense against viral and non-viral pathogens. Viruses 4: 2578–2597. doi: 10.3390/v4112578.

Chapter 2: Study of the prevalence and diversity of alphasatellites associated with banana bunchy top virus

Article 2 (draft manuscript):

A new genus of alphasatellites associated with banana bunchy top virus in South-East Asia

Valentin Guyot[1], Marie-Line Iskra-Caruana[2], Mikhail M. Pooggin[1,*]

[1] PHIM Plant Health Institute, University of Montpellier, INRAE, CIRAD, IRD, Institute Agro, 34398 Montpellier, France

[2] CIRAD, DGD-RS, 34398 Montpellier, France

[*] Correspondence: mikhail.pooggin@inrae.fr

Key words: alphasatellite, banana bunchy top virus, multipartite ssDNA virus, replication, RNAi, siRNA

Abstract

Self-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 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 South-East Asia (SEA) for wild and cultivated banana plants infected with banana bunchy top virus (BBTV, *Nanoviridae*), followed by Illumina sequencing reconstruction of their virome components revealed, in addition to six-component BBTV genome, one to three distinct alphasatellite components present in 16 of 24 BBTV-infected plants. Comparative nucleotide and Rep protein sequence analyses classified these alphasatellites into four species: two known species previously identified in SEA, which belong to the genus *Muscarsatellite* (subfamily *Petromoalphasatellitinae*), and two novel species which belong to the genus *Banaphisatellite* (*Nanoalphasatellitinae*) so far containing a single known species recently identified in Africa. The banaphisatellites were found to be most closely related to members of the genus *Fabenesatellite* of *Nanoalphasatellitinae* and the genus *Gosmusatellite* of *Geminialphasatellitinae*, both infecting dicots. This suggests a dicot origin of banaphisatellites that got independently associated with monocot-infecting BBTV strains in Africa and SEA. Analysis of conserved sequence motifs in the common regions driving replication and gene expression of alphasatellites and helper BBTV strains revealed both similarities and differences. 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 alphasatellite- and BBTV-derived small interfering RNAs. Our findings shed new light on the provenance of alphasatellites, their co-evolution with helper viruses and potential mutual benefits of their association.

Introduction

Banana bunchy top virus (BBTV, genus *Babuvirus*, family *Nanoviridae*) has a six-component circular single stranded (ss)DNA genome and infects monocots mainly from the families Musaceae and Zingiberaceae (Qazi, 2016). Banana bunchy top disease (BBTD) causes tremendous damage to bananas and plantains, the biggest dessert and cooking banana cultures in the world (Dale, 1987; Lowe *et al.*, 2020). BBTD is transmitted by the aphid vector *Pentalonia nigronervosa* in a circulative non-propagative manner (Magee, 1940; Watanabe *et al.*, 2013). This disease is present in at least 36 different countries in all continents except Americas (Kumar *et al.*, 2011). Sequenced isolates of BBTV are classified into two phylogenetic groups with distinct geography: the South East Asia (SEA) group (Japan, China, Vietnam, Philippines, Indonesia, Thailand and Taiwan) and the Pacific and India Oceans (PIO) group (Hawaii, Tonga, Samoa, Fiji, Australia, Egypt, Cameroon, Gabon, Republic of Congo, Democratic Republic of Congo, Rwanda, Burundi, Malawi, India, Sri Lanka, Pakistan and Myanmar) (Stainton *et al.*, 2015). The BBTV genome is composed of DNA-C coding for cell-cycle protein (Clink) (Lageix *et al.*, 2007), DNA-M coding for movement protein (MP) (Wanitchakorn *et al.*, 2000), DNA-N coding for nuclear shuttle protein (NSP) (Ji *et al.*, 2019) implicated in aphid transmission (Guyot *et al.*, 2022/Chapter 1 of this thesis), DNA-R coding for master replication protein (M-Rep) that mediates rolling-circle replication of all six BBTV components (Horser *et al.*, 2001), DNA-S coding for coat protein (CP) encapsidating each viral ssDNA component individually (Wanitchakorn *et al.*, 1997) and DNA-U3 coding for a protein of unknown function (Beetham *et al.*, 1999). BBTV components have a size of ca. 1 to 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 *et al.*, 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.*, 1997b). 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 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 *et al.*, 1997a).

BBTV, like other nanovirids and related plant circular ssDNA viruses of the families *Geminiviridae* and *Metaxyviridae* are frequently associated with circular ssDNA alphasatellites (family *Alphasatellitidae*) that are non-essential for viral replication and infection cycles (Briddon & Stanley, 2006; Briddon *et al.*, 2018). Alphasatellites are classified into three subfamilies: *Geminialphasatellitinae* (associated with helper viruses of the family *Geminiviridae*), *Nanoalphasatellitinae* (associated with helper viruses of the genus *Nanovirus* of family *Nanoviridae*) and *Petromoalphasatellitinae* (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. This Rep enables self-replication of alphasatellite DNA but cannot mediate trans-replication of helper virus genome components (Horser *et al.*, 2001). Alphasatellites depend on their helper viruses for encapsidation, movement and transmission by insect vectors. Impact 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 *Subclovsatellite*, *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 loads in aphids (Mansourpour *et al.*, 2022). In the case of BBTV, a newly emerging alphasatellite associated with BBTV isolates from 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 virus-derived 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 Rep- and movement protein-encoding components. Notably, the plant RNA interference (RNAi)-based antiviral defenses was found to silence DRC alphasatellite gene expression at both transcriptional and posttranscriptional levels, generating highly-abundant 21, 22 and 24 nucleotide (nt) siRNAs, suggesting that alphasatellite may serve as a decoy protecting its helper virus from antiviral RNAi (Guyot *et al.*, 2022). Importantly, DRC alphasatellite is the first satellite reported to be associated with BBTV isolates from the PIO phylogenetic group and represents a new species provisionally named *Banaphisatellite musae1* in a new genus *Banaphisatellite*. It is more related to alphasatellites of the subfamily *Nanoalphasatellitinae* infecting dicots rather than to other BBTV alphasatellites of the subfamily

Petromoalphasatellitinae infecting monocots and previously identified only in South-East Asia (Guyot *et al.*, 2022). More specifically, the latter alphasatellites were discovered first in Taiwan in 1994 and later in China and Vietnam, and were classified into two distinct genera of *Petromoalphasatellitinae* – *Babusatellite* and *Muscarsatellite* (Supplementary Table S1).

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* of subfamily *Nanoalphasatellitinae*. Besides comparative analysis of complete nucleotide sequences of the newly discovered and previously-reported alphasatellites and their helper viruses, we performed clustering analysis of alphasatellite Rep proteins, which reveals 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 get independently associated with monocot-infecting BBTV strains 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 their helper BBTV from PIO and SEA. Finally, we compared the relative abundances of alphasatellites and their helper BBTV components, and the profiles of alphasatellite- and BBTV-derived siRNAs, which reveals the impact of alphasatellites on helper virus replication and evasion of RNAi-based antiviral defenses.

Results and Discussion

Prevalence and diversity of BBTV alphasatellites in Vietnam, Laos and China

Leaf samples of wild and cultivated *Musa* plants were collected in Northern Vietnam (n = 11), Laos (n = 10), and the south of Yunnan province in China (n = 5) during surveys conducted in 2018 and 2019 (Supplementary Figure S1 and Supplementary Table S2). 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

resulting 125 nt paired-end 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 genomes: these viral reads likely represent cross-contamination from the other samples multiplexed and sequenced in one flowcell of Illumina HiSeq2500. These two virus-free samples served as negative controls to establish the cross-contamination threshold. In addition, BLASTn analysis revealed that 16 of the 24 BBTV-infected samples contain a total of 28 complete circular genomes of alphasatellites (1 to 3 per sample) (Supplementary Table S2). Notably, alphasatellites were present in all 11 samples from Vietnam, 2 of 9 samples from Laos and 3 of 4 samples from China, indicating their high prevalence in SEA. Moreover, among the samples coinfected with BBTV and alphasatellites, two samples contained complete circular genomes of two distinct badnaviruses (genus *Badnavirus*, family *Caulimoviridae*) and three samples contained circular molecules representing defective viral DNA components lacking protein-coding capacity (Supplementary Table S2). 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 3 alphasatellites and from BBTV DNA-R. In a follow up analysis of alphasatellites and their helper viruses from SEA, we also included Illumina-sequenced BBTV isolates from PIO described in our previous study, which originated from DRC (n = 5), Gabon (n = 3), Benin (n = 1), Malawi (n = 1) and New Caledonia (n = 1) and which included two samples (ALYU-21 and JGF-1) from DRC with two distinct isolates of *Banaphisatellite musae1* (Guyot *et al.*, 2022; Supplementary Tables S1 and S2).

Pairwise comparison and phylogenetic analysis of full-length nucleotide sequences of the new alphasatellites from Vietnam, Laos and China and previously-reported alphasatellites (Varsani *et al.*, 2021) revealed that, based on the current species and genus demarcation criteria (81 % and 68 % sequence identity, respectively), the new satellites represent four different species, including two known species of BBTV alphasatellites - banana bunchy top alphasatellite 2 (BBTA2) and BBTA3 - both classified into the genus *Muscarsatellite* of subfamily *Petromoalphasatellitinae*, and two novel species that we classified together with *Banaphisatellite musae1* (also named BBTA4) into the tentative genus *Banaphisatellite* of subfamily *Nanoalphasatellitinae* (Figure 1 and Supplementary Table S2). We propose to name

these new species *Banaphisatellite musae2* and *Banaphisatellite musae3*. Keeping in line with old nomenclature and numbering of BBTV alphasatellites, the new alphasatellites will hereafter be named banana bunchy top alphasatellite 5 (BBTA5) and banana bunchy top alphasatellite 6 (BBTA6), respectively. BBTA5 is represented with 8 isolates and is prevalent in Vietnam (6 of the 11 alphasatellite-containing samples) and Laos (each of the 2 alphasatellite-containing samples) but absent in China. BBTA6 is represented with 5 isolates and is present in Vietnam (3 of the 11 alphasatellite-containing isolates) and China (2 of the 4 alphasatellite-containing isolates) but absent in Laos. The new isolates of previously-identified alphasatellites of the genus *Muscarsatellite* - BBTA2 and BBTA3 – are prevalent in Vietnam (nine BBTA2 and two BBTA3 isolates) and China (three BBTA2 and one BBTA3) but absent in Laos and are frequently found in coinfections with banaphisatellites (8 samples in Vietnam and 2 samples in China) (Supplementary Table S2). Among 17 alphasatellite-containing samples, 5 samples contain only banaphisatellites (ALYU-21, 35, 40, 42, 43), 3 samples only muscarsatellites (ALYU-36, 37, 56) and 9 samples contain two or more alphasatellites with at least one member from both *Muscarsatellite* and *Banaphisatellite* genera (ALYU-25, 26, 29, 32, 33, 34, 39, 54, 55). In both Vietnam and China, mixed infections of alphasatellites are more frequent than infections with single alphasatellites (Supplementary Table S2).

Previously, alphasatellites representing the babusatellite BBTA1 and the muscarsatellite BBTA3 species were reported only from in Haikou (Hainan province, China) and Taiwan, while those representing the muscarsatellite BBTA2 were reported from Haikou and Vietnam (Supplementary Table S1). In our survey, new isolates of BBTA2 and BBTA3 were detected in the North of Vietnam and the South of Yunnan province of China, which are in close geographic vicinity of each other (Supplementary Figure S1 and Supplementary Table S2). However, our sampling in the Central and Northern Laos as well as in China close to the border with Northern Laos did not identify any known BBTV alphasatellites and only 2 samples from Central Laos (both very close to the border with Thailand) contained BBTA5 (Supplementary Figure S1 and Supplementary Table S2). This indicates that BBTV alphasatellites are less prevalent in Laos, likely due to less intensive trade and exchange of banana propagules with its neighbour countries.

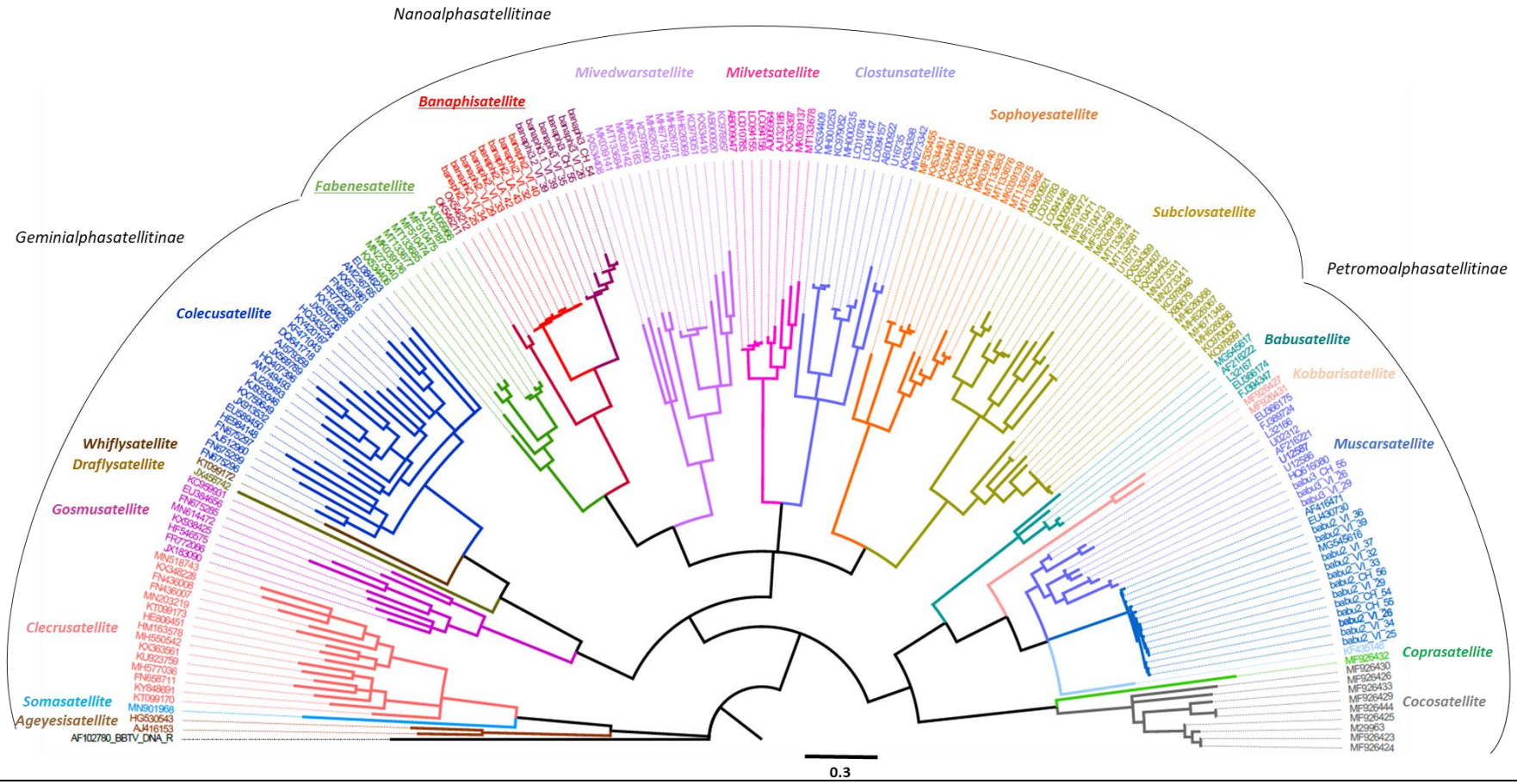


Figure 1. Phylogenetic analysis of complete nucleotide sequences of alphasatellites (family *Alphasatellitidae*)

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 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 Figure 1 of that paper) and (Varsani *et al.*, 2021). In comparison with the phylogenetic tree shown in Guyot et al. (2022), bootstrap values < 60 were collapsed.

Using Sequence Demarcation Tool (SDT) v1.2, intraspecies pairwise sequence identity of isolates of the banaphisatellite BBTA5 and BBTA6 species was found to be > 98 % and 93-98 %, respectively, and of the muscarsatellite BBTA2 and BBTA3 species > 96 % and 77-98 %, respectively (Supplementary Figure S2). Thus, more prevalent BBTV alphasatellites - the muscarsatellite BBTA2 and the banaphisatellite BBTA5 – appear to have higher genetic stability than less prevalent ones, i.e. the banaphisatellite BBTA6 and the muscarsatellite BBTA3. Likewise, the babusatellite BBTA1, which was not detected in our surveys, has low genetic stability among its previously identified isolates (86-98 %; Supplementary Figure S2).

Considering sampling chronology of new isolates of the previously identified alphasatellites, sequence analysis of 12 new isolates of BBTA2 shows high genetically stability of this species over the last 17 years since the previous reports of its isolates in 2002 and 2008 in Vietnam (AF416471 and EU430730) and in 2013 in China (Haikou, Hainan) (MG545616). In contrast, our new isolates of BBTA3 that was previously reported only from Taiwan and China (Haikou, Hainan) are strongly divergent: our isolates ALYU26 and ALYU29 from Vietnam share < 78 % pairwise identity with the isolates L32166 and U02312 (1994, Taiwan) and ca. 80-81 % identity with the isolates EU366175 (2007, Taiwan), FJ389724 (2009, Taiwan) and HQ616080 (2012, Haikou, Hainan, China), unlike our isolate ALYU55 from the Yunnan province of China sharing 83-97 % identity with the previous isolates from China and Taiwan (Supplementary Figure S2). The BBTA3 isolates ALYU26 and ALYU29 should therefore be classified into a sub-species or strain of BBTA3.

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*, *Petromoalphasatellitinae* and *Geminialphasatellitinae*. 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).

The results of clustering analysis of Rep proteins confirmed the nucleotide sequence-based classification of banaphisatellites in three species and their close evolutionary links with fabenesatellites (genus *Fabenesatellite*, subfamily *Nanoalphasatellitinae*) (Figure 2 vs Figure 1). Indeed, at the most stringent threshold (P ≤ 1E-81), isolates of each of the three species - BBTA4 (n = 2), BBTA5 (n = 8) and BBTA6 (n = 5) – are linked to each other within each species but not between species (Figure 2A; Supplementary Figure S3A), while at a less stringent threshold (P ≤ 1E-72) all isolates of all three species become interconnected within the genus *Banaphisatellite* (Figure 2B; Supplementary Figure S3B). Notably at the latter threshold, two isolates of BBTA4 (DRC alphasatellite) become connected to some but not all fabenesatellites classified in a single species Faba bean necrotic yellows alphasatellite 2 (FBNYA2), the only species in the genus *Fabenesatellite*, whereas at a lower stringency (P ≤ 1E-63) all alphasatellites of the *Fabenesatellite* and *Banaphisatellite* genera become interconnected, with none of them being linked to other genera (Figure 2C; Supplementary Figure S3C). Finally, further decrease in stringency (P ≤ 1E-52) resulted in appearance of evolutionary links of both banaphisatellites and fabenesatellites to four of the 7 alphasatellites from the genus *Gosmusatellite* of subfamily *Geminialphasatellitinae* (Figure 2D; Supplementary Figure S3D). Notably, the gosmusatellites being interconnected in one cluster at more stringent thresholds, at the less stringent threshold (P ≤ 1E-52) become connected to other clusters including the cluster of *Banaphisatellite* and *Fabenesatellite* genera of *Nanoalphasatellitinae* and the cluster of *Colecusatellite*, *Draflysatellite* and *Whiflysatellite* genera of *Geminialphasatellitinae* (Figure 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). Other 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) and Vernonia yellow vein Fujian alphasatellite (KC959931, JF733780, KF471041) – have a wider dicot host range comprising Malvaceae (MN614472, EU384656), Asteraceae (KC959931), Urticaceae (LN880828), Caricaceae (KF471041) and Fabaceae (KC959931) families.

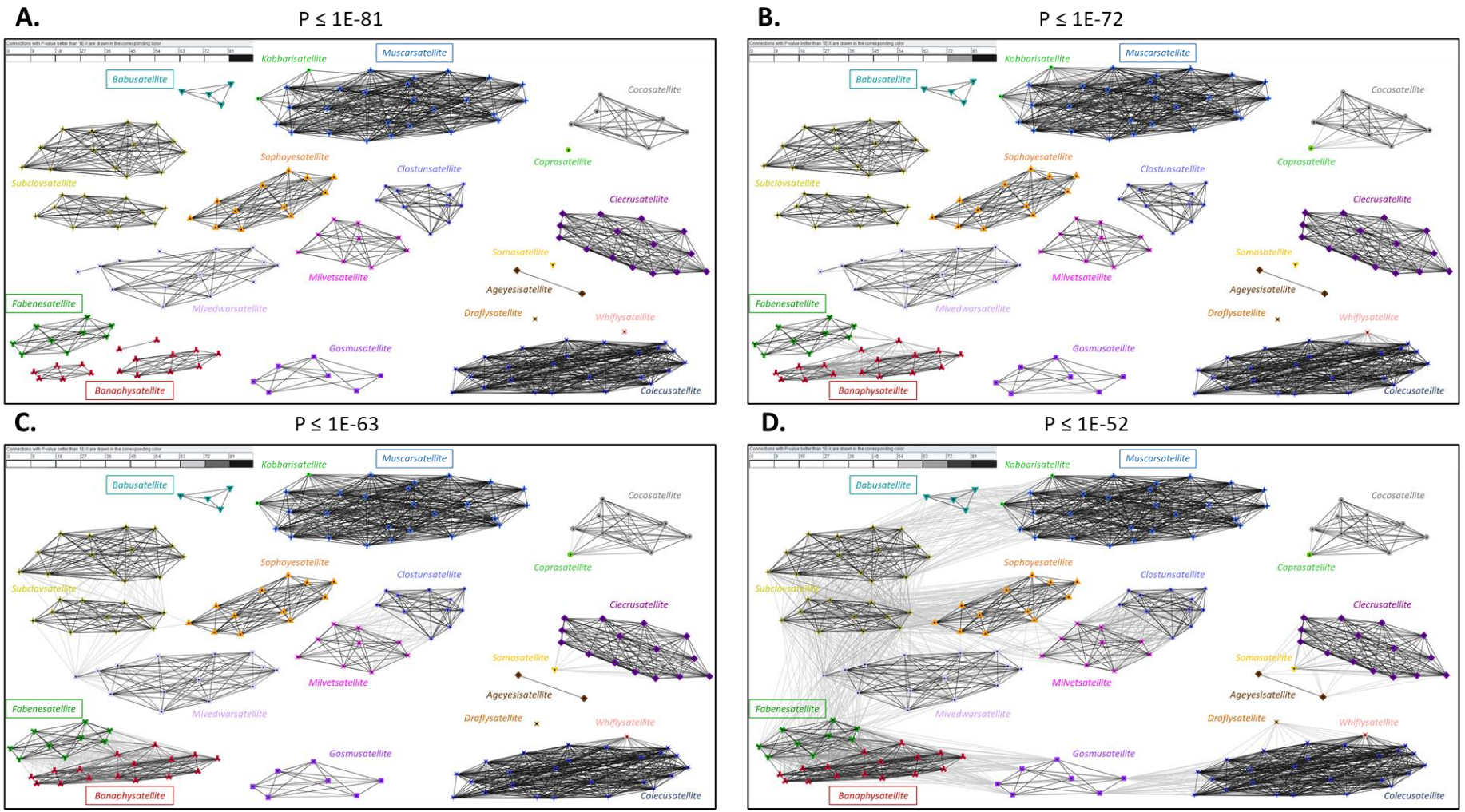


Figure 2. Sequence similarity-based clustering analysis of Rep proteins encoded by alphasatellites (family *Alphasatellitidae*).

Rep proteins of the alphasatellites classified by Briddon 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 color 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 Fruchterman-Reingold force-directed layout algorithm): P ≤ 1E-81 (strongest link/black) **(A)**, P ≤ 1E-72 (weaker link/dark grey) **(B)**, P ≤ 1E-63 (weaker link/grey) **(C)**, P ≤ 1E-52 (weakest link/light grey) **(D)**.

Note 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 ($P \leq 1E\text{-}52$), fabenesatellites (but not banaphisatellites) become connected to other genera of *Nanoalphasatellitinae*, namely *Midvesatellite* and *Subclovsatellite* (Figure 2D). The genus *Subclovsatellite* is composed of two distinct clusters, likely representing two strains or subspecies, with one cluster being linked to the *Fabenesatellite*, *Midvesatellite* and *Sophoyesatellite* genera of *Nanoalphasatellitinae* and another being linked to the *Babusatellite* and *Muscarsatellite* genera of *Petromoalphasatellitinae* (Figure 2D) that contain all previously-identified alphasatellites associated with BBTV SEA isolates representing BBTA1, BBTA2 and BBTA3 species. Thus, Rep proteins of the babusatellite BBTA1 and the muscarsatellites BBTA2 and BBTA3 are only distantly related to Rep proteins of the banaphisatellites BBTA4, BBTA5 and BBTA6 that have much closer evolutionary links to Rep proteins of dicot-infecting fabenesatellites and gosmusatellites. This implies the provenance of banaphisatellites from a dicot-infecting ancestor associated with a helper virus from the genus *Nanovirus* of family *Nanoviridae*, which contains helper viruses of fabenesatellites, or from the genus *Begomovirus* of family *Geminiviridae*, which contains 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 using SDTv1.2 showed their clear differences in each component of BBTV genome, with the highest pairwise identity between the groups being in DNA- R (89-93 %), followed by DNA-S (87-91 %), DNA-N (86-90 %), DNA-C (85-89 %), DNA-M (84-88 %) and DNA-U3 (79-83 %) (Figure 3). This confirms previous findings establishing the two phylogenetic groups with a distinct geographical delineation (Stainton *et al.*, 2015). In addition, within each group we found distinct subgroups, most evident for DNA-M and DNA-N of SEA isolates, which are split in two subgroups in each case with respectively 94-95 % and 92-93 % between-subgroup identities and > 98 % and > 99 % within-subgroup identities. These subgroups are not shared between components, although a DNA-N subgroup represented with 5 isolates from the Northern Laos (ALUY-47-51) is part of a larger DNA-M subgroup additionally represented with all 4 isolates from China (ALYU-53-56) and 6 of 9 isolates from Vietnam (ALYU-32, 34,

35, 37, 39, 40) (Figure 3). We found no link between those subgroups and the presence or absence of alphasatellites or any particular alphasatellite species. In the case of PIO isolates, a notable example is the sample ALYU-32 from Gabon, which lacks DNA-U3 and whose DNA-S shares 94-95 % pairwise identities with DNA-S of other isolates from Gabon, DRC, Benin, Malawi and New Caledonia that share > 99% identity with each other. Importantly, the only 2 isolates from DRC associated with the banaphisatellite BBTA4 (ALYU-21 and JGF-1) do not differ substantially from any other DRC isolates or other PIO isolates (except ALYU-32 from Gabon) in any component of the BBTV genome (Figure 3).

Our analysis of BBTV protein-coding sequences revealed that, unlike the components encoding M-Rep (DNA-R), CP (DNA-S), Clink (DNA-C), MP (DNA-M) and NSP (DNA-N) proteins - all conserved in both PIO and SEA isolates, DNA-U3 has no ORF common between PIO and SEA. Our PIO isolates of DNA-U3 share one conserved ORF encoding a small (77 amino acid) protein of unknown function, which has previously been recognized in other PIO isolates, although its length was not identical between isolates (Guyot *et al.*, 2022). Our SEA isolates of DNA-U3 share a different conserved ORF encoding a smaller (39 amino acid) protein of unknown function, which is also conserved in some isolates previously reported from China and Thailand, and two other ORFs (encoding 40 and 55 amino acid proteins) which are, however, interrupted by premature stop codons or frameshifting mutations in some isolates (Supplementary sequence analysis dataset to be prepared). Taken together, DNA-U3 components of SEA and PIO isolates do not possess any conserved protein-coding capacity, suggesting either distinct functions in the PIO and SEA BBTV strains, or a non-coding function of this component of BBTV.

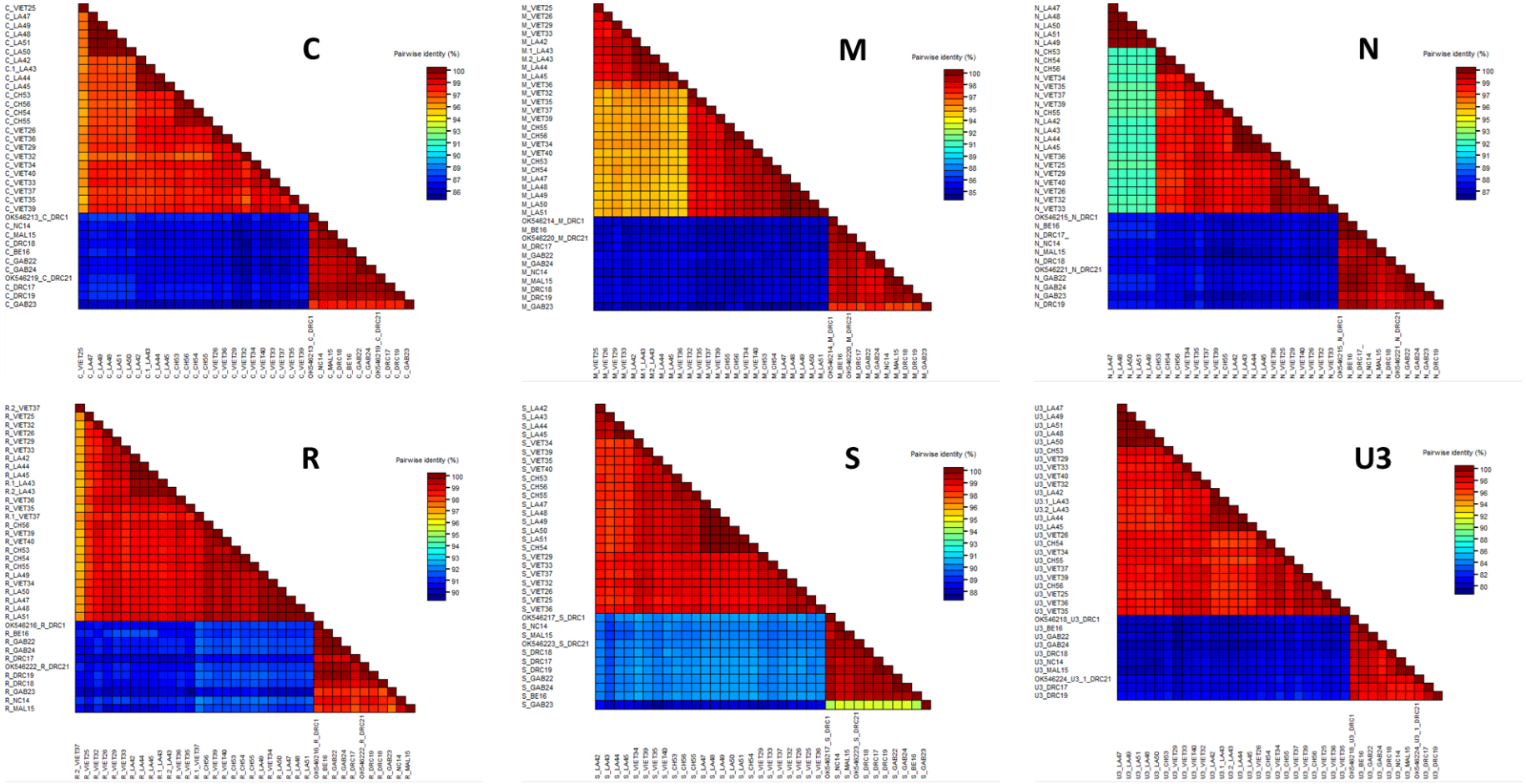


Figure 3. Pairwise sequence comparison of BBTV genome components from SEA and PIO.

The nucleotide sequences of BBTV DNA-C, DNA-M, DNA-N, DNA-R, DNA-S, DNA-U3 were compared using Sequence Demarcation Tool (SDT) v1.2 with Muscles (Muhire et al. 2014) and their pairwise identities (in %) were plotted as heatmap diagrams. Note that U3 component was not detected in the BBTV-infected sample ALYU-23 from Gabon.

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

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

Alignment of the CR-M sequences of BBTV components of our SEA and PIO isolates revealed that they range in size from 66 to 92 nts with the exception for ALYU-23 from Gabon and ALYU-35 from Vietnam, whose DNA-N components have longer CR-M with an identical 17 nt insert at their 5'-ends (a direct repeat of the downstream sequence) (Supplementary Figure S4), and that they all contain a highly conserved GC-rich element at their 3'-ends that can potentially form a small stem-loop structure - AGGGCCGHAGGCCCGT (an inverted repeat is underlined; H = C, A, or T) (Figure 4 and Supplementary Figure S4). Sequences upstream of this universally-conserved element fall in two distinct clades - SEA and PIO – and share high identities and lengths within these clades with the exception for DNA-R components that have a deletion of 10 and 26 nts, respectively, at the 5'-end (Supplementary Figure S4). Previously, the deletion of 26 nts in DNA-R CR-M has been reported for a PIO isolate (Burns *et al.*, 1995; Hafner *et al.*, 1997a). Differences in CR-M sequences of PIO and SEA isolates have been reported, although not specified, by Karan et al. (1994).

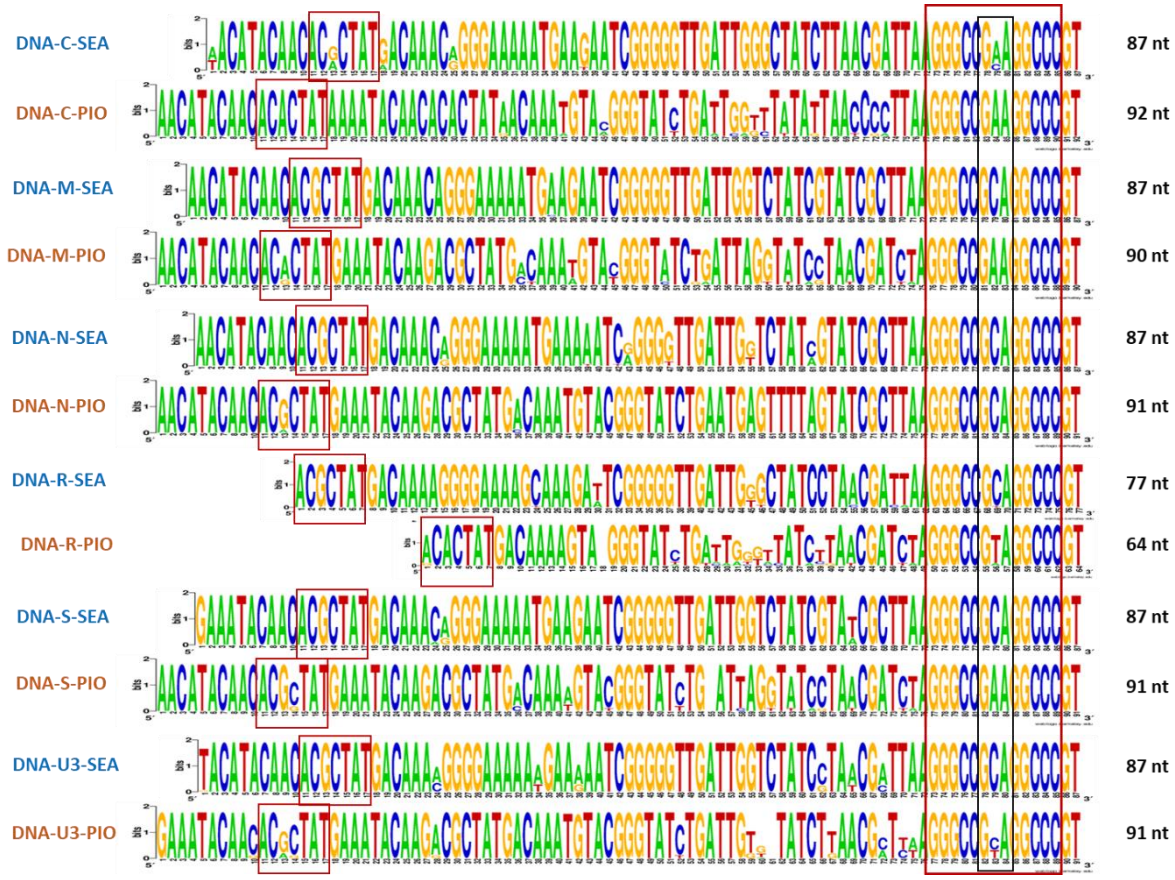


Figure 4. Logo of the common-region major (CR-M) sequences of the SEA and PIO isolates of BBTV.

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.

We next inspected if BBTV alphasatellites possess any region homologous to the CR-M of their helper BBTV and located upstream of the CR-SL. As a result, CR-M sequences with a characteristic GC-rich stem-loop region (GGGCCGHAGGCCC) at the 3´-end were identified in all BBTV alphasatellite species except BBTA4 (Figure 5). This suggests that the banaphisatellite BBTA4 recently emerged in Africa (Guyot *et al.*, 2022) is distinct from other two species of banaphisatellites (BBTA5 and BBTA6) that are prevalent in SEA and raises a question on a mechanism of priming of complementary DNA synthesis in this particular case. Interestingly, the A and GT nucleotides flanking the stem-loop structure in all BBTV components (AGGGCCGHAGGCCCGT) were also conserved in some but not all the five alphasatellite species. Moreover, the sequences upstream of the stem-loop region, which are conserved in the BBTV components of SEA isolates were highly similar to those of all isolates of the muscarsatellites BBTA2 and BBTA3 and the banaphisatellite BBTA5 as well as four of the 6 isolates of the banaphisatellites BBTA6 (Figure 5). In contrast, other two isolates of BBTA6 and all five isolates of the babusatellite BBTA1 (identified only in previous surveys) had deletions or alterations in the conserved sequence TCGGGGGTTGATTG and further upstream purine-rich sequences, except for the motif ACRCTAT (R = G or A) preserved at the 5´-end of CR-M of all alphasatellites (Figure 5). The same motif is present at the 5´-end of CR-M of all six BBTV components of both SEA and PIO isolates (Figure 4 and Supplementary Figure S4), suggesting its involvement in priming the complementary DNA synthesis.

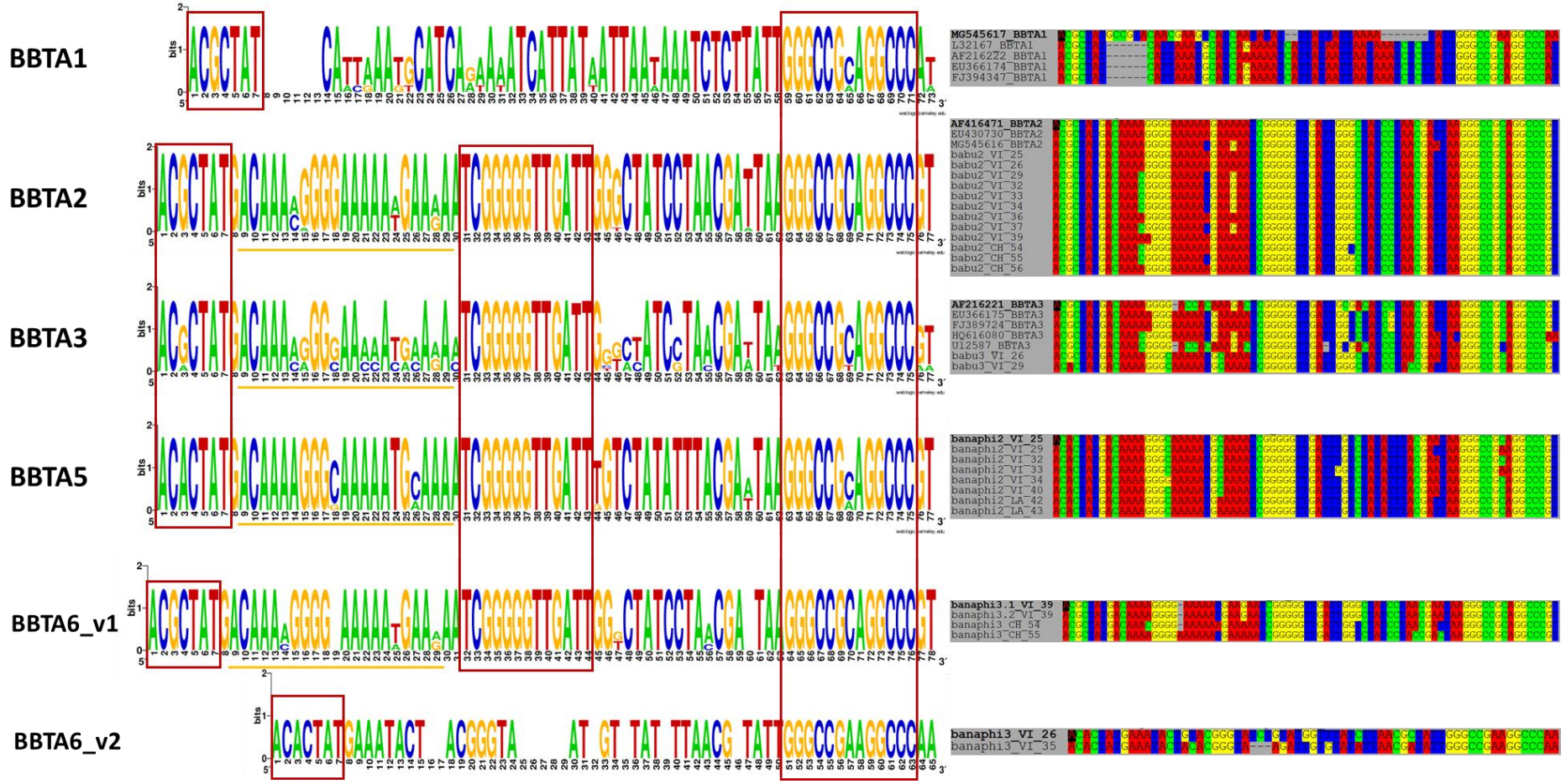


Figure 5. Logo and alignment of the common-region major (CR-M) sequences of SEA and PIO isolates of BBTV alphasatellites.

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. The alignments are shown on the right side from each logo.

Alignment of the CR-SL sequences of BBTV components of our PIO and SEA isolates revealed previously-described cis-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 iterons – the 2 adjacent iterons 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 (Figure 6 and Supplementary Figure S6A). The distance between the iteron R and the stem-loop structure base is conserved (26 nts) with the exception for DNA-N of both SEA and PIO isolates (9 nts) and DNA-U3 of SEA isolates (90 nts; note that DNA-U3 of PIO isolates has the second iteron GTCCC at the same distance of 90 nts; Supplementary Figure S6B).

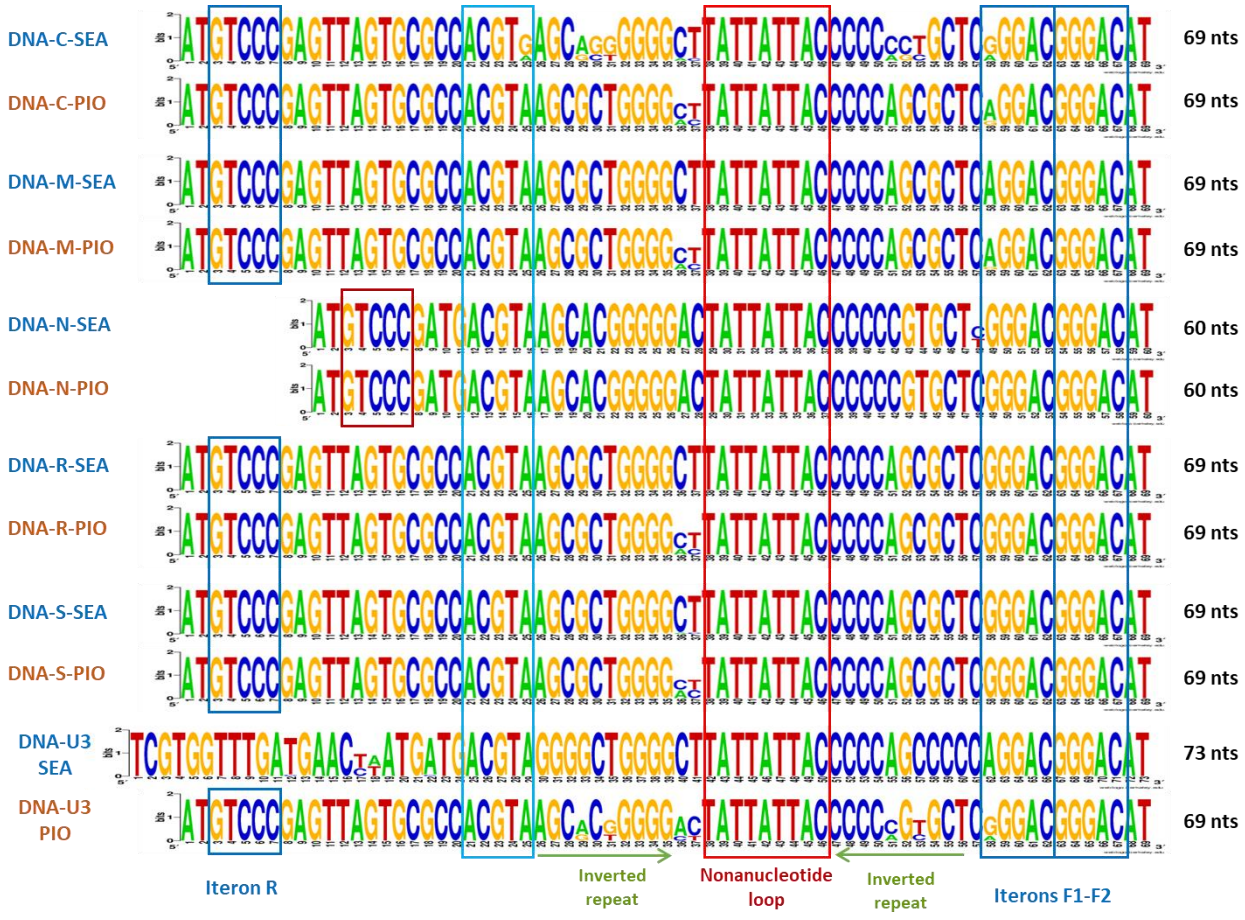


Figure 6. Logo of the common-region stem-loop (CR-SL) sequences of the SEA and PIO isolates of BBTV.

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 and named.

Thus, our PIO and SEA isolates share all the previously-described conserved cis-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. 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 its 3'-terminal nucleotide (ACTGG) in all isolates from Vietnam and China and 4 of 9 isolates from Laos (Figure 6 and Supplementary Figure S6A).

Inspection of the CR-SL sequences of BBTV alphasatellites revealed (i) the invariant nonanucleotide TAGTATTAC that is preserved in most alphasatellites (Stainton *et al.*, 2017) and that differs at one position from the helper virus nonanucleotide (TATTATTAC), (ii) the inverted repeats flanking the nonanucleotide sequence, (iii) three putative 5 nt iterons R, F1 and F2 and (iv) the TATA-box of Pol II promoter driving transcription of Rep mRNA (Figure 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) and, in both cases, comprise the iterons R and F1 sequences (Figure 7). This is unlike the helper BBTV having all the three iterons outside the stem-loop structure (Figure 6). Note, however, that two of the 3 previously reported isolates (but not our 12 isolates) of BBTA2 have an insertion of G in the iteron F1 (TCCGGC) (Supplementary Figure S7) and in these 2 isolates we identified alternative iterons F1' (GCGCA), F2' (GCGCA) and R' (TGCGC) (Figure 7). Another notable exception for the within-species iteron identity is BBTA1 for which one of the 5 previously-reported isolates has a different version of both the iteron R (AGGCA vs AGGAG) and iterons F1 and F2 (TGCCT vs CTCCT) (Supplementary Figure S7). The alterations of 2 nucleotides in each of the three iterons (underlined) do not affect their nature as inverted (R) and direct (F1 an F2) repeats, thus supporting their functionality in alphasatellite replication. Other species of alphasatellites have invariant iterons in all isolates within each species (Figure 7 and Supplementary Figure 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 the 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 the iterons of the babusatellite BBTA1.

Another notable difference between the CR-SL sequences of BBTV alphasatellites and their helper BBTV is that the iterons F1 and F2 of BBTV alphasatellites are not adjacent to each other and are separated by 3 nts in all three banaphisatellite species or 8 nts in the babusatellite BBTA1 and both muscarsatellite species (Figure 7). This indicates differences in iteron recognition domains of alphasatellite 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 alphasatellite Rep transcription (as demonstrated for BBTA4 in Guyot et al. 2022) is located just upstream of the stem-loop structure in all BBTV alphasatellites, 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 alphasatellite 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 *et al.*, 1994).

Our inspection of the CR-SL sequences of alphasatellites from the genus *Fabenesatellite* 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 well conserved between banaphisatellites and fabenesatellites (Supplementary Figure S7B-C), further supporting their close evolutionary link.

Taking our findings for the CR-SL 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 an iteron sequence and an 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 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 *et al.*, 2001), and suggests that BBTV M-Rep may not be able to mediate trans-replication of any BBTV alphasatellite.

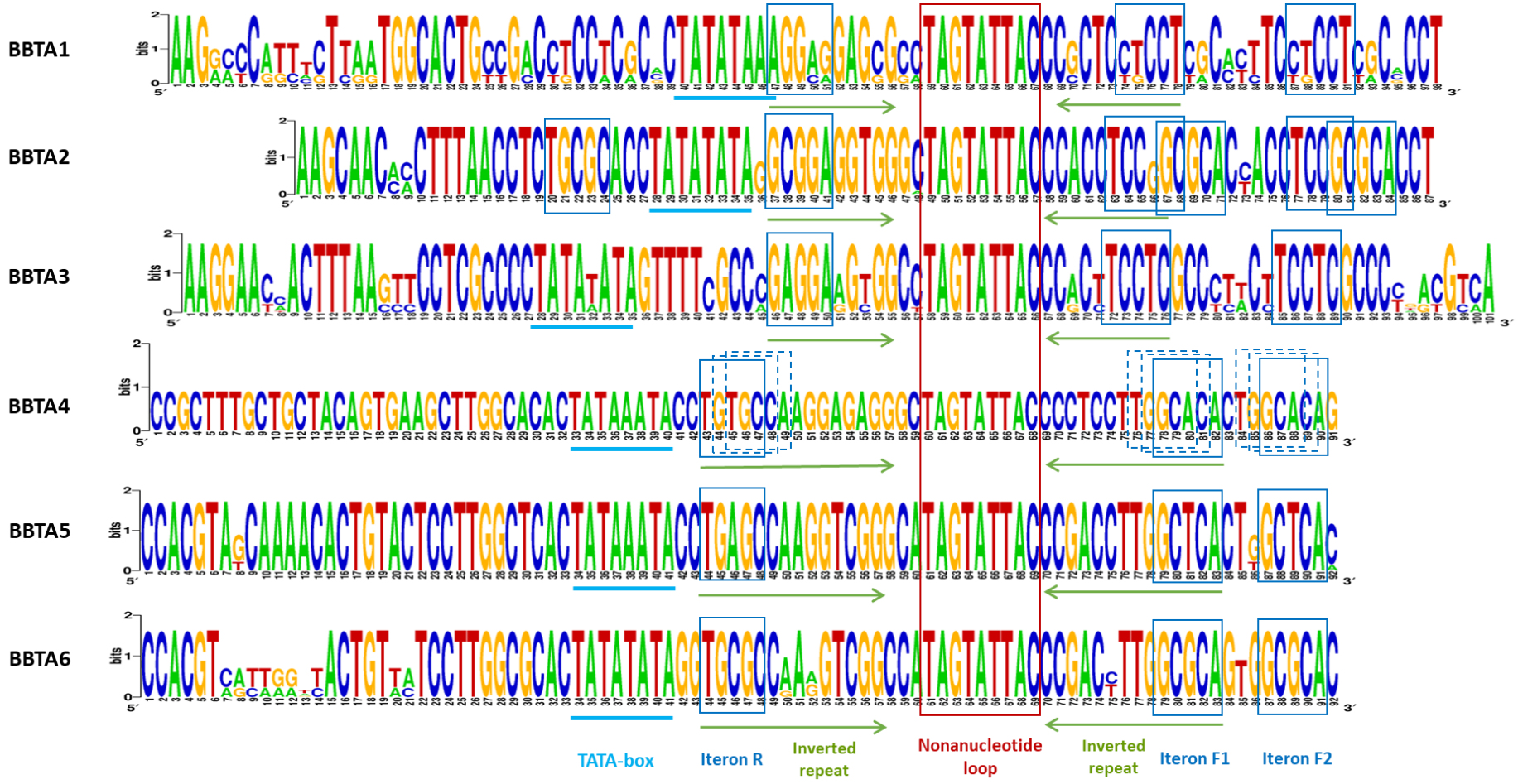


Figure 7. Logo of the common-region stem-loop (CR-SL) sequences of BBTV alphasatellites.

The CR-SL sequences of the alphasatellites 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. Their consensus sequence (logo) was generated using WebLogo v2.8.2. The conserved sequence motifs/cis-acting elements are boxes and named.

Relative abundances of alphasatellites and their helper viruses

To further investigate the interactions between alphasatellites and their helper viruses, we compared relative abundance (frequency) of six BBTV components (helper virus DNA/genome formula) and six BBTV components and alphasatellites (viral DNA formula) 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. Previously, we have established viral DNA formulas for the BBTV isolates 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 and found that both methods give similar results (Guyot *et al.*, 2022). Another study comparing the viral genome formula 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 BBTV genome formulas in PIO (n = 9) and SEA (n = 23) isolates revealed differences in relative abundance of DNA-M and DNA-C, whose median frequencies were higher in SEA isolates (Kruskal-Wallis P = 0.017 and P = 0.005, respectively), and of DNA-U3 whose median frequency was lower in SEA isolates (Kruskal-Wallis P = 0.052) (Figure 8A; Supplementary Figure S8). This is despite a big variation in the BBTV formulas within each phylogenetic group. A notable example is the SEA isolate ALYU33 that differs from all other SEA and PIO isolates in that its DNA-S is more abundant than any other BBTV genome component (Supplementary Figure S8). The observed differences in BBTV genome formula between PIO and SEA isolates can potentially be due to the presence of alphasatellites in the majority of SEA isolates. Indeed, our previous study has revealed that banaphisatellite BBTA4 accumulates at high levels, thereby reducing accumulation of the helper virus and altering its genome formula (Guyot *et al.*, 2022). We therefore compared BBTV genome formulas of the isolates without alphasatellites and the isolates with one, two or three alphasatellites. The results did not reveal any statistically significant differences of BBTV genome formulas, although the median values differ substantially (Figure 8B).

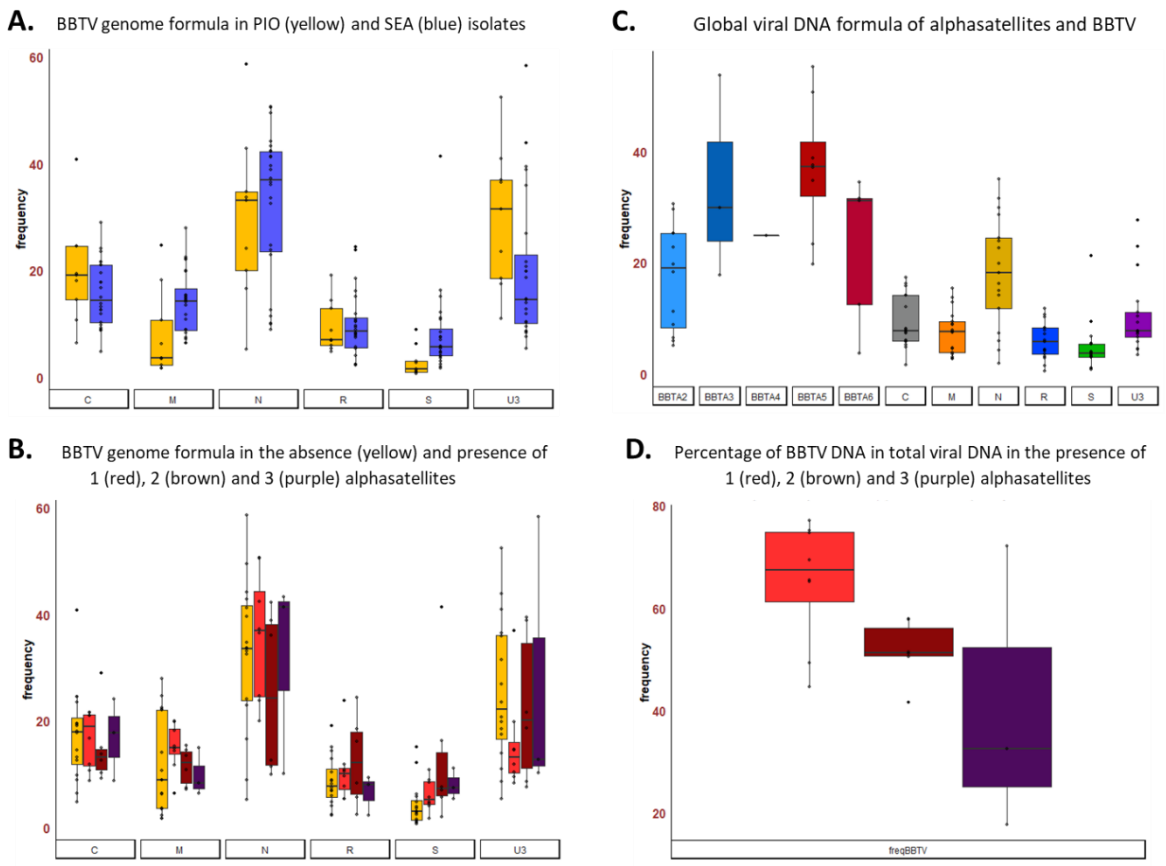


Figure 8. Relative abundance of BBTV genome components and alphasatellites.

(A) BBTV genome formula in PIO (yellow) and SEA (blue) isolates. **(B)** BBTV genome formula in the absence (yellow) and presence of one (red), two (brown) and three (purple) alphasatellites. **(C)** Global viral DNA formula 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.

Calculation of the median viral DNA formulas in the isolates with alphasatellites revealed that all five alphasatellites have a high median frequency that is comparable (BBTA2) or exceeding (BBTA3-6) the median frequency of DNA-N, the most abundant component of the helper virus in the presence of alphasatellites (Figure 8C). 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 Figure S9A-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 3 isolates (Supplementary Figure

S9D). All the banaphisatellites from PIO (BBTA4) and SEA (BBTA5 and BBTA6) accumulated at high frequency in all their single infections (n = 7), which either exceeded the frequency of any BBTV components (BBTA5 and BBTA6) or was comparable to the frequencies of DNA-N and DNA-U3 (BBTA4) (Supplementary Figure S9B). Previously, 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 Figure S9C). Likewise, in coinfections with two alphasatellites (all 3 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 Figure S9C). A previous qPCR analysis of two SEA isolates of BBTV associated with one (BBTA2) or two (BBTA1 and BBTA2) alphasatellites has revealed that in the single infection the muscarsatellite BBTA2 accumulates at the highest level, followed by DNA-N, while in the double infection the babusatellite BBTA1 accumulates at the highest level, followed by the muscarsatellite BBTA2 and DNA-U3 (Yu *et al.*, 2019).

Since in all our isolates one or more alphasatellites accumulated at high levels the percentage of the helper virus DNA in total viral DNA was found to be 45 to 77 % (median ca. 67 %) in the presence of one alphasatellite, 43 to 57 % (median ca. 50 %) in the presence of two alphasatellites and 20 to 70 % (median ca. 30 %) in the presence of three alphasatellites (Figure 8D). This is consistent our previous study in which the percentage of helper virus DNA was found to be ca. 65 % 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% in the presence of BBTA4. By extrapolation, the high proportion of alphasatellite DNA in all our PIO and SEA isolates would imply that all BBTV alphasatellites have the ability to efficiently compete with the helper virus for components of the host replication machinery and thereby reduce loads of their helper BBTV in banana plants as we proposed previously. Moreover, coinfections 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 have previously established for BBTA4 (Guyot *et al.*, 2022).

Interactions of BBTV alphasatellites and their helper virus with RNAi-based antiviral defences generating siRNAs

To investigate the plant RNAi responses to co-infections with BBTV alphasatellites and their helper viruses, we employed Illumina sequencing of small RNA populations from six leaf samples from Vietnam 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 respective virome components. The mapped viral reads representing each virome component or combinations thereof were then counted and normalized per million of total 19-25 nt 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 % (ca. 4,500 to 15,500 RPM), while the percentage of combined alphasatellite-derived sRNAs ranged from ca. 0.2 to 0.7 % (1,500 to 7,000 RPM) (Figure 9A). Notably, the ratio of BBTV- to alphasatellite-derived sRNAs was comparable in all samples: alphasatellite sRNAs constituted 25 to 31 % of the total BBTV + alphasatellite sRNAs. This is consistent with our previous findings for banaphisatellite BBTA4 whose sRNAs constituted on average 27 % 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 4 samples) and BBTA6 (present in one sample) accumulated at lower levels, comparable to those of BBTA3 (Figure 9B-C). Comparison of alphasatellite sRNA and DNA formulas revealed an inverse correlation for. the alphasatellites BBTA3, BBTA5 and BBTA6 spawning less abundant sRNAs and accumulating their DNA at higher levels, with the exception for one of the 4 samples with BBTA5 (ALYU-33) where its DNA accumulated at a level comparable with that of BBTA2 (Figure 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 (Figure 9B). Comparison of sRNA and DNA formulas 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 5 samples where it is the second best producer of viral sRNAs (Figure 10 and Supplementary Figure S10).

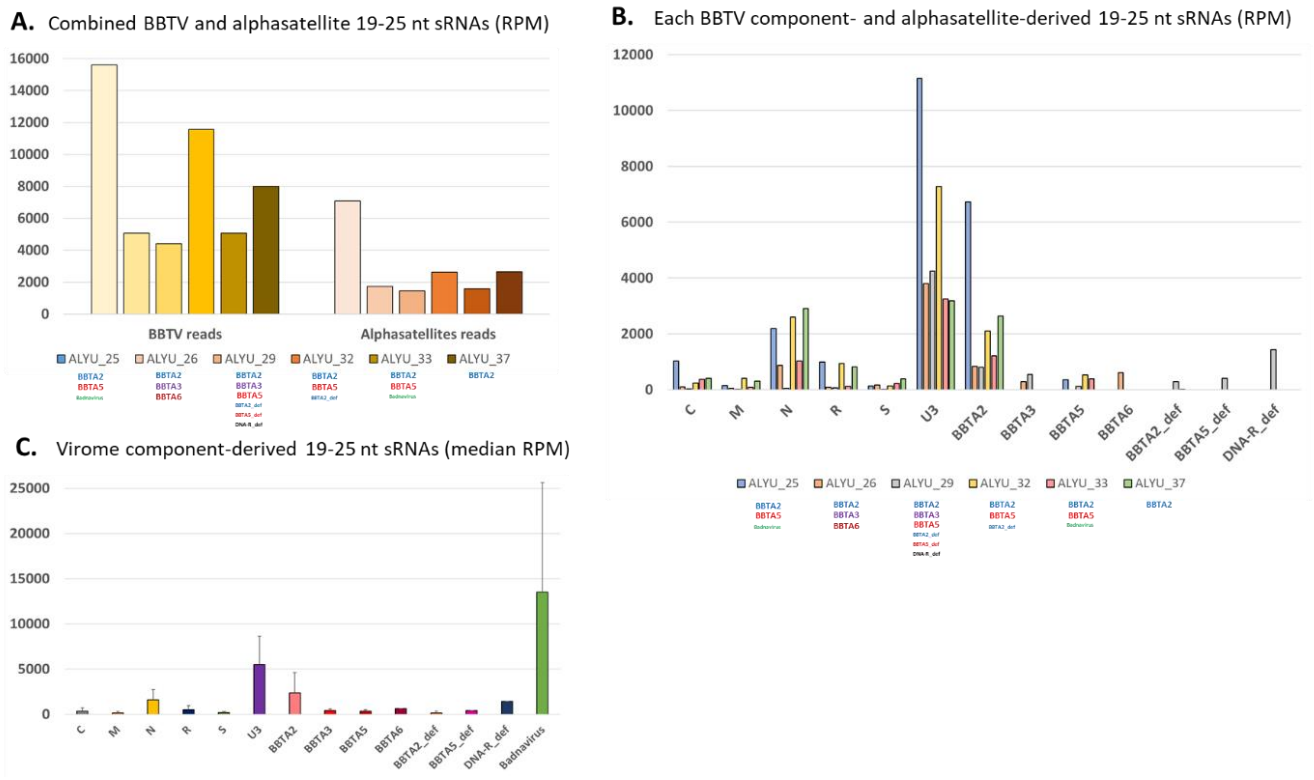


Figure 9. Counts of BBTV- and alphasatellite-derived 19-25 nt sRNAs in reads per million (RPM).

(A) Combined BBTV and alphasatellite 19-25 nt sRNAs in reads per million of total 19-25 nt sRNAs (RPM). **(B)** Counts for each BBTV component- and alphasatellite-derived 19-25 nt sRNAs (RPM). **(C)** Virome component-derived 19-25 nt sRNAs (median RPM).

Interestingly, the defective BBTA2 and BBTA5 molecules present together with defective DNA-R in the sample 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 (Figure 9B). While the sRNA formula of the three defective molecules and non-defective BBTA2 and DNA-R resembled well the DNA formula of these 5 virome components, non-defective BBTA5 being the weakest producer of sRNAs accumulated its DNA at higher levels than any of the latter components (Figure 10A-B), indicating its ability to evade RNAi.

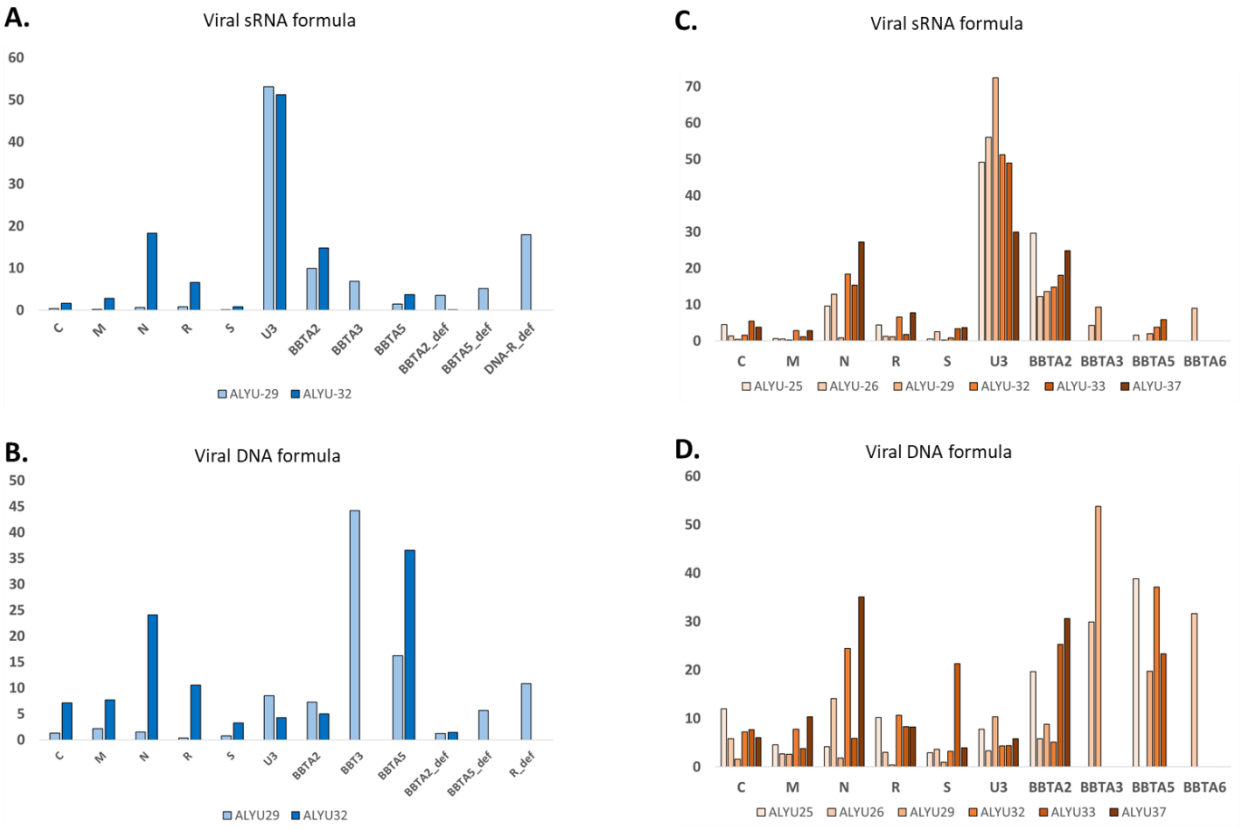


Figure 10. Comparison of viral sRNA and DNA formulas.
(A) and **(C)** Viral sRNA formulas. **(B)** and **(D)** Viral DNA formulas.

Size-class profiling of viral sRNAs revealed that BBTV and all alphasatellites spawn predominantly 21, 22 and 24-nt sRNAs (Figure 11 and Supplementary Figure S11). This is consistent with our previous findings for BBTV and banaphisatellite BBTA4 (Guyot *et al.*, 2022) and indicates that banana Dicer-like (DCL) 4, DCL2 and DCL3 activities generate respectively 21, 22 and 24 nt siRNAs that can potentially silence BBTV and alphasatellite gene expression at both transcriptional and post-transcriptional levels (Pooggin, 2018; Guyot *et al.*, 2022). Additionally, we found that the defective molecules derived from DNA-R, BBTA2 and BBTA5 also spawn predominantly 21, 22 and 24 nt sRNAs (Figure 9), suggesting that these circular non-coding molecules also generate double-stranded RNA precursors processed by DCL4, DCL2 and DCL3 into siRNAs of respective size-classes and may therefore serve as decoys diverting RNAi from BBTV and alphasatellites.

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

Interestingly, badnaviruses present in two samples were also found to spawn sRNAs whose accumulation levels were comparable to or exceeded those of other virome components (Figure 9B). In contrast to BBTV and alphasatellites, both badnaviruses spawned predominantly 21 and 22 nt sRNA, while 24 nt sRNAs were underrepresented. These results are consistent with our previous finding for 6 distinct badnaviruses in persistently infected banana plants, which spawned predominantly 21 and 22 nt siRNAs with the exception for two badnaviruses spawning 24 nts siRNAs at levels exceeding those of 22 nt (but not 21 nt) siRNAs (Rajeswaran et al., 2014).

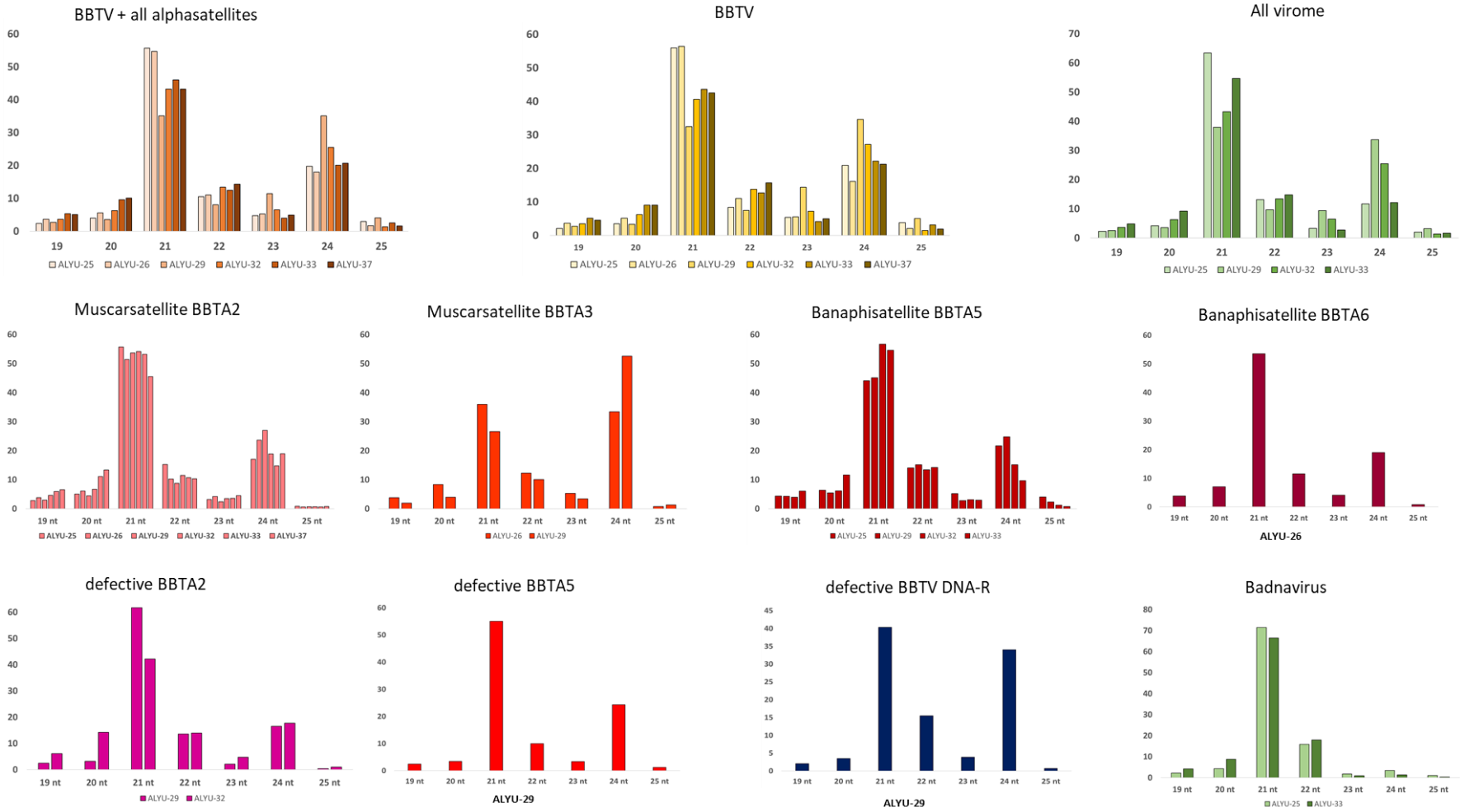


Figure 11. Relative abundance of 19-, 20-, 21-, 22-, 23-, 24- and 25-nt classes of viral sRNAs (size profile) derived from BBTV, alphasatellites and other virome components.

Conclusions and open questions

Our survey in South-East Asia revealed that BBTV alphasatellites are prevalent in Vietnam and China and less prevalent in Laos. We reconstructed complete genomes of new genetic variants of the muscarsatellite species BBTA2 and BBTA3 (genus *Muscarsatellite*, subfamily *Petromoalphasatellitinae*) previously identified in Vietnam, China and Taiwan as well as two new alphasatellite species - BBTA5 and BBTA6 - 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 species into a new genus of the subfamily *Nanoalphasatellitinae* - *Banaphisatellite* - and found that the 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 South-East Asia (BBTA5 and BBTA6) indicates that alphasatellites in their evolution and adaptation to new hosts (dicot vs monocots) may frequently change helper viruses within one family and even across families.

Previously, BBTV isolates from Vietnam, China and Taiwan were found to be associated with the muscarsatellites BBTA2 and BBTA3 and the babusatellite BBTA1 (genus *Babusatellite*, subfamily *Petromoalphasatellitinae*) (Supplementary Table S1) but not with any other alphasatellites. This suggests that the banaphisatellites BBTA5 and BBTA6 have emerged in SEA only recently, similar to the banaphisatellite 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 (Guyot *et al.*, 2022; Stainton *et al.*, 2015; Mukwa *et al.*, 2016). 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 BBTV alphasatellites and their helper viruses from PIO and SEA revealed that BBTA4 (the only satellite from PIO) does not possess any CR-M, while the CR-M sequences are well conserved in all other BBTV

alphasatellite species and in both PIO and SEA strains of BBTV. 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 enough time to co-evolve with its monocot-infecting helper BBTV in Congo-DRC. The high prevalence of the banaphisatellite BBTA5 and BBTA6 species in SEA would suggest their better adaptation to the helper BBTV. In contrast to CR-M, the CR-SL sequences are well conserved between the three species of banaphisatellites that share the high sequence identity not only in the stem-loop structure but also in the iterons and the Pol II promoter elements. This implies the conserved mechanisms of the initiation of rolling-circle replication by banaphisatellite Rep proteins and of the Pol II transcription of Rep-mRNA, both pointing at the potential for trans-replication between alphasatellite species within this genus, and highlighting their evolutionary links and the common origin. Likewise, the two species in the genus *Muscarsatellite* (BBTA2 and BBTA3) share all the conserved CR-SL elements and other sequence features between each other but not with the babusatellite BBTA1 or the banaphisatellites.

Our analysis of the relative abundances (DNA formulas) of BBTV alphasatellites and their helper virus components revealed that alphasatellite DNA accumulates at high levels in all plants, constituting 30 to 80 % of total viral (BBTV+alphasatellite) DNA. This indicates that self-replicating alphasatellites, being unable to mediate trans-replication of their helper virus components (as supported by our comparative analysis of the CR-SL sequences), can efficiently compete with their helper viruses for the host replication machinery and likely reduce the helper virus loads, as we have established earlier 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 the 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 RNA 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 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 the 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 new light on the provenance of alphasatellites, their co-evolution with helper ssDNA viruses and potential mutual benefits of alphasatellite-helper virus association.

Materials &Methods

Surveys in Vietnam, Laos and China

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

Total DNA extraction from banana leaf samples

Dried leaf tissue (100 mg) was ground in liquid nitrogen and 500 µl extraction buffer [100 mM Tris-HCl pH 8.0, 1.4 M NaCl, 20 mM EDTA, 2 % alkyltrimethylamonium bromide (MATAB), 1 % polyethyleneglycol 6000, 0.5 % sodium sulfite] pre-heated at 74°C and supplemented with 0.4 µl RNase (100 mg/ml) was added to the frozen powder. The mixture was vortexed for 20 sec, 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 13000 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 the appearance of a hank and then spun at 13000 rpm and 4°C for 30 min. The pellet was washed twice with 500 µl of 70 % ethanol, air dried and dissolved in 100 µl of milli-Q water.

Rolling circle amplification (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 µl Sample buffer and 1 µl total DNA extracted from banana leaf tissues were mixed and heated at 95°C for 3 min. The samples were cooled and 5 µl Reaction buffer and 0.2 µl enzyme mix were added, followed by incubation at 30°C for 18 hours. 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 ng of the cleaned undigested RCA products were taken for Illumina sequencing at Fasteris AG (www.fasteris.com). Libraries were prepared using Nextera XT standard DNA protocol and all libraries were multiplexed and sequenced in one flowcell of HiSeq2500 with a 2x 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, https://www.ebi.ac.uk/~zerbino/velvet/) 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 BWA (Burrows and Wheeler 1994) and visualisation using MISIS-2 (Seguin et al. 2016).

Total RNA extraction from banana leaves

RNA extraction was performed as described previously (Rajeswaran et al. 2014). Briefly, 0.5 dried leaf tissue was ground in liquid nitrogen and mixed with 2.5 ml extraction buffer [100 mM Tris-HCl (pH 7.5), 500 mM NaCl, 25 mM EDTA (pH 8.0), 1.5 % SDS, 2 % polyvinylpyrrolidone, 0.7 % beta-mercaptoethanol (added just before use)]. The mixture was incubated for 10 min at room temperature and then centrifuged at 3700 rpm for 15 min at 4°C. The supernatant was taken and mixed with 1/3 volume of 5 M sodium acetate (pH 6.0) and kept on ice for 30 min, followed by centrifugation at 13000 rpm for 15 min at 4°C. The supernatant was taken and mixed vigorously with an equal volume of phenol/chloroform/isoamyl alcohol (25:24:1), followed by centrifugation at 13000 rpm for 10 min at 4°C. The aqueous phase was taken and mixed vigorously with one volume of chloroform/isoamyl alcohol (24:1), followed

by centrifugation as described above. The aqueous phase was collected and mixed with one volume of isopropanol, followed by incubation for 5 min at room temperature and for 30 min at -80°C. RNA was pelleted by centrifugation for 15 min at a maximum speed at 4°C. The pellet was washed with 500 µl of 75 % ethanol, air-dried and dissolved in 50 µl RNase-free water.

Illumina sequencing and bioinformatics analysis of viral small RNA-ome

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 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 flowcell of NovaSeq 6000. Following de-multiplexing and adapter trimming, the resulting reads from each library were mapped using Burrow-Wheeler Aligner (BWA) 0.7.12 (Burrow and Wheeler 1994) onto the reference sequences of alphasatellites, BBTV and other components of the virome reconstructed by Illumina sequencing of RCA amplified viral DNA (as described above). Mapped viral reads were sorted by polarity (forward, reverse), by size (15 to 34 nts) and 5´-terminal nucleotide identity (5'U, 5'A, 5'G, 5'C) and counted, followed by normalization in reads per million (RPM) of total (viral + non-viral) reads. 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.

Phylogenetic analysis of alphasatellites and helper viruses

Reference sequences of all isolates of alphasatellites available in the NCBI Genbank in Feb 2022 were downloaded, adjusted to begin with the conserved nonanucleotide sequence TAGTATTAC and aligned to the reconstructed sequences of new alphasatellites using Sequence Demarcation Tool (SDT) v1.2 with Muscles (Muhire et al. 2014) to calculate their pairwise identities. A maximum likelihood phylogenetic tree was constructed using Seaview (Gouy et al. 2010) with Muscles 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).

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).

Acknowledgements

We thank Ly Ngoc Sam (Institute of Tropical Biology, Vietnam), Tien Ding Trieu (Nomafsi, Vietnam), Oudomephone Insisiendmay (Cabinet of Lao Academy of Science, Laos) and Ting Zhang (Kunming Institute of Botany, China) for their assistance with surveys in Vietnam, Laos and China, and Matthieu Chabannes for preparation of Nagoya protocol-based agreements and assistance in leaf sampling. The work was supported through BforBB project funds from Agropolis (Montpellier) to Marie-Line-Iskra Caruana and the project partners, INRAE SPE department funds (project ViroMix) to Mikhail M. Pooggin and a PhD scholarship from Institute Agro (Montpellier) granted to Valentin Guyot.

Data availability

Viral genome sequences obtained in this study were deposited in the NCBI Genbank with the accession numbers (to be done).

Competing Interests Statement

The authors declare no competing interests.

References

Beetham PR, Harding RM, Dale JL. **1999**. Banana bunchy top virus DNA-2 to 6 are monocistronic. *Archives of Virology* **144**: 89–105.

Briddon RW, Martin DP, Roumagnac P, Navas-Castillo J, Fiallo-Olivé E, Moriones E, Lett JM, Zerbini FM, Varsani A. **2018**. Alphasatellitidae: a new family with two subfamilies for the classification of geminivirus- and nanovirus-associated alphasatellites. *Archives of Virology* **163**: 2587–2600.

Briddon RW, Stanley J. **2006**. Subviral agents associated with plant single-stranded DNA viruses. *Virology* **344**: 198–210.

Burrows M, Wheeler DJ. 1994. A Block-Sorting Lossless Data Compression Algorithm. Technical Report 124, Digital Equipment Corporation, Digital Equipment Corporation Press, Palo Alto, CA, U.S.A. Pages 1-24.

Burns TM, Harding RM, Dale JL. **1995**. The genome organization of banana bunchy top virus: Analysis of six ssDNA components. *Journal of General Virology* **76**: 1471–1482.

Crooks, G. E., Hon, G., Chandonia, J. M., & Brenner, S. E. (2004). WebLogo: A sequence logo generator. *Genome Research*, *14*(6), 1188–1190. https://doi.org/10.1101/gr.849004

Dale JL. **1987**. Banana bunchy top: An economically important tropical plant virus disease. *Advances in Virus Research* **33**: 301–325.

Eagle PA, Orozco BM, Hanley-Bowdoin L. **1994**. A DNA sequence required for geminivirus replication also mediates transcriptional regulation. *The Plant Cell* **6**: 1157–1170.

Frickey T, Lupas A. 2004. CLANS: A Java application for visualizing protein families based on pairwise similarity. Bioinformatics 20: 3702–3704. https://doi.org/10.1093/bioinformatics/bth444

Gallet R, Fabre F, Michalakis Y, Blanc S. **2017**. The Number of Target Molecules of the Amplification Step Limits Accuracy and Sensitivity in Ultradeep-Sequencing Viral Population Studies. *Journal of Virology* **91**: 561–578.

Gouy M, Guindon S, Gascuel O. 2010. Sea view version 4: A multiplatform graphical user interface for sequence alignment and phylogenetic tree building. Mol Biol Evol. 27: 221–224. https://doi.org/10.1093/molbev/msp259.

Guyot V, Rajeswaran R, Chu HC, Karthikeyan C, Laboureau N, Galzi S, Mukwa LFT, Krupovic M, Kumar PL, Iskra-Caruana ML, Pooggin MM. 2022. A newly emerging alphasatellite affects banana bunchy top virus replication, transcription, siRNA production and transmission by aphids. *PLoS Pathog.* **18**: e1010448. doi: 10.1371/journal.ppat.1010448. Online ahead of print.

Hafner GJ, Harding RM, Dale JL. **1997a**. A DNA primer associated with banana bunchy top virus. *Journal of General Virology* **78**: 479–486.

Hafner GJ, Stafford MR, Wolter LC, Harding RM, Dale JL. **1997b**. Nicking and joining activity of banana bunchy top virus replication protein in vitro. *Journal of General Virology* **78**: 1795–1799.

Herrera-Valencia VA, Dugdale B, Harding RM, Dale JL. **2006**. An iterated sequence in the genome of Banana bunchy top virus is essential for efficient replication. *Journal of General Virology* **87**: 3409–3412.

Horser CL, Harding RM, Dale JL. **2001**. Banana bunchy top nanovirus DNA-1 encodes the 'master' replication initiation protein. *Journal of General Virology* **82**: 459–464.

Ji X long, Yu N tong, Qu L, Li B bin, Liu Z xin. **2019**. Banana bunchy top virus (BBTV) nuclear shuttle protein interacts and re-distributes BBTV coat protein in Nicotiana benthamiana. *3 Biotech* **9**: 1–6.

Karan M, Harding RM, Dale JL. 1994. Evidence for two groups of banana bunchy top virus isolates. J Gen Virol 75: 3541–3546. doi: 10.1099/0022-1317-75-12-3541.

Kumar PL, Hanna R, Alabi OJ, Soko MM, Oben TT, Vangu GHP, Naidu RA. **2011**. Banana bunchy top virus in sub-Saharan Africa: Investigations on virus distribution and diversity. *Virus Research* **159**: 171–182.

Lageix S, Catrice O, Deragon J-M, Gronenborn B, Pélissier T, Ramírez BC. **2007**. The Nanovirus-Encoded Clink Protein Affects Plant Cell Cycle Regulation through Interaction with the Retinoblastoma-Related Protein. *Journal of Virology* **81**: 4177–4185.

Lowe S, Browne M, Boudjelas S, Poorter M De. **2020**. 100 of the World's Worst Invasive Alien Species: A Selection From The Global Invasive Species Database. In: Encyclopedia of Biological Invasions. 715–716.

Magee CJP. **1940**. Transmission Studies on the Banana Bunchy-top Virus. *Journal of the*

Australian Institute of Agricultural Science **6**.

Mansourpour M, Gallet R, Abbasi A, Blanc S, Dizadji A, Zeddam J-L. **2022**. Effects of an Alphasatellite on the Life Cycle of the Nanovirus Faba Bean Necrotic Yellows Virus. *Journal of Virology* **96**.

Muhire, B. M., Varsani, A., & Martin, D. P. (2014). SDT: A virus classification tool based on pairwise sequence alignment and identity calculation. PLoS ONE, 9(9), e108277. https://doi.org/10.1371/journal.pone.0108277

Mukwa LF, Gillis A, Vanhese V, Romay G, Galzi S, Laboureau N, Kalonji-Mbuyi A, Iskra-Caruana ML, Bragard C. 2016. Low genetic diversity of Banana bunchy top virus, with a sub-regional pattern of variation, in Democratic Republic of Congo. Virus Genes 52: 900-905. doi: 10.1007/s11262-016-1383-1.

Pooggin MM. **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, Seguin J, Chabannes M, Duroy P-O, Laboureau N, Farinelli L, Iskra-Caruana M-L, Pooggin MM. 2014. Evasion of short interfering RNA-directed antiviral silencing in Musa acuminata persistently infected with six distinct banana streak pararetroviruses. J Virol 88: 11516-11528. doi: 10.1128/JVI.01496-14.

Seguin J, Otten P, Baerlocher L, Farinelli L, Pooggin MM. 2016. MISIS-2: A bioinformatics tool for in-depth analysis of small RNAs and representation of consensus master genome in viral quasispecies. J Virol Methods 233: 37-40. doi: 10.1016/j.jviromet.2016.03.005.

Stainton D, Martin DP, Collings DA, Varsani A. **2017**. Comparative analysis of common regions found in babuviruses and alphasatellite molecules. *Archives of Virology* **162**: 849–855.

Stainton D, Martin DP, Muhire BM, Lolohea S, Halafihi M, Lepoint P, Blomme G, Crew KS, Sharman M, Kraberger S, *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.

Varsani A, Martin DP, Randles JW, Vetten HJ, Thomas JE, Fiallo-Olivé E, Navas-Castillo J, Lett JM, Zerbini FM, Roumagnac P, *et al.* 2021. Taxonomy update for the family

Alphasatellitidae: new subfamily, genera, and species. *Archives of Virology* **166**: 3503–3511.

Wanitchakorn R, Harding RM, Dale JL. **1997**. Banana bunchy top virus DNA-3 encodes the viral coat protein. *Archives of Virology* **142**: 1673–1680.

Wanitchakorn R, Harding RM, Dale JL. **2000**. Sequence variability in the coat protein gene of two groups of banana bunchy top isolates. *Archives of Virology* **145**: 593–602.

Yu NT, Xie HM, Zhang YL, Wang JH, Xiong Z, Liu ZX. **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.

Chapter 3: Host range of banana bunchy top virus and its alphasatellites

Short communication (draft manuscript)

Evidence for dicot plants as alternative hosts of banana bunchy top virus and its alphasatellites in South-East Asia

Valentin Guyot[1], Marie-Line Iskra-Caruana[2], Mikhail M. Pooggin[1], *

[1] PHIM Plant Health Institute, University of Montpellier, INRAE, CIRAD, IRD, Institute Agro, 34398 Montpellier, France

[2] CIRAD, DGD-RS, 34398 Montpellier, France

* Correspondence: Dr. Mikhail M. Pooggin; mikhail.pooggin@inrae.fr

Banana bunchy top virus (BBTV, genus *Babuvirus*, family *Nanoviridae*) causes severe disease of cultivated bananas and plantains (*Musa sp.*) and represents one of the greatest threats to global food security (Lowe *et al.*, 2020). Since the beginning of 19th century the disease resulted in devastating epidemics reducing the banana production up to 95 % in several countries in South-East Asia and Africa (Dale, 1987; Kumar & Hanna, 2009). The disease is transmitted by the banana aphids *Pentalonia nigronervosa* and *P. caladii* (Watanabe *et al.*, 2013). The BBTV genome is composed of six circular single-stranded (ss)DNA components of ca. 1 to 1.1 kilobases (kb), each encoding for one protein and each encapsidated individually in isometric 18-20 nm virions. DNA-S encodes the capsid protein, while DNA-R encodes a master replication protein (Rep) that recruits the host DNA polymerase machinery for replication of DNA-R itself and trans-replication of other BBTV components. DNA-C encodes a cell-cycle link protein that facilitates replication, while DNA-M and DNA-N encode movement and nuclear shuttle proteins. DNA-U3 encodes a small protein of unknown function (Burns *et al.*, 1995; Wanitchakorn *et al.*, 2000; Qazi, 2016). In South-East Asia, BBTV is frequently associated with one or more circular ssDNA alphasatellites (family *Alphasatellitidae*) that are

similar in size to BBTV genome components and encode a Rep-like protein that mediates self-replication of alphasatellite DNA. Alphasatellites depend on their helper viruses for movement, encapsidation and transmission (Horser *et al.*, 2001; Bell *et al.*, 2002, Guyot *et al.*, 2022a/Chapter 1 of this thesis).

The host range of BBTV has so far been reported to be restricted to plant species of the monocot families Musaceae (*Musa acuminata*, *M. balbisiana*, *M. coccinea*, *M. jackeyi*, *M. ornata*, *M. textilis*, *M. velutina* and *Ensete ventricosum*), Zingiberaceae (*Alpinia zerumbet*, *Zingiber officinale*, *Curcuma longa*, *Kaempferia galanga*), Araceae (*Colocasia esculenta*), Cannaceae (*Canna indica*) and Heliconiaceae (*Heliconia aurantiaca*) (Thomas & Iskra Caruana, 2000; Thomas *et al.*, 2003; Pinili, M.S., I. Nagshima, T.O. Dizon, 2013; Hamim *et al.*, 2017; Rahayuniati *et al.*, 2021). Studies of alternative hosts for BBTV were mainly done on plant species co-cultivated with or grown nearby banana plants and those hosting *Pentalonia* aphids. The host range of *P. nigronervosa* and *P. caladii* was also reported to be restricted the monocot families including Araceae, Cannaceae, Commelinaceae, Heliconiaceae, Musaceae, Strelitzeaceae and Zingiberaceae (Blackman & Eastop, 1994; Bagariang *et al.*, 2019) and is largely overlapping with the BBTV host range.

During the surveys conducted in South-East Asia (Vietnam, Laos and China) in 2018 and 2019, leaf samples of wild and cultivated banana plants displaying BBTD symptoms were collected together with leaf samples of non-banana plants grown in close vicinity of the banana plants (Supplementary Figure S1, Table 1) (Guyot *et al.*, 2022b/Chapter 2 of this thesis). The non-banana species included the monocot *Commelina sp.* (Commelinaceae) and the dicots *Arachis hypogaea* (Fabaceae), *Bidens pilosa* (Asteraceae), *Chromolaena odorata* (Asteraceae), *Ipomoea aquatica* (Convolvulaceae) and *Phyllanthus sp.* (Phyllanthaceae), some of which such as *B. pilosa* exhibiting virus-like chlorotic symptoms (Supplementary Figure S2). Total DNA extracted from the leaf samples (dried over silica gel after sampling) was used for enrichment of circular viral DNA by rolling circle amplification (RCA) and Illumina sequencing of the resulting RCA products, followed by (i) *de novo* assembly of 125 nt paired-end Illumina reads to reconstruct complete genomes of circular DNA virome components and (ii) verification of consensus sequences of the reconstructed genomes by read mapping and analysis (see Guyot et al. 2022b/Chapter 2 of this thesis for the methodology details).

Sample ID	Country	Plant species/Genome	BBTV genome	No. of alpha-satellites	BBTA2	BBTA3	BBTA4	BBTA5	BBTA6	defective (d) molecules	others DNA viruses
ALYU-25	Vietnam	*Musa itinerans*	full	2	1			1			Badnavirus new
ALYU-26	Vietnam	*Musa sp.*	full	3	1	1			1		
ALYU-27	Vietnam	*Commelina sp.*	full	3	1	1			1		
ALYU-28	Vietnam	*Ipomoea aquatica*	no								
ALYU-29	Vietnam	*Musa* AA sucrier Pisang mas ?	full	3	1	1		1		d-BBTA5, d-BBT2, d-DNA-R	
ALYU-30	Vietnam	*Phyllanthus sp.*	no								
ALYU-31	Vietnam	*Arachis sp.*	no								
ALYU-32	Vietnam	*Musa* sp. Sweet banana	full	2	1			1		d-BBT2	
ALYU-33	Vietnam	*Musa* AAB Chuoi Ngop	full	2	1			1			Badnavirus BSVNV
ALYU-34	Vietnam	*Musa sp.*	full	2	1			1			
ALYU-35	Vietnam	*Musa sp.*	full	1					1		
ALYU-36	Vietnam	*Musa sp.*	full	1	1						
ALYU-37	Vietnam	*Musa sp.*	full	1	1						
ALYU-38	Vietnam	*Bidens pilosa*	full	4	1	1		1	1		Cyclovirus, Microvirus
ALYU-39	Vietnam	*Musa sp.*	full	2	1				1		
ALYU-40	Vietnam	*Musa* AAA Red banana	full	1				1			
ALYU-41	Laos	*Chromolaena odorata*	no N	1				1			
ALYU-42	Laos	*Musa acuminata* Cavendish AAA	full	1				1			
ALYU-43	Laos	*Musa acuminata* Cavendish AAA	full	1				1			
ALYU-44	Laos	*Musa ornata*	full								
ALYU-45	Laos	*Musa sp.*	full								
ALYU-46	Laos	*Musa* ABB Klue Tiparot	no								
ALYU-47	Laos	*Musa yunnenensis*	full								
ALYU-48	Laos	*Musa sp.*	full								
ALYU-49	Laos	*Musa sp.*	full								
ALYU-50	Laos	*Musa* AA sucrier Kouay niew mung	full								
ALYU-51	Laos	*Musa* ABB Pisang Awak ?	full								
ALYU-52	China	*Musa acuminata* wild	no								
ALYU-53	China	*Musa yunnanensis*	full								
ALYU-54	China	*Musa acuminata* Cavendish AAA	full	2	1				1	d-BBTA6	
ALYU-55	China	*Musa acuminata* Cavendish AAA	full	3	1	1			1		
ALYU-56	China	*Musa acuminata* Cavendish AAA	full	1	1						

Table 1. Virome components identified in wild and cultivated banana (*Musa*) and non-banana plants sampled in Vietnam, Laos and China during surveys in 2018 and 2019.

In the samples of *Arachis hypogaea* (ALYU-31), *Ipomoea aquatica* (ALYU-28), *Phyllanthus sp.* (ALYU-30) and two of the 26 banana plants (ALYU-46 and ALYU-52), very low numbers of viral reads were detected, which likely represent cross-contamination from other samples multiplexed and sequenced in one flowcell of Illumina HiSeq2500; this assumption was supported by inspection of read coverage with and without mismatches (Supplementary Figure S3). These virus-free samples were taken to establish the cross-contamination threshold (0.0006 % of total reads; Supplementary Figure S4), based on which *Commelina sp.* (ALYU-27), *Bidens pilosa* (ALYU-38) and *Chromolaena odorata* (ALYU-41) were considered to be infected with BBTV and coinfected with one or more alphasatellites (Table 1 and Supplementary Figures S3 and S4).

Both *Commelina* (ALYU-27) and its neighbour banana (ALYU-26) shared 100 % identical consensus sequences of all the virome components including six components of BBTV genome and three species of BBTV alphasatellites (BBTA2, BBTA3, BBTA6), although single-nucleotide polymorphism (SNP) profiles of their virome quasispecies population (calculated using MISIS-2, Seguin *et al.*, 2016) differed substantially. Furthermore, relative abundances of their virome components (virome DNA formulas) were found to be similar but not identical (Figure 1A). The percentage of viral reads in the total (plant+viral) reads was found to be much lower in *Commelina* (0.004 %) than in its neighbour banana (0.64 %) (Supplementary Figure S4). These findings suggest that *Commelina* is a poor host for BBTV and its alphasatellites, allowing very low-titre accumulation of viral DNA compared to *Musa*.

The viromes of *Chromolaena* (ALYU-41) and its two *Musa* neighbours (ALYU-42 and ALYU-43) all contained BBTV and single alphasatellite (BBTA5). However, both BBTV and BBTA5 were represented with distinct genetic variants. Most notably, BBTV DNA-N was below the cross-contamination threshold in *Chromolaena*, both DNA-R and DNA-S differed at several distinct nucleotide positions between all 3 samples, and BBTA5 differed at single distinct nucleotide positions between all 3 samples. The virome DNA formulas differed substantially not only between *Chromolaena* and *Musa* neighbours (due to the absence of DNA-N in the former), but also between the two *Musa* neighbours (Figure 1B). The percentage of viral reads in total reads was found to be much lower in *Chromolaena* (0.002 %) than in its Musa neighbours (0.13 % in ALYU-42 and 0.22 % in ALYU-43) (Supplementary Figure S4).

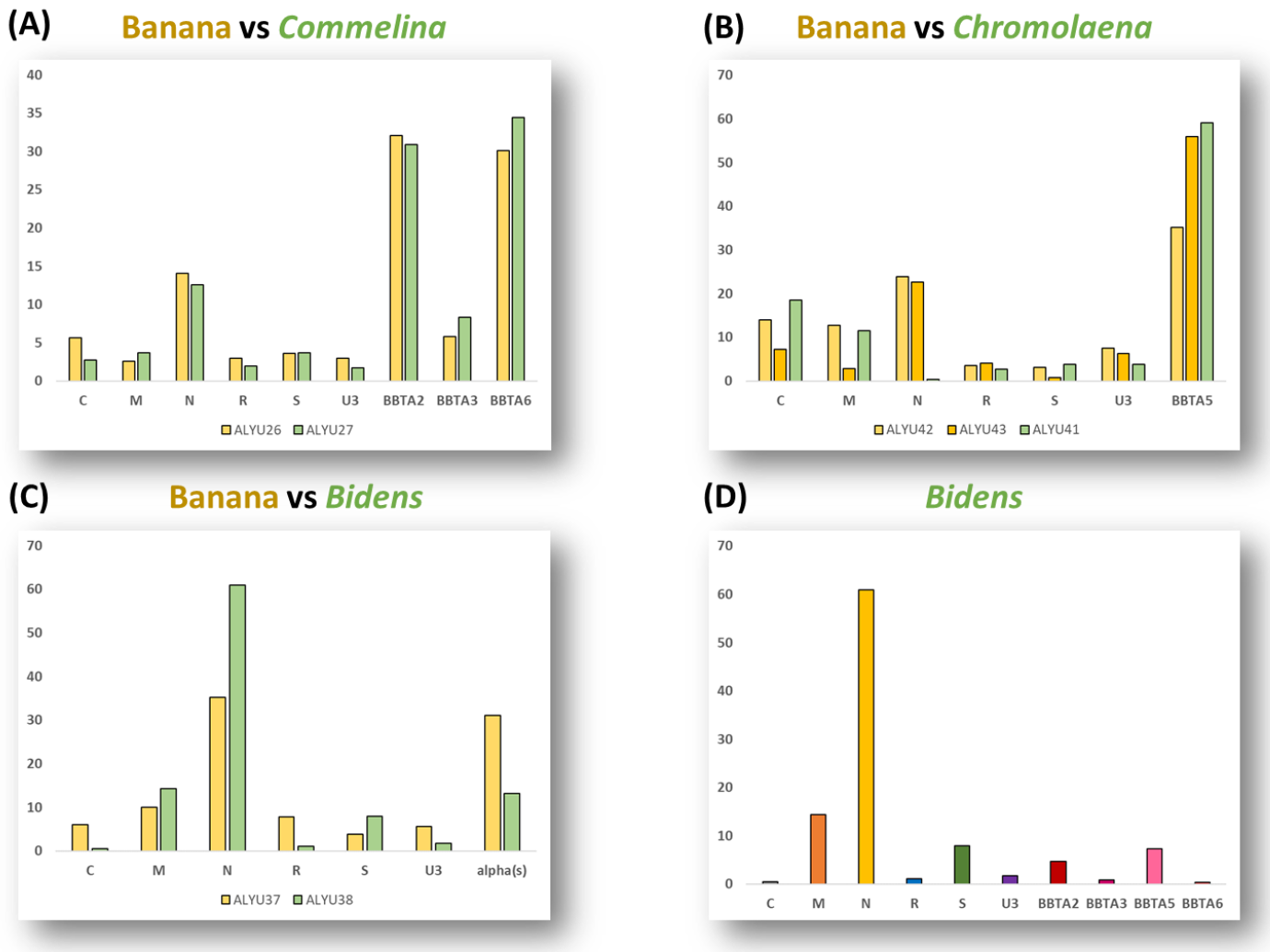


Figure 1. Relative abundance (formula) of virome components in *Commelina sp* (A), *Chromolaena odorata* (B) and *Bidens pilosa* (C&D) and their neighbour banana plants

Thus, similar to *Commelina*, *Chromolaena odorata* appeared to be a poor host for BBTV and its alphasatellite BBTA5. Previously, we have shown that DNA-N could be lost upon aphid transmission without affecting BBTV symptom development in recipient *Musa acuminata* Cavendish plants, but the virus lacking DNA-N was not transmissible by *P. nigronervosa* (Guyot et al. 2022a/Chapter 1 of this thesis). Likewise, DNA-N of faba been necrotic yellows virus (genus *Nanovirus*, family *Nanoviridae*) is essential for aphid transmission but not for disease symptom development (Grigoras *et al.*, 2018; Di Mattia *et al.*, 2020). Based on these findings, BBTV-infected *Chromolaena odorata* lacking DNA-N is unlikely to serve as a source for virus transmission, while the self-replicating alphasatellite BBTA5 being the most abundant virome component (Figure 1B) and likely encapsidated by BBTV capsid protein (expressed from DNA-S) can potentially be transmissible to other plants by viruliferous aphids carrying complete BBTV (or by virus-free aphids to BBTV-infected plants).

The viromes of *Bidens* (ALYU-38) and its *Musa* neighbour (ALYU-38) shared six-component BBTV and one alphasatellite (BBTA2), with all these 7 components having 100 % identity in their consensus sequences. In addition, the *Bidens* virome contained three more alphasatellites (BBTA3, BBTA5, BBTA6) and two Rep-encoding circular ssDNA viruses, classified as novel species in the non-plant families *Microviridae* and *Circoviridae* (manuscript in preparation); all these additional components were below the detection threshold in the *Musa* neighbour. Remarkably, the alphasatellites BBTA3 and BBTA6 identified in *Bidens* share 100% sequence identity with the respective alphasatellites identified in *Commelina* and one of its *Musa* neighbours (ALYU-26) sampled at a far-away location. Likewise, the alphasatellite BBTA5 from *Bidens* shares 100% sequence identity with BBTA5 identified in the *Musa* sample ALYU-40 at another far-away location in Vietnam (Supplementary Figure S1 and Table 1). These findings highlight the genetic stability of BBTV alphasatellites, consistent with our analysis of all 26 isolates representing the four alphasatellite species in Vietnam (Table 1; Guyot et al. 2022b/Chapter 2 of this thesis). The virome DNA formulas in *Bidens* and its *Musa* neighbour were found to be different, although in both cases DNA-N and (combined) alphasatellite DNA accumulated at high levels (Figure 1C; Supplementary Figure S1). Most notably, the combined BBTV and alphasatellite reads constituted 0.24 % of the total (plant + viral) reads in Bidens, which is more than 50 times higher than in *Commelina* and *Chromolaena* and only ~4 times lower than in the Musa neighbour (0.90 %). These findings indicate that *Bidens pilosa* is a

permissive host for BBTV and all four BBTV alphasatellite species circulating in South-East Asia.

Collectively, our findings suggest that viruliferous banana aphids could have transmitted BBTV and alphasatellites from the respective *Musa* neighbours not only to *Commelina* species representing the monocot family Commelinaceae that falls within the known host range of *Pentalonia* aphids and BBTV, but also to *Chromolaena odorata* and *Bidens pilosa* both representing the dicot family Asteraceae that was unexpected to host BBTV or banana aphids. In the latter case, even a short-term probing and salivation event would be sufficient for the release of viral particles circulating in the aphid body and accumulating in the salivary glands (Watanabe et al. 2013; Watanabe & Bressan, 2013). It remains to be investigated if banana aphids can feed on BBTV- and alphasatellite-infected *Bidens* and *Chromolaena* plants for a prolonged time to acquire BBTV and its alphasatellites and transmit them back to Musaceae hosts or among these and other Asteraceae plants.

Phylogenetic analysis of the alphasatellites associated with our BBTV isolates from Vietnam, Laos and China (Table 1) revealed that, besides new genetic variants of BBTA2 and BBTA3 species previously found to be associated with other BBTV isolates from South-East Asia and classified in the genus *Muscarsatellite* of the subfamily *Petromoalphasatellitinae* (comprising alphasatellites of several genera that infect monocots), other alphasatellites represent the two new species BBTA5 and BBTA6 that belong to a new genus *Banaphisatellite* within the subfamily *Nanoalphasatellitinae* (Guyot et al. 2022b/Chapter 2 of this thesis). Until recently, the latter subfamily was known to comprise several genera of alphasatellites that infect dicots. The first species classified in the genus *Banaphisatellite* was banana bunchy top alphasatellite 4 (BBTA4) recently discovered to be associated with BBTV isolates from the banana aphid and plant samples collected in Africa (Democratic Republic of the Congo) (Guyot et al. 2022a/Chapter 1 of this thesis). A hypothetical dicot origin of the banaphisatellite BBTA4 proposed based on its phylogeny and the ability of its clone to infect the model dicot *Nicotiana benthamiana* (Solanaceae) (Guyot et al. 2022a/Chapter 1 of this thesis) is now supported by the above-described findings that the banaphisatellite BBTA5 and BBTA6 can naturally infect the dicot plants *Bidens pilosa* and *Chromolaena odorata*.

Acknowledgements

We thank Ly Ngoc Sam (Institute of Tropical Biology, Vietnam), Tien Ding Trieu (Nomafsi, Vietnam), Oudomephone Insisiendmay (Cabinet of Lao Academy of Science, Laos) and Ting Zhang (Kunming Institute of Botany, China) for their assistance with surveys in Vietnam, Laos and China, and Matthieu Chabannes for preparation of Nagoya protocol-based agreements and assistance in leaf sampling. The work was supported through BforBB project funds from Agropolis (Montpellier) to Marie-Line Iskra Caruana and the project partners, INRAE SPE department funds (project ViroMix) to Mikhail M. Pooggin and a PhD scholarship from Institute Agro (Montpellier) granted to Valentin Guyot.

References

Bagariang W, Hidayat P, Hidayat SH. **2019**. Morphometric Analysis and Host Range of the Genus Pentalonia Coquerel (Hemiptera: Aphididae) Infesting Banana in Java. *Jurnal Perlindungan Tanaman Indonesia* **23**: 171.

Bell KE, Dale JL, Ha C V, Vu MT, Revill PA. **2002**. Characterisation of Rep-encoding components associated with banana bunchy top nanovirus in Vietnam. *Archives of Virology* **147**: 695–707.

Blackman RL, Eastop VF. **1994**. Aphids on the world's trees: an identification and information guide. *Aphids on the world's trees: an identification and information guide*: 466.

Burns TM, Harding RM, Dale JL. **1995**. The genome organization of banana bunchy top virus: Analysis of six ssDNA components. *Journal of General Virology* **76**: 1471–1482.

Dale JL. **1987**. Banana bunchy top: An economically important tropical plant virus disease. *Advances in Virus Research* **33**: 301–325.

Grigoras I, Vetten HJ, Commandeur U, Ziebell H, Gronenborn B, Timchenko T. **2018**. Nanovirus DNA-N encodes a protein mandatory for aphid transmission. *Virology* **522**: 281–291.

Guyot V, Rajeswaran R, Chu HC, Karthikeyan C, Laboureau N, Galzi S, Mukwa LFT, Krupovic M, Kumar PL, Iskra-Caruana ML, Pooggin MM. 2022. A newly emerging alphasatellite affects banana bunchy top virus replication, transcription, siRNA production and transmission by aphids. *PLoS Pathog.* **18**: e1010448. doi: 10.1371/journal.ppat.1010448.

Online ahead of print.

Hamim I, Green JC, Borth WB, Melzer MJ, Wang YN, Hu JS. **2017**. First report of Banana bunchy top virus in Heliconia spp. on Hawaii. *Plant Disease* **101**: 2153.

Horser CL, Harding RM, Dale JL. **2001**. Banana bunchy top nanovirus DNA-1 encodes the 'master' replication initiation protein. *Journal of General Virology* **82**: 459–464.

Kumar L, Hanna R. **2009**. Tackling the banana menace. *IITA - R4D Review*: 11–13.

Lowe S, Browne M, Boudjelas S, Poorter M De. **2020**. 100 of the World's Worst Invasive Alien Species: A Selection From The Global Invasive Species Database. In: Encyclopedia of Biological Invasions. 715–716.

Di Mattia J, Vernerey M-S, Yvon M, Pirolles E, Villegas M, Gaafar Y, Ziebell H, Michalakis Y, Zeddam J-L, Blanc S. **2020**. Route of a Multipartite Nanovirus across the Body of Its Aphid Vector. *Journal of Virology* **94**.

Pinili, M.S., I. Nagshima, T.O. Dizon KTN. **2013**. Cross-Transmission and New Alternate Hosts of Banana bunchy top virus. *Tropical Agriculture and Development* **57**: 1–7.

Qazi J. **2016**. Banana bunchy top virus and the bunchy top disease. *Journal of General Plant Pathology* **82**: 2–11.

Rahayuniati RF, Subandiyah S, Hartono S, Somowiyarjo S, Kurniawan REK, Prakoso AB, Crew K, Vance ME, Ray JD, Thomas JE. **2021**. Recent distribution and diversity analysis on banana bunchy top virus of banana and alternative host in Indonesia. *Tropical Plant Pathology* **46**: 506–517.

Thomas JE, Geering ADW, Dahal G, Lockhart BEL, Thottappilly G. **2003**. Banana and Plantain. In: Virus and Virus-like Diseases of Major Crops in Developing Countries. Springer, Dordrecht, 477–496.

Thomas JE, Iskra Caruana ML. **2000**. Bunchy top. In: Diseases of Banana, Abaca and Enset. 241–253.

Wanitchakorn R, Hafner GJ, Harding RM, Dale JL. **2000**. Functional analysis of proteins encoded by banana bunchy top virus DNA-4 to -6. *Journal of General Virology* **81**: 299–306.

Watanabe S, Bressan A. **2013**. Tropism, compartmentalization and retention of banana bunchy top virus (Nanoviridae) in the aphid vector Pentalonia nigronervosa. *Journal of General*

Virology **94**: 209–219.

Watanabe S, Greenwell AM, Bressan A. **2013**. Localization, concentration, and transmission efficiency of banana bunchy top virus in four asexual lineages of Pentalonia aphids. *Viruses* **5**: 758–776.

Chapter 4: Review on alphasatellites of plant ssDNA viruses

Draft manuscript

The mystery of self-replicating alphasatellites of plant ssDNA viruses

Valentin Guyot[1], **Mikhail M. Pooggin**[1, *]

[1] PHIM Plant Health Institute, University of Montpellier, INRAE, CIRAD, IRD, Institute Agro, 34398 Montpellier, France

*Correspondence: Dr. Mikhail M. Pooggin; mikhail.pooggin@inrae.fr

Introduction

Viruses and virus-like agents belong to mobile genetic elements (MGEs), are found in all domains of life - bacteria, archea and eukaryota - and are supposed to be hosted by their Last Universal Common Ancestor (LUCA) (Krupovic *et al.*, 2020a; Koonin *et al.*, 2021). Viruses are not considered as living organisms, because they rely on a host cell for their replication. Recently, viruses were redefined by Koonin et al. (2021) as "replicators that encode structural proteins encasing their own genomes". Viruses and virus-like MGEs that do not encode a capsid protein are included in the virosphere, a dynamic entity composed of the orthovirosphere comprising true viruses that is surrounded by the perivirosphere comprising other kinds of virus-like replicators such as viroids, virusoids, and satellite nucleic acids (Koonin *et al.*, 2021). Viroids are autonomous replicators, while virusoids and satellite nucleic acids rely on their helper viruses for encapsidation that provides protection and facilitates movement and transmission among hosts (Francki, 1985; Hu *et al.*, 2009; Badar *et al.*, 2021). Satellite nucleic acids should not be confused with the satellite viruses that encode a capsid protein but rely on a helper virus for replication and other aspects of their infection cycle (Krupovic *et al.*, 2016). Therefore, the satellite viruses fit to the definition of a virus, but because they need a helper virus for replication, satellite viruses belong at the periphery of the orthovirosphere (Koonin *et*

al., 2021). Notably, some satellite nucleic acids such as alphasatellites are capable of self-replication in the absence of any helper virus but, because they rely on helper viruses for encapsidation and other aspects of their infection cycle (Horser *et al.*, 2001; Briddon *et al.*, 2001; Cui *et al.*, 2005; Gnanasekaran & Chakraborty, 2018; Mubin *et al.*, 2020; Guyot et al. 2022a/Chapter 1 of this thesis), self-replicating satellites belong to the perivirosphere and are placed near the border with the orthovirosphere (Koonin *et al.*, 2021). Based on the nature and structure of their genomes, satellite nucleic acids are divided into the RNA satellites with linear double stranded (ds)RNA, linear single stranded (ss)RNA, and circular ssRNA genomes (the latter satellites are named virusoids to distinguish them from fully autonomous viroids with circular ssRNA genomes) and the DNA satellites (Badar *et al.*, 2021). All DNA satellites discovered so far have small (up to 1.3 Kb) circular ssDNA genomes and are associated with the circular ssDNA helper viruses infecting plants and belonging to the families *Geminiviridae*, *Nanoviridae* and *Metaxyviridae*. The number of DNA satellites species has been growing during the last two decades thanks to advances in deep sequencing technology and metagenomics approaches (Ferro *et al.*, 2017; Bornancini *et al.*, 2020). Since DNA satellites have low sequence homology with their helper viruses, PCR based approached may not be applicable for their detection, while approaches based on rolling circle amplification (RCA) of circular DNA followed by Illumina sequencing of RCA products and de novo assembly of sequencing reads enable discovery of new satellites and their helper viruses. Moreover, *Geminiviridae* is the second largest plant virus family which also explains the diversity of DNA satellites associated with geminivirids (Fiallo-Olivé *et al.*, 2021). The circular ssDNA satellites are classified into alphasatellites (family *Alphasatellitidae*) that encode a replication protein (Rep) and betasatellites and deltasatellites (both in the family *Tolecusatellitidae*) that rely on helper viruses for replication. So far only alphasatellites were found to be associated with all the three plant ssDNA virus families. Unlike betasatellites that encode a pathogenicity determinant and suppressor of RNA silencing (betaC1 protein) (Zhou, 2013; Lozano *et al.*, 2016), deltasatellites and alphasatellites do not appear to play a major role in helper virus symptom development (Briddon *et al.*, 2004; Briddon & Stanley, 2006; Fiallo-Olivé *et al.*, 2016). Alphasatellite-encoded Rep mediates self-replication of its own genome but not trans-replication of the *bona fide* genome components of helper viruses (Timchenko *et al.*, 1999, 2000; Horser *et al.*, 2001). Like other DNA satellites, alphasatellites depend on their helper viruses for movement, encapsidation and plant-to-plant transmission by insect vectors (Saunders & Stanley, 1999; Mansoor *et al.*, 2001). The evolutionary origin of alphasatellites is unknown, although they resemble DNA-R of the multicomponent nanovirids, which encodes a

master-Rep (M-Rep) protein that mediates replication of DNA-R itself and trans-replication of other components of nanovirid genomes. Recently-discovered genome composition of coconut foliar decay virus (CFDV) revealed that up to 8 alphasatellites are associated with this virus. Moreover, the *bona fide* genome component of CFDV encoding M-Rep, likely mediating trans-replication of the putative coat protein and movement protein-encoding components, phylogenetically resembles members of the family *Alphasatellitidae* rather than DNA-R components of nanovirids (Gronenborn *et al.*, 2018). Thus, this multicomponent ssDNA virus is now considered as a founding member of a unique genus *Cofodevirus* in the new viral family *Metaxyviridae* ('metaxy' = Greek for 'in-between') standing in-between the alphasatellite family *Alphasatellitidae* and the viral families *Nanoviridae* and *Geminiviridae* (https://talk.ictvonline.org/taxonomy/p/taxonomy-history?taxnode_id=202012153). In this review, we describe history, diversity, classification, genome organisation, replication and gene expression of alphasatellites as well as their evolution and interactions with helper viruses and their impact on helper virus replication, gene expression, evasion of antiviral defences and transmission by insect vectors.

History, classification and host range of alphasatellites

Alphasatellites were discovered during the 1990s to be associated with plant viruses from the families *Nanoviridae* (Wu Rey Yuh *et al.*, 1994; Yeh *et al.*, 1994) and *Geminiviridae* (Mansoor *et al.*, 1999; Saunders & Stanley, 1999; Briddon *et al.*, 2004) and in 2018 to be part of the coconut foliar decay virus complex whose alphasatellite-like DNA encoding M-Rep was first reported in 1990 (Rohde *et al.*, 1990; Gronenborn *et al.*, 2018). Despite *Geminiviridae* is the second largest family of plant viruses, presently split into 14 genera, geminialphasatellites (subfamily *Geminialphasatellitinae*) have so far been found in association with helper geminivirids from only two genera *Begomovirus* and *Mastrevirus* in both monocot and dicot hosts (Kumar *et al.*, 2014; Briddon *et al.*, 2018; Hamza *et al.*, 2018; Claverie *et al.*, 2020). The latter two genera are better studied because they comprise viruses responsible for permanent crop losses around the world (Varma & Malathi, 2003) and consequently are the biggest genera of geminiviruses comprising 445 and 45 species to date, respectively (Fiallo-Olivé *et al.*, 2021). It is therefore likely that yet-to-be-discovered alphasatellites can potentially be associated with other genera of *Geminiviridae*. Alphasatellites associated with begomoviruses were initially designated as DNA-1 because their genome organisation is more related to that of DNA-1 (now

known as DNA-R) of nanovirids and is supposed to be derived from nanovirid DNA-R (Briddon & Stanley, 2006; Zhou, 2013). On the other hand, alphasatellites associated with nanovirids were initially designated as DNA components encoding Rep-like proteins to distinguish them from DNA-R encoding the master Rep (M-Rep) the only viral protein essential for viral replication and able to initiate trans-replication of the *bona fide* genome components of nanovirids (see below). Later, both were named alphasatellites to distinguish them from other satellites that do not encode Rep or Rep-like proteins and were gathered into the family *Alphasatellitidae* (Briddon *et al.*, 2018). *Alphasatellitidae* is currently divided into three sub-families (see Figure 1, Chapter 2): *Geminialphasatellitinae* comprising geminivirid-associated alphasatellites separated into 7 genera, *Nanoalphasatellitinae* comprising dicot nanovirid (genus *Nanovirus*)-associated alphasatellites separated into 6 genera, and *Petromoalphasatellitinae* comprising monocot nanovirid (genus *Babuvirus*)-associated alphasatellites and cofodevirid (genus *Cofodevirus*)-associated alphasatellites separated into 5 genera (Varsani *et al.*, 2021). Importantly, this classification is based on pairwise identity of nucleotide sequences, host range and helper virus identity. Recently, we have shown that banana bunchy top virus (BBTV), belonging to the genus *Babuvirus* and infecting monocots mainly of the families Musaceae and Zingiberaceae is naturally associated not only with alphasatellites of the subfamily *Petromoalphasatellitinae* but also with alphasatellites phylogenetically more related to those of the subfamily *Nanoalphasatellitinae* (Guyot *et al.*, 2022a and 2022b/Chapters 2 and 3 of this thesis). This suggests that alphasatellites associated with viruses infecting dicots can expand their host range and adapt to viruses infecting monocots and vice versa. In support of this hypothesis, the monocot-infecting geminivirid wheat dwarf India virus from the genus *Mastrevirus* was found to be naturally associated with two alphasatellites of dicot origin: (i) cotton leaf curl Multan alphasatellite (CLCuMA) previously identified in a complex with cotton leaf curl Multan virus from the genus *Begomovirus*, and (ii) guar leaf curl alphasatellite previously identified in a complex with guar leaf curl virus from the genus *Begomovirus* (Kumar *et al.*, 2014). However, the latter alphasatellites are classified into the same subfamily *Geminialphasatellitinae*, unlike those associated with BBTV that belong to two different subfamilies. It remains to be investigated through more extensive field surveys how frequent in their evolution could alphasatellites swap genera, families and higher taxa of their helper viruses and their host organisms.

Genome organisation and replication of alphasatellites and their helper viruses

Geminivirids have twinned (geminate) icosahedral virions each containing a single circular single-stranded genomic DNA of ~ 2,600–3,600 nucleotides. They are classified in 14 genera, 13 of which having a monopartite genome and one genus (*Begomovirus*) comprising members with both monopartite and bipartite genomes (Roumagnac *et al.*, 2022). So far, alphasatellites were found to be associated with monopartite and bipartite begomoviruses infecting dicots as well as the monocot-infecting monopartite mastrevirus (genus *Mastrevirus*) as mentioned above. Notably, bipartite begomoviruses have a DNA-A component encoding Rep, coat protein and other functions on both virion and complementary strands, which is similar to monopartite begomoviruses and other geminivirids, and a unique DNA-B component encoding two additional proteins with movement functions – nuclear shuttle protein (NSP) and movement protein (MP) - on the virion and the complementary strands, respectively. DNA-A of bipartite begomoviruses can establish systemic infection in the absence of DNA-B, albeit with less severe (or no) symptoms and at lower DNA titres compared to co-infections with DNA-A and DNA-B (Karthikeyan *et al.*, 2016; Iqbal *et al.*, 2017). Hence, begomoviral DNA-B can also be considered as a satellite depending on DNA-A for replication, encapsidation and transmission by the insect vector (the whitefly *Bemisia tabaci*, family *Aleyrodidae*, order Hemiptera, for both monopartite and bipartite begomoviruses). Mastreviruses have a monopartite genome of ca. 2.7 kb, encoding Rep, CP and other proteins on both virion and complementary strands, infect mainly monocots and some dicots and are transmitted by leafhoppers (family *Cicadellidae*, order Hemiptera) (Muhire *et al.*, 2013; Fiallo-Olivé *et al.*, 2021). Nanovirids have a multipartite genome comprising 6 (genus *Babuvirus*) to 8 (genus *Nanovirus*) circular ssDNA components, each encoding one protein, and each encapsidated individually inside isometric 17-20 nm virions (Gronenborn, 2004; Thomas *et al.*, 2021). Nanovirids infect either dicot (genus *Nanovirus*) or monocot (genus *Babuvirus*) hosts and are transmitted by aphids (order Hemiptera) in circulative, non-propagative manner (Wang & Blanc, 2021).

Both nanovirids and geminivirids are restricted to plant phloem tissue and replicate in the nucleus of companion cells by rolling circle replication (RCR) mediated by viral Rep and the host DNA replication machinery via a circular double-stranded (ds)DNA intermediate that also serves as a template for Pol II-mediated transcription of viral mRNAs (King *et al.*, 2011; Pooggin 2013).

Nanovirid-associated alphasatellites have ssDNA genome of ~1-1.1 kb, a size of their helper virus components, which would likely enable encapsidation of alphasatellite DNA by helper virus coat protein in icosahedral virions of similar size. In contrast, geminivirid-associated alphasatellites have a ssDNA genome of ca. ~1.3-1.4 kb, a half-size of their helper viruses, raising a question on their encapsidation in twinned icosahedral virions like helper virus DNA or in smaller "split-twin" icosahedral virions, or small icosahedral virions similar to those of nanovirids. It is assumed that alphasatellites replicate via the RCR mechanism similar to their helper viruses and are also phloem-restricted in their host plants (Alberter *et al.*, 2005; Briddon & Stanley, 2006). Geminialphasatellite- and nanoalphasatellite-encoded Rep proteins have a comparable size and the difference in their genome sizes is mainly due to a ~0.2 Kb insertion which is adenine-rich in most but not all geminialphasatellites (Claverie *et al.*, 2020). This insertion is supposed to be an evolutionary adaptation of progenitors of geminialphasatellites to reach a half size of helper geminivirids for better encapsidation, because an icosahedral virion geometry would impose size restriction on encapsidated DNA molecules (King *et al.*, 2011). All alphasatellites have a monocistronic genome organisation with a single Rep-encoding ORF on the virion strand, which is similar to DNA-R of nanovirids. The Rep ORF is preceded by the Pol II promoter with a TATA box and followed by the Pol II terminator with a polyadenylation signal (Guyot *et al.*, 2022a).

Alphasatellite Rep is evolutionary related to Rep proteins of all circular ssDNA viruses hosted by eukaryota and bacteria and designated as circular Rep-encoding single stranded DNA (CRESS-DNA) viruses, which are classified into the phylum *Cressdnaviricota* (Zhao *et al.*, 2019; Krupovic *et al.*, 2020b), suggesting the conservation of a Rep-mediated RCR mechanism in this phylum and the family *Alphasatellitidae*. As established for geminivirids and nanovirids, Rep protein binds the origin of replication composed of (i) a stem-loop secondary structure with the conserved nonanucleotide sequence TATTATTAC in the loop and (ii) short iterated sequence repeats (iterons) that are not conserved among different viruses and thought to provide sequence specificity for Rep binding and replication of its own virus but not of other viruses (Ilyina & Koonin, 1992; Argüello-Astorga *et al.*, 1994). At the replication initiation step, Rep recognizes nonanucleotide sequence and nicks it between 7th and 8th nucleotides and then gets attached to the 5' end of the cleaved virion strand to open the dsDNA intermediate and allow the host DNA polymerase to initiate synthesis of a new virion strand using the complementary strand as the template (see Figure 8 in the Introduction section). Then, Rep nicks the extended virion strand at the same position and, following displacement of the pre-existing virion strand,

ligates the termini of the released monomer, yielding the circular ssDNA product of RCR. Both binding and catalytic activities (nicking and end-joining ligation) of Rep are ATP-independent but need two specific bivalent ions: magnesium (Mg^{2+}) and manganese (Mn^{2+}) (Heyraud-nitschke *et al.*, 1995; Laufs *et al.*, 1995; Hafner *et al.*, 1997; Gronenborn, 2004; Jeske, 2009). In nanovirids, the origin of replication is part of the common region stem-loop (CR-SL) sequence shared by all the genome components to enable self-replication of DNA-R encoding M-Rep and trans-replication of other components (Burns *et al.*, 1995). In BBTV (genus *Babuvirus*, family *Nanoviridae*), 5 nt long iterons - two tandem iterons in the forward orientation (called F1 and F2) and a single iteron in the reverse orientation, located respectively downstream and upstream of the stem-loop structure - are essential for trans-replication of viral genome components mediated by M-Rep (Herrera-Valencia *et al.*, 2006). Alphasatellites also possess CR-SL sequences with the conserved nonanucleotide TAGTATTAC that differs from the helper virus nonanucleotide at the 3rd nucleotide position and the iterated sequences that are much divergent of their helper viruses, possibly explaining why alphasatellite Rep is not able to mediated trans-replication of its helper virus components (Londoño *et al.*, 2010; Stainton *et al.*, 2017; Guyot et al. 2022b/Chapter 2 of this thesis). By mediating replication of their own DNA only, alphasatellites can accumulate at high levels exceeding those of any component of the helper virus (Guyot *et al.*, 2022a and 2022b/Chapters 1 and 2 of this thesis).

Besides replication, Rep proteins of geminivirids have been implicated in several other functions such as regulation of viral transcription, suppression of antiviral RNA interference (RNAi)-based antiviral defenses, and interaction with viral coat protein (Ruhel & Chakraborty, 2019). This suggests that alphasatellite Rep can also have similar functions potentially impacting helper virus replication, transcription, evasion of RNAi, encapsidation and transmission by insect vectors.

Diversity of geminialphasatellites and their impact on helper geminivirids

One of the first discovered geminialphasatellites is ageratum yellow vein alphasatellite (AYVA, genus *Ageyesisatellite*, Genbank accession AJ416153) naturally associated with the Old World (OW) monopartite begomovirus ageratum yellow vein virus (AYVV) and transmitted by the whitefly *B. tabaci*. Under the laboratory conditions in the model dicot host *Nicotiana benthamiana* AYVA was able to establish systemic coinfection with the OW bipartite begomovirus african cassava mosaic virus (ACMV) and the New World (NW) bipartite

begomovirus tomato golden mosaic virus (TGMV) as well as beet curly top virus (BCTV) of the genus *Curtovirus*, but not with bean yellow dwarf virus of the genus *Mastrevirus*. In addition, this alphasatellite in a complex with helper BCTV was transmitted from plant to plant (*Beta vulgaris*) by the leafhopper *Circulifer tenellus* (Saunders & Stanley, 1999; Saunders *et al.*, 2002). This indicated that alphasatellites can potentially swap helper viruses within one genus and between genera (*Begomovirus* and *Curtovirus*). Reversely, a single begomovirus can be associated with more than one alphasatellite. For example, the OW monopartite begomovirus alternanthera yellow vein virus was found to be associated with two different alphasatellites (in coinfections with two different betasatellites): hibiscus leaf curl alphasatellite (HLCuA, *Colecusatellite* AM884370-AM9302[44/45/46/48]) and potato leaf curl alphasatellite (PotLCuA), now known as chilli leaf curl Multan alphasatellite (CLCuMA, *Colecusatellite*, FN432360) (Mubin *et al.*, 2010). In earlier reports, geminialphasatellites were found to be naturally associated only with OW monopartite begomoviruses, and always in complex with betasatellites (Saunders *et al.*, 2000; Pei-Jun WU & Xue-Ping ZHOU, 2005; Briddon & Stanley, 2006; Ha *et al.*, 2008; Xie *et al.*, 2010). But later, geminialphasatellites were reported in natural coinfections with NW bipartite begomoviruses in Brazil (*Clecrusatellite*), Venezuela (*Clecrusatellite*), Argentina (*Clecrusatellite*) and Puerto Rico (*Draflysatellite*), in all cases without betasatellites (Paprotka *et al.*, 2010; Romay *et al.*, 2010; Rosario *et al.*, 2013; Bornancini *et al.*, 2020), as well as with an OW bipartite begomovirus and a betasatellite in India (spine gourd mosaic alphasatellite, SGMA, *Colecusatellite*, MG571523) (Venkataravanappa *et al.*, 2019). Alphasatellites can also be naturally associated with OW monopartite begomoviruses without betasatellites. For example, tomato yellow leaf curl virus (TYLCV) was found to be associated with sida yellow vein China alphasatellite (SiYVCNA, *Colecusatellite*, KJ4660[49/50]) in Japan (Shahid *et al.*, 2015). While being able to establish systemic coinfection with a helper curtovirus in the model plant *N. benthamiana* as mentioned above, geminialphasatellites have been found in natural coinfections only with helper viruses of *Begomovirus* and *Mastrevirus* genera (Varsani *et al.*, 2021). It remains to be investigated if geminivirids of other genera are also naturally associated with alphasatellites.

In two reports (Kumar *et al.*, 2014; Hamza *et al.*, 2018) the mastreviruses wheat dwarf India virus (WDIV) and chickpea chlorotic dwarf virus (CpCDV) were found to be associated with the alphasatellites previously discovered in association with OW monopartite begomoviruses, namely cotton leaf curl Multan alphasatellite (CLCuMA, *Colecusatellite*, KC305093 & MF289498) associated with both mastreviruses and guar leaf curl alphasatellite (GLCuA,

Colecusatellite, KC3050[95/96]) associated with WDIV. These findings indicate that geminialphasatellites such as the colecusatellites CLCuMA and CLCuMA not only can naturally swap/jump the helper virus genera - *Begomovirus* versus *Mastrevirus* - but also the host plant families and clades - the dicots cotton (Malvaceae) and guar (Fabaceae) versus the monocot wheat (Poaceae). More recently, nine alphasatellites classified in a single species of a new genus (*Somasatellite*) were reported to be naturally associated with the mastrevirus maize streak virus (MSV) in sorghum (Poaceae) (Claverie *et al.*, 2020; Varsani *et al.*, 2021)

Although the impact of alphasatellites on their helper geminivirids and costs versus benefits of their associations are still debated, recent studies begin to elucidate their selective advantage. Geminialphasatellites appeared to be phenotypically silent in natural infections (Briddon *et al.*, 2018). More advanced studies, however, have revealed that they can have both negative and positive impacts on helper virus symptom development and severity and affect helper virus accumulation in different host plants as well as transmission from plant to plant by insect vectors. For example, Pei-Jun WU & Xue-Ping ZHOU (2005) showed that co-inoculation of the OW monopartite begomovirus tobacco curly shoot virus with tobacco curly shoot alphasatellite (TbCSA, *Colecusatellite*, AJ579349) could reduce viral titer and modulate symptoms in *N. benthamiana*. Kon et al. (2009) showed that cotton leaf curl Gezira alphasatellite (*Colecusatellite*, EU589450) could reduce the accumulation levels of four different OW monopartite begomoviruses in *N. benthamiana*, without attenuating disease symptoms. Idris et al. (2011) showed that ageratum yellow vein Singapore alphasatellite (*Ageyesisatellite*, FJ956707) co-inoculated in *N. benthamiana* either with two different OW monopartite begomoviruses - tomato yellow leaf curl virus Oman strain (TYLCV-OM) or Tomato leaf curl Oman virus (ToLCOMV) - in complex with tomato leaf curl betasatellite (ToLCB) resulted in reduction in betasatellite DNA accumulation and attenuation of disease symptoms, without affecting helper virus DNA accumulation. Luo et al. (2019) showed by transcriptome profiling in *N. benthamiana* infected by the begomovirus/betasatellite complex [OW monopartite tomato yellow leaf curl China virus (TYLCCNV) and tomato yellow leaf curl China betasatellite (TYLCCNB)] that the presence of tomato yellow leaf curl China alphasatellite (TYLCCNA, *Colecusatellite*) reduced the number of differentially expressed genes (DEGs), compared to the condition without alphasatellite. Silencing of six of the DEGs increased both symptom severity and viral DNA accumulation in *N. benthamiana*, suggesting that TYLCCNA attenuates begomovirus-betasatellite complex infection by impacting expression of these genes (Luo *et al.*, 2019). Kumar et al. (2021) showed that cotton leaf curl

Multan alphasatellite (*Colecusatellite*, MG373558) and gossypium darwinii symptomless alphasatellite (*Colecusatellite*, MG373559), both associated with yellow mosaic disease of hollyhock (*Alcea rosea*), could negatively impact their helper begomovirus, the OW monopartite cotton leaf curl Multan virus-India (CLCuMuV-IN): presence of both satellites reduced viral DNA accumulation and delayed and attenuated disease symptoms in *N. benthamiana*. Interestingly, Shafiq et al. (2021) reported that coinoculation of the OW bipartite begomovirus watermelon chlorotic stunt virus (WmCSV) with ageratum enation alphasatellite (AEA, *Colecusatellite*, KF267445 (Akhtar *et al.*, 2014)) slightly affected symptoms and reduced the titres of both DNA-A and DNA-B of the helper virus.

A potentially positive role of geminialphasatellites was first proposed by Nawaz-ul-Rehman et al (2010) who demonstrated a silencing suppressor activity of Rep proteins encoded by gossypium darwinii symptomless alphasatellite (*Colecusatellite*, EU384626) and gossypium davidsonii symptomless alphasatellite (*Gosmusatellite*, EU384653, originally named gossypium mustillinum symptomless alphasatellite). RNA interference (RNAi) and gene silencing play a major role in antiviral defenses, generating of virus-derived 21, 22, and 24 nucleotide (nt) small interfering RNA (siRNAs) that restrict viral gene expression at both transcriptional and post-transcriptional levels (Pooggin *et al.*, 2003; Zvereva & Pooggin, 2012; Pooggin, 2013, 2018). Notably, both alphasatellites exhibited silencing suppression activities in the presence of the OW monopartite begomovirus cotton leaf curl Rajasthan virus (CLCuRaV) and a betasatellite. Additionally, alphasatellite Rep was shown to interact with the helper virus Rep and C4 proteins in yeast two-hybrid assays (Nawaz-ul-Rehman *et al.*, 2010). Abbas et al. (2019) showed that Rep encoded by cotton leaf curl Multan alphasatellite (CLCuMuA, *Colecusatellite*, KY652676) could suppress transcriptional gene silencing (TGS) by restoring expression of a transcriptionally-silent GFP transgene in *N. benthamiana*. Consistent with potential silencing suppression of alphasatellite Rep, euphorbia mosaic alphasatellite (EuMV, *Clecrusatellite*, FN436008) enhanced symptoms severity of euphorbia mosaic virus strain MG1 (EuMV-MG1) (Paprotka *et al.*, 2010). Likewise, Kumar et al. (2014) showed that co-inoculation of wheat plants with the mastrevirus wheat dwarf India virus (WDIV) and cotton leaf curl Multan alphasatellite (CLCuMA, *Colecusatellite*, KC305093) increased viral DNA accumulation and induced more severe symptoms, although the symptoms differed from those observed in the field. In addition, CLCuMA reduced accumulation of WDIV-derived siRNAs, possibly due to its silencing suppressor activity (Kumar *et al.*, 2014). Mar et al. (2017) characterized euphorbia yellow mosaic alphasatellite (EuYMA,

Clecrusatellite, FN436008) naturally associated with the NW bipartite begomovirus euphorbia yellow mosaic virus (EuYMV) in *Euphorbia heterophylla* (Euphorbiaceae) in Brazil (Paprotka *et al.*, 2010). Co-inoculation of this alphasatellite and its helper virus in five different host plant species revealed that the presence of alphasatellite could increase symptom severity in both natural (*E. heterophylla*) and model (*Arabidopsis thaliana* and *N. benthamiana*) hosts, confirming the results of Paprotka et al. (2010) for the natural host. Furthermore, EuYMA affected helper virus DNA accumulation in a host-dependent manner and interfered with helper virus transmission by the whitefly *B. tabaci* MEAM1 (Mar *et al.*, 2017). Interestingly, EuYMA (isolate KX348227) was also detected naturally in *Sida spp.* (Malvaceae) (Ferro *et al.*, 2017), supporting the notion on plasticity of alphasatellites in their host range. In a follow-up study, Nogueira et al. (2021) investigated EuYMA isolate KX348227 and another clecrusatellite tomato yellow spot alphasatellite (ToYSA, KX348228) naturally associated with the NW bipartite begomovirus tomato yellow spot virus (ToYSV) by co-inoculation of three different hosts *N. benthamiana* (Solanaceae), *Solanum lycopersicum* (Solanaceae) and *Leonurus sibiricus* (Lamiaceae) with all combinations of the two alphasatellites and three begomoviruses, including EuYMV, ToYSV and the NW bipartite tomato severe rugose virus (ToSRV). All the combinations were infectious in *N. benthamiana* and *S. lycopersicum*, but only the complex of ToYSV and ToYSA could infect *L. sibiricus*, indicating that alphasatellite can increase the helper virus host range. Depending on alphasatellite-virus combination, both alphasatellites could either increase symptom severity or have no impact on symptoms. With exception for the ToSRV-EuYMA complex in *N. benthamiana* where helper virus DNA accumulation was elevated by alphasatellite, both alphasatellites either decreased or do not alter helper virus DNA loads. Moreover, ToYSA was shown to interfere with transmission of ToYSV by the whitefly *B. tabaci* (Nogueira *et al.*, 2021).

The above-described findings illustrate the complexity of geminialphasatellite-helper virus interactions depending on the host plant and the alphasatellite isolate and genus and show the negative impact alphasatellites on vector transmission. Furthermore, the monopartite vs bipartite nature of helper viruses and the presence of betasatellites can also influence the outcome of alphasatellite-helper virus interactions such as helper virus DNA accumulation and symptom severity. Notably, betasatellite-encoded betaC1 protein is both the pathogenicity determinant and a silencing suppressor (Gnanasekaran *et al.*, 2019).

Besides the impacts on helper viruses described above, geminialphasatellites were demonstrated to depend on helper viruses for systemic movement, suggesting that they need to

be encapsidated by helper virus coat protein for efficient cell-to-cell movement and long distance movement via phloem (Saunders & Stanley, 1999; Saunders *et al.*, 2000; Idris *et al.*, 2011). Amin et al. (2011) showed using blot hybridization that both helper virus and alphasatellite are targets of antiviral RNAi that generates helper virus- and alphasatellite-derived siRNAs. Consistent with this finding, geminialphasatellites could be used as vectors for virus-induced gene silencing (VIGS) (Huang *et al.*, 2011; Shahid *et al.*, 2019). Interestingly, to construct the alphasatellite-based VIGS vector Shahid et al. (2019) deleted the conserved adenine-rich region without affecting alphasatellite replication and movement. Huang et al. (2013) showed that tobacco curly shoot alphasatellite was able to recombine with tomato yellow leaf curl China betasatellite and ageratum yellow vein China betasatellite, but such recombination appeared to be a rare event. Deltasatellites can potentially be a relic of alphasatellite recombination with others genetic elements.

Role of alphasatellites associated with nanovirids

Nanovirids have a multipartite genome comprising 6 (genus *Babuvirus*) to 8 (genus *Nanovirus*) components whose relative abundance or frequency, named genome formula, differs between the host plant and the insect vector as well as between different hosts (Sicard *et al.*, 2013, 2015, 2016, Guyot et al., 2022a and 2022b/Chapters 1 and 2 of this thesis). Moreover, DNA components of the nanovirus genome are not necessarily present in the same phloem cell and are unequally distributed between neighbouring cells possibly for independent gene expression and trans-replication by mobile M-Rep (or M-Rep expressed from mobile Rep-mRNA) (Sicard *et al.*, 2019). Subterranean clover stunt alphasatellite 1 (SCSA1, *Subclovsatellite*, AJ005968) naturally associated with 8-component faba bean necrotic yellows virus (FBNYV, *Nanovirus*) (Katul *et al.*, 1998) was found to accumulate at relatively high levels (12 to 31% of total viral DNA) and to modify FBNYV genome formula in the dicot *Vicia faba* (Fabaceae) (Mansourpour *et al.*, 2022). However, the total viral DNA accumulation in the host plant and symptom servery were not altered by SCSA1. Interestingly, accumulation levels of helper virus and alphasatellite DNA positively correlated in individual host plants. Furthermore, FBNYV genome formula in the aphid vector *Acyrthosiphon pisum*, which differed from that in the host plant, was also modified by SCSA1 and helper virus DNA accumulation in the vector was found to be about 3 time lower in the presence of alphasatellite. Surprisingly, despite the strong reduction of helper

virus DNA loads in aphids, virus transmission rate was increased in the presence of SCSA1 (Mansourpour *et al.*, 2022).

In another study (Guyot et al. 2022a/Chapter 1 of this thesis), banana bunchy top alphasatellite 4 (BBTA4, *Banaphisatellite*, OK546211) associated with BBTV (genus *Babuvirus*) in monocot banana (Musaceae) plants in Africa did not modify BBTV disease symptoms under the laboratory conditions and could be lost during virus transmission from plant to plant. This alphasatellite accumulated at high levels, being the most abundant component in plants and the third most abundant in the aphid vector *Pentalonia nigronervosa*, and thereby reduced helper virus accumulation in the host plants and viruliferous aphids and interfered with virus transmission by aphids. Furthermore, BBTA4 modified BBTV genome formulas in plants and aphids and negatively impacted transcription rate of mRNA-N encoding a putative aphid transmission factor and increased the rates of siRNA production from Rep- and movement protein-encoding components. Alphasatellite itself spawned the most abundant siRNAs and had the lowest transcription rate, suggesting that it can potentially serve as a decoy diverting the antiviral RNAi machinery from the helper virus (Guyot et al. 2022a/Chapter 1 of this thesis). Phylogenetically, BBTA4 is a founder of a new genus *Banaphisatellite* in the subfamily *Nanoalphasatellitinae*, and is more related to alphasatellites infecting dicot plants (genus *Fabenesatellite*) than to the BBTV alphasatellites BBTA1, BBTA2 and BBTA3 previously identified in South-East Asia and classified in the *Babusatellite* (BBTA1) and *Muscarsatellite* (BBTA2 and BBTA3) genera of the subfamily *Petromoalphasatellitinae*. In support of a dicot origin of BBTA4, BBTV and BBTA4 were able to establish systemic infection of the model dicot *N. benthamiana*, followed by recovery and symptomless persistence of alphasatellite in the absence of helper virus, likely due to its ability to self-replicate. Finally, alphasatellite DNA was found to be encapsidated by BBTV coat protein in both banana and *N. benthamiana* (Guyot et al. 2022a/Chapter 1 of this thesis). So far, this study is one of the most comprehensive for any alphasatellite.

Consistent with the findings of Guyot et al. 2022a for the banaphisatellite BBTA4 and the RNAi decoy hypothesis, the babusatellite BBTA1 (MG545617, genus *Babusatellite, Petromoalphasatellitinae*) and BBTA2 (MG545616, *Muscarsatellite Petromoalphasatellitinae*) identified in BBTV-infected banana plants in China were found to accumulate their DNA at high levels, while alphasatellite transcripts were found to be low-abundance compared to BBTV transcripts (Yu *et al.*, 2019). Likewise, the muscarsatellite BBTA2 and BBTA3 and the banaphisatellites BBTA5 and BBTA6 found in single and mixed

infections in wild and cultivated banana plants in Vietnam accumulated their (combined) DNA at high levels and produced abundant siRNAs (Guyot et al. 2022b/Chapter 2 of this thesis).

Coconut foliar decay virus

Coconut foliar decay virus (CFDV, genus *Cofodevirus*) is a single member in the family *Metaxyviridae* (Walker *et al.*, 2021). This phloem-restricted virus is responsible of coconut foliar decay disease of the monocot coconut (*Cocos nucifera*, family Arecaceae) (Randles *et al.*, 1992, 1999) and is transmitted from plant to plant by *Myndus taffini* (*Cixiidae*, Hemiptera) in a persistent circulative manner (Wefels *et al.*, 2015). CFDV is restricted to Vanuatu islands and has a low mutation rate based on sequence analysis of its isolates collected in 1989/90 and 2013, making this virus unique. Gronenborn *et al.* (2018) have reconstructed its multipartite ssDNA genome composition: by extensive deep-sequencing analysis they found two variants of DNA-S (1263 and 1286 nts) both encoding a CP (214 and 217 amino acids, respectively) sharing similarity with the coat protein of grapevine red blotch virus (genus *Grablovirus*, *Geminiviridae*), DNA-gamma, a very abundant viral DNA of 641 nts encoding a putative MP with similarities to babuviral MP; and nine distinct alphasatellites of ca. 1.3 kb. In the absence of nanovirus-like DNA-R, one of the coconut foliar decay alphasatellites (CFDA), now named CFDV DNA-R (MF926434), was postulated to encode M-Rep based on similarity of the sequences flanking its origin of replication with those of DNA-S and DNA-gamma, suggesting that this M-Rep can mediate self-replication of CFDV DNA-R and trans-replication of DNA-S and DNA-gamma (https://talk.ictvonline.org/taxonomy/p/taxonomy-history?taxnode_id=202012153). Other CFDV alphasatellites, were classified in the three new genera of subfamily *Petromoalphasatellitinae*: *Kobbarisatellite* (MF926427, MF926431), *Coprasatellite* (MF926432) and *Cocosatellite* (MF9264[23/24/25/26/29/30/33/44] & M29963) (Varsani *et al.*, 2021). CFDV DNA-R is most closely related to the cocosatellite CFDA2 (isolate MF926426) and other cocosatellites. No specific role was attributed yet to these petromoalphasatellites. Interestingly, CFDV DNA-R and all CFDAs have a size similar to that of geminialphasatellites, while CFDV DNA-gamma is similar in size to geminideltasatellites (Lozano *et al.*, 2016).

The case of CFDV shows that alphasatellite Rep can potentially mediate trans-replication of other DNA components provided the adaptation of their origins of replication through mutations in cis-elements such as the stem-loop structure and iterons responsible for efficient Rep binding and initiation of replication (see Guyot et al. 2022b/Chapter 2 of this thesis).

Hence, by adopting a CP-encoding component, any alphasatellite can potentially evolve into a true virus and move from the perivirosphere to the orthovirosphere. Conversely, a DNA-R component of any multipartite ssDNA virus having lost the CP-encoding component can potentially evolve into an alphasatellite and leave the orthovirosphere for the perivirosphere.

References

Abbas Q, Amin I, Mansoor S, Shafiq M, Wassenegger M, Briddon RW. **2019**. The Rep proteins encoded by alphasatellites restore expression of a transcriptionally silenced green fluorescent protein transgene in Nicotiana benthamiana. *VirusDisease* **30**: 101–105.

Akhtar S, Khan AJ, Singh AS, Briddon RW. **2014**. Identification of a disease complex involving a novel monopartite begomovirus with beta- and alphasatellites associated with okra leaf curl disease in Oman. *Archives of Virology* **159**: 1199–1205.

Alberter B, Ali Rezaian M, Jeske H. **2005**. Replicative intermediates of Tomato leaf curl virus and its satellite DNAs. *Virology* **331**: 441–448.

Amin I, Hussain K, Akbergenov R, Yadav JS, Qazi J, Mansoor S, Hohn T, Fauquet CM, Briddon RW. **2011**. Suppressors of RNA silencing encoded by the components of the cotton leaf curl begomovirus-betasatellite complex. *Molecular Plant-Microbe Interactions* **24**: 973–983.

Argüello-Astorga GR, Guevara-González RG, Herrera-Estrella LR, Rivera-Bustamante RF. **1994**. Geminivirus Replication Origins Have a Group-Specific Organization of Iterative Elements: A Model for Replication. *Virology* **203**: 90–100.

Badar U, Venkataraman S, AbouHaidar M, Hefferon K. **2021**. Molecular interactions of plant viral satellites. *Virus Genes* **57**: 1–22.

Bornancini VA, Irazoqui JM, Flores CR, Vaghi Medina CG, Amadio AF, López Lambertini PM. **2020**. Reconstruction and characterization of full-length begomovirus and alphasatellite genomes infecting pepper through metagenomics. *Viruses* **12**.

Briddon RW, Bull SE, Amin I, Mansoor S, Bedford ID, Rishi N, Siwatch SS, Zafar Y, Abdel-Salam AM, Markham PG. **2004**. Diversity of DNA 1: A satellite-like molecule associated with monopartite begomovirus-DNA β complexes. *Virology* **324**: 462–474.

Briddon RW, Mansoor S, Bedford ID, Pinner MS, Saunders K, Stanley J, Zafar Y, Malik

KA, Markham PG. **2001**. Identification of DNA components required for induction of cotton leaf curl disease. *Virology* **285**: 234–243.

Briddon RW, Martin DP, Roumagnac P, Navas-Castillo J, Fiallo-Olivé E, Moriones E, Lett JM, Zerbini FM, Varsani A. **2018**. Alphasatellitidae: a new family with two subfamilies for the classification of geminivirus- and nanovirus-associated alphasatellites. *Archives of Virology* **163**: 2587–2600.

Briddon RW, Stanley J. **2006**. Subviral agents associated with plant single-stranded DNA viruses. *Virology* **344**: 198–210.

Burns TM, Harding RM, Dale JL. **1995**. The genome organization of banana bunchy top virus: Analysis of six ssDNA components. *Journal of General Virology* **76**: 1471–1482.

Claverie S, Varsani A, Hoareau M, Filloux D, Roumagnac P, Martin DP, Lefeuvre P, Lett JM. **2020**. Sorghum mastrevirus-associated alphasatellites: new geminialphasatellites associated with an African streak mastrevirus infecting wild Poaceae plants on Reunion Island. *Archives of Virology* **165**: 1925–1928.

Cui X, Li G, Wang D, Hu D, Zhou X. **2005**. A Begomovirus DNAβ-Encoded Protein Binds DNA, Functions as a Suppressor of RNA Silencing, and Targets the Cell Nucleus. *Journal of Virology* **79**: 10764–10775.

Ferro CG, Silva JP, Xavier CAD, Godinho MT, Lima ATM, Mar TB, Lau D, Zerbini FM. **2017**. The ever increasing diversity of begomoviruses infecting non-cultivated hosts: new species from Sida spp. and Leonurus sibiricus, plus two New World alphasatellites. *Annals of Applied Biology* **170**: 204–218.

Fiallo-Olivé E, Lett J-M, Martin DP, Roumagnac P, Varsani A, Zerbini FM, Navas-Castillo J. **2021**. ICTV Virus Taxonomy Profile: Geminiviridae 2021. *Journal of General Virology* **102**: 001696.

Fiallo-Olivé E, Tovar R, Navas-Castillo J. **2016**. Deciphering the biology of deltasatellites from the New World: maintenance by New World begomoviruses and whitefly transmission. *New Phytologist* **212**: 680–692.

Francki RI. **1985**. Plant virus satellites. *Annual review of microbiology* **39**: 151–174.

Gnanasekaran P, Chakraborty S. **2018**. Biology of viral satellites and their role in pathogenesis. *Current Opinion in Virology* **33**: 96–105.

Gnanasekaran P, KishoreKumar R, Bhattacharyya D, Vinoth Kumar R, Chakraborty S. **2019**. Multifaceted role of geminivirus associated betasatellite in pathogenesis. *Molecular Plant Pathology* **20**: 1019–1033.

Gronenborn B. **2004**. Nanoviruses: Genome organisation and protein function. In: Veterinary Microbiology. Elsevier, 103–109.

Gronenborn B, Randles JW, Knierim D, Barrière Q, Vetten HJ, Warthmann N, Cornu D, Sileye T, Winter S, Timchenko T. **2018**. Analysis of DNAs associated with coconut foliar decay disease implicates a unique single-stranded DNA virus representing a new taxon. *Scientific Reports* **8**.

Ha C, Coombs S, Revill P, Harding R, Vu M, Dale J. **2008**. Molecular characterization of begomoviruses and DNA satellites from Vietnam: Additional evidence that the New World geminiviruses were present in the Old World prior to continental separation. *Journal of General Virology* **89**: 312–326.

Hafner GJ, Stafford MR, Wolter LC, Harding RM, Dale JL. **1997**. Nicking and joining activity of banana bunchy top virus replication protein in vitro. *Journal of General Virology* **78**: 1795–1799.

Hamza M, Tahir MN, Mustafa R, Kamal H, Khan MZ, Mansoor S, Briddon RW, Amin I. **2018**. Identification of a dicot infecting mastrevirus along with alpha- and betasatellite associated with leaf curl disease of spinach (Spinacia oleracea) in Pakistan. *Virus Research* **256**: 174–182.

Herrera-Valencia VA, Dugdale B, Harding RM, Dale JL. **2006**. An iterated sequence in the genome of Banana bunchy top virus is essential for efficient replication. *Journal of General Virology* **87**: 3409–3412.

Heyraud-nitschke F, Schumacher S, Laufs J, Schaefer S, Schell J, Gronenborn B. **1995**. Determination of the origin cleavage and joining domain of geminivirus rep proteins. *Nucleic Acids Research* **23**: 910–916.

Horser CL, Harding RM, Dale JL. **2001**. Banana bunchy top nanovirus DNA-1 encodes the 'master' replication initiation protein. *Journal of General Virology* **82**: 459–464.

Hu CC, Hsu YH, Lin NS. **2009**. Satellite RNAs and satellite viruses of plants. *Viruses* **1**: 1325–1350.

Huang C, Xie Y, Zhao L, Ren H, Li Z. **2013**. A naturally occurring defective DNA satellite associated with a monopartite begomovirus: Evidence for recombination between alphasatellite and betasatellite. *Viruses* **5**: 2116–2128.

Huang CJ, Zhang T, Li FF, Zhang XY, Zhou XP. **2011**. Development and application of an efficient virus-induced gene silencing system in Nicotiana tabacum using geminivirus alphasatellite. *Journal of Zhejiang University: Science B* **12**: 83–92.

Idris AM, Shahid MS, Briddon RW, Khan AJ, Zhu JK, Brown JK. **2011**. An unusual alphasatellite associated with monopartite begomoviruses attenuates symptoms and reduces betasatellite accumulation. *Journal of General Virology* **92**: 706–717.

Ilyina T V., Koonin E V. 1992. Conserved sequence motifs in the initiator proteins for rolling circle DNA replication encoded by diverse replicons from eubacteria, eucaryotes and archaebacteria. *Nucleic Acids Research* **20**: 3279–3285.

Iqbal Z, Shafiq M, Ali I, Mansoor S, Briddon RW. **2017**. Maintenance of cotton leaf curl multan betasatellite by tomato leaf curl New Delhi virus—Analysis by mutation. *Frontiers in Plant Science* **8**: 2208.

Jeske H. **2009**. Geminiviruses. *Current Topics in Microbiology and Immunology* **331**: 185–226.

Karthikeyan C, Patil BL, Borah BK, Resmi TR, Turco S, Pooggin MM, Hohn T, Veluthambi K. **2016**. Emergence of a latent indian cassava mosaic virus from cassava which recovered from infection by a non-persistent Sri Lankan cassava mosaic virus. *Viruses* **8**: 264.

Katul L, Timchenko T, Gronenborn B, Vetten HJ. **1998**. Ten distinct circular ssDNA components, four of which encode putative replication-associated proteins, are associated with the faba bean necrotic yellows virus genome. *Journal of General Virology* **79**: 3101–3109.

King A, Lefkowitz E, Adams M, Carstens E. **2011**. Virus taxonomy: ninth report of the International Committee on Taxonomy of Viruses.

Kon T, Rojas MR, Abdourhamane IK, Gilbertson RL. **2009**. Roles and interactions of begomoviruses and satellite DNAs associated with okra leaf curl disease in Mali, West Africa. *Journal of General Virology* **90**: 1001–1013.

Koonin E V., Dolja V V., Krupovic M, Kuhn JH. **2021**. Viruses Defined by the Position of the Virosphere within the Replicator Space. *Microbiology and Molecular Biology Reviews*.

Krupovic M, Dolja V V., Koonin E V. 2020a. The LUCA and its complex virome. *Nature Reviews Microbiology* **18**: 661–670.

Krupovic M, Kuhn JH, Fischer MG. **2016**. A classification system for virophages and satellite viruses. *Archives of Virology* **161**: 233–247.

Krupovic M, Varsani A, Kazlauskas D, Breitbart M, Delwart E, Rosario K, Yutin N, Wolf YI, Harrach B, Zerbini FM, *et al.* 2020b. Cressdnaviricota : a Virus Phylum Unifying Seven Families of Rep-Encoding Viruses with Single-Stranded, Circular DNA Genomes. *Journal of Virology* **94**.

Kumar J, Kumar J, Singh SP, Tuli R. **2014**. Association of Satellites with a Mastrevirus in Natural Infection: Complexity of Wheat Dwarf India Virus Disease. *Journal of Virology* **88**: 7093–7104.

Kumar M, Zarreen F, Chakraborty S. **2021**. Roles of two distinct alphasatellites modulating geminivirus pathogenesis. *Virology Journal* **18**: 249.

Laufs J, Traut W, Heyraud F, Matzeit V, Rogers SG, Schell J, Gronenborn B. **1995**. In vitro cleavage and joining at the viral origin of replication by the replication initiator protein of tomato yellow leaf curl virus. *Proceedings of the National Academy of Sciences of the United States of America* **92**: 3879–3883.

Londoño A, Riego-Ruiz L, Argüello-Astorga GR. **2010**. DNA-binding specificity determinants of replication proteins encoded by eukaryotic ssDNA viruses are adjacent to widely separated RCR conserved motifs. *Archives of Virology* **155**: 1033–1046.

Lozano G, Trenado HP, Fiallo-Olivé E, Chirinos D, Geraud-Pouey F, Briddon RW, Navas-Castillo J. **2016**. Characterization of non-coding DNA satellites associated with sweepoviruses (Genus Begomovirus, Geminiviridae) - Definition of a distinct class of begomovirus-associated satellites. *Frontiers in Microbiology* **7**.

Luo C, Wang ZQ, Liu X, Zhao L, Zhou X, Xie Y. **2019**. Identification and analysis of potential genes regulated by an alphasatellite (TYLCCNA) that contribute to host resistance against tomato yellow leaf curl China virus and its betasatellite (TYLCCNV/TYLCCNB) infection in Nicotiana benthamiana. *Viruses* **11**: 442.

Mansoor S, Amin I, Hussain M, Zafar Y, Bull S, Briddon RW, Markham PG. **2001**. Association of a Disease Complex Involving a Begomovirus, DNA 1 and a Distinct DNA Beta

with Leaf Curl Disease of Okra in Pakistan. *Plant Disease* **85**: 922–922.

Mansoor S, Khan SH, Bashir A, Saeed M, Zafar Y, Malik KA, Briddon R, Stanley J, Markham PG. **1999**. Identification of a novel circular single-stranded DNA associated with cotton leaf curl disease in Pakistan. *Virology* **259**: 190–199.

Mansourpour M, Gallet R, Abbasi A, Blanc S, Dizadji A, Zeddam J-L. **2022**. Effects of an Alphasatellite on the Life Cycle of the Nanovirus Faba Bean Necrotic Yellows Virus. *Journal of Virology* **96**.

Mar TB, Mendes IR, Lau D, Fiallo-Olivé E, Navas-Castillo J, Alves MS, Zerbini FM. **2017**. Interaction between the new world begomovirus Euphorbia yellow mosaic virus and its associated alphasatellite: Effects on infection and transmission by the whitefly Bemisia tabaci. *Journal of General Virology* **98**: 1552–1562.

Mubin M, Ijaz S, Nahid N, Hassan M, Younus A, Qazi J, Nawaz-ul-Rehman MS. **2020**. Journey of begomovirus betasatellite molecules: from satellites to indispensable partners. *Virus Genes* **56**: 16–26.

Mubin M, Shahid MS, Tahir MN, Briddon RW, Mansoor S. **2010**. Characterization of begomovirus components from a weed suggests that begomoviruses may associate with multiple distinct DNA satellites. *Virus Genes* **40**: 452–457.

Muhire B, Martin DP, Brown JK, Navas-Castillo J, Moriones E, Zerbini FM, Rivera-Bustamante R, Malathi VG, Briddon RW, Varsani A. **2013**. A genome-wide pairwise-identity-based proposal for the classification of viruses in the genus Mastrevirus (family Geminiviridae). *Archives of Virology* **158**: 1411–1424.

Nawaz-ul-Rehman MS, Nahid N, Mansoor S, Briddon RW, Fauquet CM. **2010**. Post-transcriptional gene silencing suppressor activity of two non-pathogenic alphasatellites associated with a begomovirus. *Virology* **405**: 300–308.

Nogueira AM, Nascimento MB, Barbosa TMC, Quadros AFF, Gomes JPA, Orílio AF, Barros DR, Zerbini FM. **2021**. The association between new world alphasatellites and bipartite begomoviruses: Effects on infection and vector transmission. *Pathogens* **10**.

Paprotka T, Metzler V, Jeske H. **2010**. The first DNA 1-like α satellites in association with New World begomoviruses in natural infections. *Virology* **404**: 148–157.

Pei-Jun WU, Xue-Ping ZHOU. **2005**. Interaction between a Nanovirus-like Component and

the Tobacco Curly Shoot Virus/Satellite Complex. *Acta Biochimica et Biophysica Sinica* **37**: 25–31.

Pooggin MM. **2013**. How can plant DNA viruses evade siRNA-directed DNA methylation and silencing? *International Journal of Molecular Sciences* **14**: 15233–15259.

Pooggin MM. **2018**. Small RNA-omics for plant virus identification, virome reconstruction, and antiviral defense characterization. *Frontiers in Microbiology* **9**: 2779.

Pooggin M, Shivaprasad P V., Veluthambi K, Hohn T. **2003**. RNAi targeting of DNA virus in plants. *Nature Biotechnology* **21**: 131–132.

Randles, W. J, MILLER DC, MORIN JP, ROHDE W, HANOLD D. **1992**. Localisation of coconut foliar decay virus in coconut palm. *Annals of Applied Biology* **121**: 601–617.

Randles JW, Wefels E, Hanold D, Miller DC, Morin JP, Rohde W. **1999**. Detection and diagnosis of coconut foliar decay disease. In: Springer, Dordrecht, 247–258.

Rohde W, Randles JW, Langridge P, Hanold D. **1990**. Nucleotide sequence of a circular single-stranded DNA associated with coconut foliar decay virus. *Virology* **176**: 648–651.

Romay G, Chirinos D, Geraud-Pouey F, Desbiez C. **2010**. Association of an atypical alphasatellite with a bipartite New World begomovirus. *Archives of Virology* **155**: 1843–1847.

Rosario K, Padilla-Rodriguez M, Kraberger S, Stainton D, Martin DP, Breitbart M, Varsani A. **2013**. Discovery of a novel mastrevirus and alphasatellite-like circular DNA in dragonflies (Epiprocta) from Puerto Rico. *Virus Research* **171**: 231–237.

Roumagnac P, Lett JM, Fiallo-Olivé E, Navas-Castillo J, Zerbini FM, Martin DP, Varsani A. **2022**. Establishment of five new genera in the family Geminiviridae: Citlodavirus, Maldovirus, Mulcrilevirus, Opunvirus, and Topilevirus. *Archives of Virology* **167**: 695–710.

Ruhel R, Chakraborty S. **2019**. Multifunctional roles of geminivirus encoded replication initiator protein. *VirusDisease* **30**: 66–73.

Saunders K, Bedford ID, Briddon RW, Markham PG, Wong SM, Stanley J. **2000**. A unique virus complex causes Ageratum yellow vein disease. *Proceedings of the National Academy of Sciences of the United States of America* **97**: 6890–6895.

Saunders K, Bedford ID, Stanley J. **2002**. Adaptation from whitefly to leafhopper transmission of an autonomously replicating nanovirus-like DNA component associated with

ageratum yellow vein disease. *Journal of General Virology* **83**: 907–913.

Saunders K, Stanley J. **1999**. A nanovirus-like DNA component associated with yellow vein disease of Ageratum conyzoides: Evidence for interfamilial recombination between plant DNA viruses. *Virology* **264**: 142–152.

Shafiq M, Sattar MN, Shahid MS, Al-Sadi AM, Briddon RW. **2021**. Interaction of watermelon chlorotic stunt virus with satellites. *Australasian Plant Pathology* **50**: 117–128.

Shahid MS, Ikegami M, Briddon RW, Natsuaki KT. **2015**. Characterization of Tomato yellow leaf curl virus and associated alphasatellite infecting Cucurbita maxima in Japan. *Journal of General Plant Pathology* **81**: 92–95.

Shahid MS, Mansoor S, Briddon RW. **2019**. Use of the cotton leaf curl Multan alphasatellite as a silencing or expression vector. *Acta Virologica* **63**: 36–44.

Sicard A, Michalakis Y, Gutiérrez S, Blanc S. **2016**. The Strange Lifestyle of Multipartite Viruses. *PLoS Pathogens* **12**: e1005819.

Sicard A, Pirolles E, Gallet R, Vernerey MS, Yvon M, Urbino C, Peterschmitt M, Gutierrez S, Michalakis Y, Blanc S. **2019**. A multicellular way of life for a multipartite virus. *eLife* **8**.

Sicard A, Yvon M, Timchenko T, Gronenborn B, Michalakis Y, Gutierrez S, Blanc S. **2013**. Gene copy number is differentially regulated in a multipartite virus. *Nature Communications* **4**: 1–8.

Sicard A, Zeddam J-L, Yvon M, Michalakis Y, Gutiérrez S, Blanc S. **2015**. Circulative Nonpropagative Aphid Transmission of Nanoviruses: an Oversimplified View. *Journal of Virology* **89**: 9719–9726.

Stainton D, Martin DP, Collings DA, Varsani A. **2017**. Comparative analysis of common regions found in babuviruses and alphasatellite molecules. *Archives of Virology* **162**: 849–855.

Thomas JE, Gronenborn B, Harding RM, Mandal B, Grigoras I, Randles JW, Sano Y, Timchenko T, Vetten HJ, Yeh HH, *et al.* 2021. ICTV virus taxonomy profile: Nanoviridae. *Journal of General Virology* **102**: 001544.

Timchenko T, Katul L, Sano Y, De Kouchkovsky F, Vetten HJ, Gronenborn B. **2000**. The master rep concept in nanovirus replication: Identification of missing genome components and

potential for natural genetic reassortment. *Virology* **274**: 189–195.

Timchenko T, de Kouchkovsky F, Katul L, David C, Vetten HJ, Gronenborn B. **1999**. A Single Rep Protein Initiates Replication of Multiple Genome Components of Faba Bean Necrotic Yellows Virus, a Single-Stranded DNA Virus of Plants. *Journal of Virology* **73**: 10173–10182.

Varma A, Malathi VG. **2003**. Emerging geminivirus problems: A serious threat to crop production. *Annals of Applied Biology* **142**: 145–164.

Varsani A, Martin DP, Randles JW, Vetten HJ, Thomas JE, Fiallo-Olivé E, Navas-Castillo J, Lett JM, Zerbini FM, Roumagnac P, *et al.* 2021. Taxonomy update for the family Alphasatellitidae: new subfamily, genera, and species. *Archives of Virology* **1**: 1–9.

Venkataravanappa V, Lakshminarayana Reddy CN, Shankarappa KS, Reddy MK. **2019**. Association of tomato leaf curl new delhi virus, betasatellite, and alphasatellite with mosaic disease of spine gourd (Momordica dioica roxb. willd) in india. *Iranian Journal of Biotechnology* **17**: 17–29.

Walker PJ, Siddell SG, Lefkowitz EJ, Mushegian AR, Adriaenssens EM, Alfenas-Zerbini P, Davison AJ, Dempsey DM, Dutilh BE, García ML, *et al.* 2021. Changes to virus taxonomy and to the International Code of Virus Classification and Nomenclature ratified by the International Committee on Taxonomy of Viruses (2021). *Archives of Virology* **166**: 2633–2648.

Wang XW, Blanc S. **2021**. Insect Transmission of Plant Single-Stranded DNA Viruses. *Annual Review of Entomology* **66**: 389–405.

Wefels E, Morin JP, Randles JW. **2015**. Molecular evidence for a persistent-circulative association between Coconut foliar decay virus and its vector Myndus taffini. *Australasian Plant Pathology* **44**: 283–288.

Wu Rey Yuh, You Li Ru, Soong Tai Sen. **1994**. Nucleotide sequences of two circular single-stranded DNAs associated with banana bunchy top virus. *Phytopathology* **84**: 952–958.

Xie Y, Wu P, Liu P, Gong H, Zhou X. **2010**. Characterization of alphasatellites associated with monopartite begomovirus/betasatellite complexes in Yunnan, China. *Virology Journal* **7**.

Yeh HH, Su HJ, Chao YC. **1994**. Genome characterization and identification of viral-associated dsDNA component of banana bunchy top virus. *Virology* **198**: 645–652.

Yu NT, Xie HM, Zhang YL, Wang JH, Xiong Z, Liu ZX. **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.

Zhao L, Rosario K, Breitbart M, Duffy S. **2019**. Eukaryotic Circular Rep-Encoding Single-Stranded DNA (CRESS DNA) Viruses: Ubiquitous Viruses With Small Genomes and a Diverse Host Range. In: Advances in Virus Research. Adv Virus Res, 71–133.

Zhou X. **2013**. Advances in understanding begomovirus satellites. *Annual Review of Phytopathology* **51**: 357–381.

Zvereva AS, Pooggin MM. **2012**. Silencing and innate immunity in plant defense against viral and non-viral pathogens. *Viruses* **4**: 2578–2597.

Discussion and Perspectives

In this PhD thesis, we describe the discovery and in depth molecular and biological characterization of three novel species of BBTV alphasatellites (BBTA) that we classified into a new genus of the subfamily *Nanoalphasatellitinae* of family *Alphasatellitidae*, and proposed to name *Banaphisatellite* (Chapters 1 and 2). These banaphisatellites were found to be associated with BBTV isolates from Africa (BBTA4) and South-East Asia (BBTA5 and BBTA6), which belong to the distinct phylogenetic groups (PIO and SEA, respectively) and, based on our comparative sequence analysis, represent distinct strains of BBTV. These strains differ in *cis*-elements of the common regions CR-SL and CR-M implicated in Rep-mediated replication of viral DNA and in coding capacities of DNA-U3 (Chapter 2). Sequence similarity-based clustering analysis of banaphisatellite Rep proteins showed their close evolutionary links to Rep proteins encoded by alphasatellites from both the genus *Fabenesatellite* of subfamily *Nanoalphasatellitinae* and the genus *Gosmusatellite* of subfamily *Geminialphasatellitinae*, both infecting dicot plants. In contrast, banaphisatellite Rep proteins have only indirect (distant) links to Rep proteins encoded by BBTV alphasatellites of the *Babusatellite* (BBTA1) and *Muscarsatellite* (BBTA2 and BBTA3) genera previously identified in banana plants in South-East Asia and classified into the subfamily *Petromoalphasatellitinae* that comprises alphasatellites infecting perennial monocots such as bananas, cardamom and cocoa (Chapters 1 and 2). These findings let us hypothesize that banaphisatellites originate from dicot host(s) and, more generally, that alphasatellites in their evolution may swap/jump the helper viruses' genera (*Nanovirus* for fabenesatellites to *Babuvirus* for banaphisatellites) and even families (*Geminiviridae* for gosmusatellites to *Nanoviridae* for banaphisatellites) as well as the host plants' families (Fabaceae for fabenesatellites to Musaceae for banaphisatellites) and even clades (dicot to monocot). In support of this hypothesis we found that, under laboratory conditions, cloned banaphisatellite BBTA4 and its helper BBTV could establish systemic infection in the model dicot *N. benthamiana* (Solanaceae), followed by recovery and symptomless persistence of alphasatellite likely due to its ability to self-replicate (Chapter 1). Moreover, our survey of non-banana plants grown in close vicinity of BBTV and alphasatellite-infected banana plants in South-East Asia revealed that *Bidens pilosa* and *Chromolaena odorata*, both from the dicot family Asteraceae, host BBTV and the banaphisatellites BBTA5 and BBTA6 as well as the muscarsatellites BBTA2 and BBTA3, further supporting the dicot origin of banaphisatellites and indicating the ability of the muscarsatellites infecting monocot bananas (Musaceae) to jump to dicots (Asteraceae) (Chapter 3). It remains to be investigated

through further surveys in Democratic Republic of the Congo (DRC), where BBTA4 was discovered, and other BBTV-affected regions in Africa and elsewhere in PIO, whether or not dicot plants co-cultivated or naturally grown nearby banana plants can host BBTV and alphasatellites or other ssDNA viruses associated with alphasatellites that can potentially jump to bananas. Dhed'a (2019) reported that bananas are often cultivated with other agronomically important plants in DRC, and specified that high-altitude bananas are co-cultivated with plant species from the family Fabaceae. The surveys in South-East Asia (Chapters 2 and 3) primarily aimed at exploring the genetic diversity of wild banana (*Musa*) species in forests and other non-cultivated areas, where dicot and monocot plants grow heterogeneously, making the host jumps of viruses and satellites conceivable. Our findings that Asteraceae plant species host BBTV and its alphasatellites should make this plant family one of the most promising for future surveys in Africa and elsewhere. As an alternative to the candidate family approach, a large-scale non-targeted RCA metagenomics surveys could also be conducted for discovery of new hosts of known alphasatellites as well as unknown alphasatellites.

Following our evidence for the model dicot *N. benthamiana* to be a host for BBTV and BBTA4, further experimental validation of the host plant (and helper virus) jumps of monocot-infecting alphasatellites using dicot plants will be required to understand the molecular mechanisms underlying adaptation of monocot alphasatellites to dicots and their ssDNA viruses (and vice versa). For examples the infectious clone of banaphisatellite BBTA4 (and its modified derivatives) can be used for co-inoculation experiments with the available infectious clones of faba bean necrotic stunt virus and faba bean necrotic yellow virus (both from the genus *Nanovirus*, family *Nanoviridae*) in their dicot Fabaceae hosts, and of tomato yellow leaf curl virus (genus *Begomovirus*, family *Geminiviridae*) in its dicot Solanaceae hosts. These experiments should take into account size constraints for alphasatellite encapsidation by helper virus coat protein and other sequence features (putative packaging signals) required for encapsidation in virions of different geometry (isometric in *Nanoviridae* and twinned in *Geminiviridae*) as well as *cis*-acting sequence elements implicated in replication and transcription of alphasatellite and helper virus DNA (see Chapters 1 and 2). Experiments on aphid transmission of a monocot alphasatellite-dicot helper virus complex (if established) would then be required for final validation of the helper virus and host plant jumps. Our preliminary co-inoculation efforts using cloned BBTA4 and faba bean necrotic yellows virus in *Vicia faba* (Fabaceae) did not result in systemic movement of alphasatellite DNA. On the

other hand, using an antiviral RNAi-deficient *dcl234* mutant line of the model dicot *Arabidopsis thaliana* (Brassicaceae), we were able to detect BBTA4 DNA in leaf tissues, following short-term inoculation with viruliferous banana aphids. However, banana aphids did not survive on *A. thaliana* plants for more than 2-3 days, making this model system inapplicable for transmission tests with banana aphids. Nonetheless, the Arabidopsis system can be useful for establishing a complex of BBTA4 (or other alphasatellites) with geminiviruses such as cabbage leaf curl virus (genus *Begomovirus*) transmitted by the whitefly *Bemisia tabaci*.

A common role (if any) of alphasatellites in their co-evolution of helper viruses and, in particular, costs versus benefits of their association with helper viruses remain to be further investigated. Our findings for BBTA4 under the laboratory conditions revealed obvious costs for the helper BBTV such as alphasatellite-mediated interference with virus replication *in planta* and virus transmission by banana aphids. However, our limited surveys in DRC showed persistent association of this alphasatellite with helper BBTV over more than 4 years under the field conditions (Chapter 1). Under our controlled conditions, BBTA4 reduced helper virus accumulation in young seedlings of *M. acuminata* cultivar Cavendish (triploid AAA) and thereby interfered with aphid transmission. But in the wild nature with bigger banana plants representing various intra- and interspecific hybrids with different ploidies and grown under a perpetual flow of biotic and abiotic stress (see Introduction), BBTA4 and other alphasatellites can be beneficial for their helper BBTV. In fact, our surveys in South-Est Asia showed high prevalence of BBTV alphasatellites in wild nature (Chapters 2 and 3). Timchenko et al. (2006) have proposed that alphasatellites could reduce virulence of nanoviruses and thereby prolong the lifespan of infected plants and increase the chance for an aphid vector to find an infected plant and acquire the virus. It can be further proposed that, in turn, self-replicating alphasatellites reducing helper virus virulence and loads and potentially persisting in plants recovered from virus infection can also "wait" for the arrival of a viruliferous vector carrying another helper virus or for a virus-free vector to be transmitted to a new plant infected with the same or another helper virus. In support of this hypothesis we detected BBTA4 DNA following recovery of *N. benthamiana* plants from BBTV infection (Chapter 1). Another benefit for helper viruses would be the ability of alphasatellites to accumulate at high levels and serve as a decoy diverting the antiviral RNAi and gene silencing machinery from actively-transcribed helper virus. Our Illumina sequencing analysis of the viral transcriptome and sRNA-ome profiles in BBTV-infected banana plants in the presence and absence of BBTA4 support the decoy

hypothesis in that the highly-abundant alphasatellite DNA was found to be the least transcriptionally active virome component that generates the most abundant viral siRNAs of 21-22 and 24 nt classes implicated in both transcriptional and post-transcriptional gene silencing (Chapter 1).

As mentioned in our review manuscript on alphasatellites (Chapter 4) alphasatellites can recombine with other virome components. Interestingly, Fu et al. (2009) discovered the chimeras BBTV alphasatellite and DNA-U3, which were also observed in our samples of BBTV- and alphasatellite-infected banana plants (data not shown). These chimeras have alphasatellite-derived CR-SL and the remaining sequence (including ORF) from DNA-U3. It is thus conceivable that an alphasatellite could give rise to a new multipartite ssDNA virus by recombining with and thereby picking up coat protein (CP)- and movement protein (MP)-coding sequences from other viruses. It can be proposed that such recombination events could have given rise to cocoa foliar decay virus (genus *Cofodevirus*, family *Metaxyviridae*) whose M-Rep-encoding component is a *bona fide* alphasatellite, most closely related cocosatellite CFDA2 (genus *Cocosatellite*, subfamily *Petromoalphasatellitinae*) and whose MP- and CP-encoding components share with the M-Rep component the highly similar CR-SL sequences with identical iterons (Chapter 4). This hypothesis remains to be validated experimentally. However, construction of infectious clones of multipartite ssDNA viruses and their delivery to the host plant cells is a difficult endeavour. Our efforts to deliver the infectious clones of BBTV and BBTA4 (showing infectivity in *N. benthamiana*) to *Musa acuminata* Cavendish were not successful so far (data not shown).

Besides establishing evolutionarily links of banaphisatellites with fabene- and gosmusatellites, our clustering analysis of alphasatellite Rep proteins (see Chapter 2) let us to propose changes in classification of other genera in the family *Alphasatellitidae*. Thus, at the p-value ≤ 1E-72 that represents a genus demarcation threshold the genus *Subclovsatellite* is divided into two distinct groups which should be considered as two separate genera (Chapter 2 – Figure 2B). Note, however, at this threshold the banaphisatellite BBTA4 (but not other banaphisatellites) have links with some fabenesatellites, highlighting their close relatedness. The division of subclovsatellites into the two subgroups is also evident on the phylogenetic trees (Chapter 2 – Figure 1; Briddon *et al.*, 2018; Varsani *et al.*, 2021). We propose to split the

genus *Subclovsatellite* into two genera based on both the phylogenetic analysis of nucleotide sequences and clustering analysis of Rep proteins. Furthermore, at the more stringent threshold (P ≤ 1E-81), which represents the species demarcation threshold for banaphisatellites, two member of the genus *Kobbarisatellite* (both associated with cocoa foliar decay virus) are linked with members of the genus *Muscarsatellite* (associated with BBTV) (Chapter 2 – Figure 2A). We therefore propose to combine kobbarisatellites and muscarsatellites in one genus. This proposal is also supported by the phylogenetic tree cladding the two genera together (Chapter 2 – Figure 1). Likewise, based on the phylogeny and Rep clustering, we propose to combine in one genus the alphasatellites of *Coprasatellite* and *Cocosatellite* genera of subfamily *Petromoalphasatellitinae* (both associated with coconut foliar decay virus) and also in one genus the alphasatellites of *Whiflysatellite* and *Colecusatellite* genera of subfamily *Geminialphasatellitinae*. Interestingly, at the least stringent p-value applied in our analysis (≤ 1E-54, Figure 2D), the *Geminialphasatellitinae* genera are still separated into two unlinked subgroups, one comprising *Clecrusatellite*, *Somasatellite* and *Ageyesatellite* genera, and another comprising *Gosmusatellite*, *Colecusatellite*, *Whiflysatellite* and *Draflysatellite* genera, with the latter group having an indirect link with *Banaphisatellite* and *Fabenesatellite* via the genus *Gosmusatellite*. As the phylogeny is also supporting a clear separation of these two subgroups (Chapter 2 - Figure 1), the subfamily *Geminialphasatellitinae* could be split in two subfamilies. Likewise, the subfamily *Petromoalphasatellitinae* could be split in two subfamilies, one comprising the *Coprasatellite* and *Cocosatellite* genera (combined in one genus as we propose) and another grouping the genus *Babusatellite* and the genera *Kobbarisatellite* and *Muscarsatellite* (also combined in one genus) (Chapter 2 - Figures 1 and 2B). In conclusion, the combination of comparative nucleotide sequence phylogenetic and Rep protein clustering analyses of the family *Alphasatellitidae* should be applied to refine its current taxonomy and further understand the origin and evolution of alphasatellites.

Multipartite ssDNA viruses such as BBTV are known to cause tremendous damage to agronomical plants. Further investigation of these viruses and their alphasatellites can contribute to designing new biotech methods for control of the viral diseases at both infection and transmission steps. Self-replicating alphasatellites reducing the load and transmission rate of a helper virus could be used as a tool for disease control, especially at an early stage of viral infection.

References of Introduction & Discussion

Abbas Q, Amin I, Mansoor S, Shafiq M, Wassenegger M, Briddon RW. **2019**. The Rep proteins encoded by alphasatellites restore expression of a transcriptionally silenced green fluorescent protein transgene in *Nicotiana benthamiana*. *VirusDisease* **30**: 101–105.

Akbergenov R, Si-Ammour A, Blevins T, Amin I, Kutter C, Vanderschuren H, Zhang P, Gruissem W, Meins F, Hohn T, *et al.* 2006. Molecular characterization of geminivirus-derived small RNAs in different plant species. *Nucleic Acids Research* **34**: 462–471.

Alberter B, Ali Rezaian M, Jeske H. **2005**. Replicative intermediates of Tomato leaf curl virus and its satellite DNAs. *Virology* **331**: 441–448.

Amin I, Ilyas M, Qazi J, Bashir R, Yadav JS, Mansoor S, Fauquet CM, Briddon RW. **2011**. Identification of a major pathogenicity determinant and suppressors of RNA silencing encoded by a South Pacific isolate of Banana bunchy top virus originating from Pakistan. *Virus Genes* **42**: 272–281.

Anhalt MD, Almeida RPP. **2008**. Effect of temperature, vector life stage, and plant access period on transmission of Banana bunchy top virus to banana. *Phytopathology* **98**: 743–748.

Aregger M, Borah BK, Seguin J, Rajeswaran R, Gubaeva EG, Zvereva AS, Windels D, Vazquez F, Blevins T, Farinelli L, *et al.* 2012. Primary and Secondary siRNAs in Geminivirus-induced Gene Silencing (S-W Ding, Ed.). *PLoS Pathogens* **8**: e1002941.

Argent GCG. **1976**. The wild bananas of Papua New Guinea. *Notes from the Royal Botanic Garden Edinburgh* **35**: 77–114.

Argüello-Astorga GR, Guevara-González RG, Herrera-Estrella LR, Rivera-Bustamante RF. **1994**. Geminivirus Replication Origins Have a Group-Specific Organization of Iterative Elements: A Model for Replication. *Virology* **203**: 90–100.

Bagariang W, Hidayat P, Hidayat SH. **2019**. Morphometric Analysis and Host Range of the Genus *Pentalonia* Coquerel (Hemiptera: *Aphididae*) Infesting Banana in Java. *Jurnal Perlindungan Tanaman Indonesia* **23**: 171.

Banerjee A, Roy S, Behere GT, Roy SS, Dutta SK, Ngachan S V. 2014. Identification and characterization of a distinct banana bunchy top virus isolate of Pacific-Indian Oceans group from North-East India. *Virus Research* **183**: 41–49.

Beetham PR, Hafner GJ, Harding RM, Dale JL. **1997**. Two mRNAs are transcribed from banana bunchy top virus DNA-1. *Journal of General Virology* **78**: 229–236.

Beetham PR, Harding RM, Dale JL. **1999**. Banana bunchy top virus DNA-2 to 6 are monocistronic. *Archives of Virology* **144**: 89–105.

Blackman RL, Eastop VF. **1994**. Aphids on the world's trees: an identification and information guide. *Aphids on the world's trees: an identification and information guide*: 466.

Blanc S, Michalakis Y. **2016**. Manipulation of hosts and vectors by plant viruses and impact of the environment. *Current Opinion in Insect Science* **16**: 36–43.

Blevins T, Rajeswaran R, Aregger M, Borah BK, Schepetilnikov M, Baerlocher L, Farinelli L, Meins F, Hohn T, Pooggin MM. **2011**. Massive production of small RNAs from a non-coding region of Cauliflower mosaic virus in plant defense and viral counter-defense. *Nucleic Acids Research* **39**: 5003–5014.

Blevins T, Rajeswaran R, Shivaprasad P V., Beknazariants D, Si-Ammour A, Park HS, Vazquez F, Robertson D, Meins F, Hohn T, *et al.* 2006. Four plant Dicers mediate viral small RNA biogenesis and DNA virus induced silencing. *Nucleic Acids Research* **34**: 6233–6246.

Borges F, Martienssen RA. **2015**. The expanding world of small RNAs in plants. *Nature Reviews Molecular Cell Biology* **16**: 727–741.

Briddon RW, Martin DP, Roumagnac P, Navas-Castillo J, Fiallo-Olivé E, Moriones E, Lett JM, Zerbini FM, Varsani A. **2018**. *Alphasatellitidae*: a new family with two subfamilies for the classification of geminivirus- and nanovirus-associated alphasatellites. *Archives of Virology* **163**: 2587–2600.

Briddon RW, Stanley J. **2006**. Subviral agents associated with plant single-stranded DNA viruses. *Virology* **344**: 198–210.

Burns TM, Harding RM, Dale JL. **1994**. Evidence that banana bunchy top virus has a multiple component genome. *Archives of Virology* **137**: 371–380.

Burns TM, Harding RM, Dale JL. **1995**. The genome organization of banana bunchy top virus: Analysis of six ssDNA components. *Journal of General Virology* **76**: 1471–1482.

Clérot D, Bernardi F. **2006**. DNA Helicase Activity Is Associated with the Replication Initiator Protein Rep of Tomato Yellow Leaf Curl Geminivirus. *Journal of Virology* **80**: 11322–

11330.

Crous PW, Braun U. **2003**. *Mycosphaerella* and its anamorphs: 1. *Names published in Cercospora and Passalora*.

Dale JL. **1987**. Banana bunchy top: An economically important tropical plant virus disease. *Advances in Virus Research* **33**: 301–325.

Dale JL, Phillips DA, Parry JN. **1986**. Double-stranded RNA in Banana Plants with Bunchy Top Disease. *Journal of General Virology* **67**: 371–375.

Denham TP, Haberle SG, Lentfer C, Fullagar R, Field J, Therin M, Porch N, Winsborough B. **2003**. Origins of agriculture at Kuk Swamp in the highlands of New Guinea. *Science* **301**: 189–193.

Dhed'a B. **2019**. La culture des bananiers et bananiers plantains en République Démocratique du Congo. Support didactique. *Presse Universitaire UNIKIS*: 1–82.

Dietzgen RG, Thomas JE. **1991**. Properties of virus-like particles associated with banana bunchy top disease in Hawaii, Indonesia and Tonga. *Australasian Plant Pathology* **20**: 161–165.

Ding SW, Voinnet O. **2007**. Antiviral Immunity Directed by Small RNAs. *Cell* **130**: 413–426.

Dring DM, Wardlaw CW. **1974**. Banana Diseases Including Plantains and Abaca. *Kew Bulletin* **29**: 460.

Erdmann JB, Shepherd DN, Martin DP, Varsani A, Rybicki EP, Jeske H. **2010**. Replicative intermediates of maize streak virus found during leaf development. *Journal of General Virology* **91**: 1077–1081.

Espino RC, Magnaye LV, Johns AP, Juanillo C. **1993**. Evaluation of Philippine banana cultivars for resistance to bunchy-top and Fusarium wilt. In: Proceedings: International Symposium on Recent Developments in Banana Cultivation Technology. 89–102.

Fones HN, Bebber DP, Chaloner TM, Kay WT, Steinberg G, Gurr SJ. **2020**. Threats to global food security from emerging fungal and oomycete crop pathogens. *Nature Food* **1**: 332–342.

Fontes EPB, Gladfelter HJ, Schaffer RL, Petty ITD, Hanley-Bowdoin L. **1994**. Geminivirus replication origins have a modular organization. *Plant Cell* **6**: 405–416.

Fontes EPB, Luckow VA, Hanley-Bowdoin L. **1992**. A geminivirus replication protein is a sequence-specific DNA binding protein. *Plant Cell* **4**: 597–608.

Fouré E, Tézenas Du Montcel H. **2000**. Les productions bananières : un enjeu économique majeur pour la sécurité alimentaire. *Bulletin de Liaison (Coopération Régionale pour le Développement des Productions Horticoles en Afrique)*: 23–28.

Fu HC, Hu JM, Hung TH, Su HJ, Yeh HH. **2009**. Unusual events involved in Banana bunchy top virus strain evolution. *Phytopathology* **99**: 812–822.

García-Bastidas FA, Quintero-Vargas JC, Ayala-Vasquez M, Schermer T, Seidl MF, Santos-Paiva M, Noguera AM, Aguilera-Galvez C, Wittenberg A, Hofstede R, *et al.* 2020. First report of Fusarium Wilt Tropical race 4 in Cavendish Bananas Caused by Fusarium odoratissimum in Colombia. *Plant Disease* **104**.

Geering ADW, Thomas JE. **1997**. Search for alternative hosts of banana bunchy top virus in Australia. *Australasian Plant Pathology 1997 26:4* **26**: 250–254.

Gowen SR. **1995**. Pests. *Bananas and Plantains*: 382–402.

Grigoras I, Vetten HJ, Commandeur U, Ziebell H, Gronenborn B, Timchenko T. **2018**. Nanovirus DNA-N encodes a protein mandatory for aphid transmission. *Virology* **522**: 281–291.

Gronenborn B. **2004**. Nanoviruses: Genome organisation and protein function. In: Veterinary Microbiology. Elsevier, 103–109.

Gronenborn B, Randles JW, Knierim D, Barrière Q, Vetten HJ, Warthmann N, Cornu D, Sileye T, Winter S, Timchenko T. **2018**. Analysis of DNAs associated with coconut foliar decay disease implicates a unique single-stranded DNA virus representing a new taxon. *Scientific Reports* **8**.

Hafner GJ, Harding RM, Dale JL. **1995**. Movement and transmission of banana bunchy top virus DNA component one in bananas. *Journal of General Virology* **76**: 2279–2285.

Hafner GJ, Harding RM, Dale JL. **1997a**. A DNA primer associated with banana bunchy top virus. *Journal of General Virology* **78**: 479–486.

Hafner GJ, Stafford MR, Wolter LC, Harding RM, Dale JL. **1997b**. Nicking and joining activity of banana bunchy top virus replication protein in vitro. *Journal of General Virology*

78: 1795–1799.

Häkkinen M. **2013**. Reappraisal of sectional taxonomy in Musa (Musaceae). *Taxon* **62**: 809–813.

Hamim I, Green JC, Borth WB, Melzer MJ, Wang YN, Hu JS. **2017**. First report of Banana bunchy top virus in Heliconia spp. on Hawaii. *Plant Disease* **101**: 2153.

Hanley-Bowdoin L, Settlage SB, Orozco BM, Nagar S, Robertson D. **2000**. Geminiviruses: Models for plant DNA replication, transcription, and cell cycle regulation. *Critical Reviews in Biochemistry and Molecular Biology* **35**: 105–140.

Harding RM, Burns TM, Dale JL. **1991**. Virus-like particles associated with banana bunchy top disease contain small single-stranded DNA. *Journal of General Virology* **72**: 225–230.

Harding RM, Burns TM, Hafner G, Dietzgen RG, Dale JL. **1993**. Nucleotide sequence of one component of the banana bunchy top virus genome contains a putative replicase gene. *Journal of General Virology* **74**: 323–328.

Herrera-Valencia VA, Dugdale B, Harding RM, Dale JL. **2006**. An iterated sequence in the genome of Banana bunchy top virus is essential for efficient replication. *Journal of General Virology* **87**: 3409–3412.

Herrera-Valencia VA, Dugdale B, Harding RM, Dale JL. **2007**. Mapping the 5′ ends of banana bunchy top virus gene transcripts. *Archives of Virology* **152**: 615–620.

Heslop-Harrison JS, Schwarzacher T. **2007**. Domestication, genomics and the future for banana. *Annals of Botany* **100**: 1073–1084.

Hooks CRR, Wright MG, Kabasawa DS, Manandhar R, Almeida RPP. **2008**. Effect of banana bunchy top virus infection on morphology and growth characteristics of banana. *Annals of Applied Biology* **153**: 1–9.

Horser CL, Harding RM, Dale JL. **2001**. Banana bunchy top nanovirus DNA-1 encodes the 'master' replication initiation protein. *Journal of General Virology* **82**: 459–464.

Hu JS, Wang M, Sether D, Xie W, Leonhardt KW. **1996**. Use of polymerase chain reaction (PCR) to study transmission of banana bunchy top virus by the banana aphid (Pentalonia nigronervosa). *Annals of Applied Biology* **128**: 55–64.

Jeske H. **2009**. Geminiviruses. *Current Topics in Microbiology and Immunology* **331**: 185–

226.

Jeske H, Lütgemeier M, Preiß W. **2001**. DNA forms indicate rolling circle and recombination-dependent replication of Abutilon mosaic virus. *EMBO Journal* **20**: 6158–6167.

Ji X long, Yu N tong, Qu L, Li B bin, Liu Z xin. **2019**. Banana bunchy top virus (BBTV) nuclear shuttle protein interacts and re-distributes BBTV coat protein in Nicotiana benthamiana. *3 Biotech* **9**: 1–6.

Jones D. **2019**. *Handbook of diseases of banana, abaca and enset*.

Jovel J, Preiß W, Jeske H. **2007**. Characterization of DNA intermediates of an arising geminivirus. *Virus Research* **130**: 63–70.

Karan M, Harding RM, Dale JL. **1994**. Evidence for two groups of banana bunchy top virus isolates. *Journal of General Virology* **75**: 3541–3546.

Kazlauskas D, Varsani A, Koonin E V., Krupovic M. **2019**. Multiple origins of prokaryotic and eukaryotic single-stranded DNA viruses from bacterial and archaeal plasmids. *Nature Communications* **10**: 1–12.

Kesavamoorthy RC. **1980**. Radical changes in ecosystem in the Pulney hills. In: C. R. Muthukrishnan, Abdul Chaser,&J. B. M. Md (Eds.) Proc. 13th national seminar on banana production technology, Coimbatore: TNAU. 23-28.

King A, Lefkowitz E, Adams M, Carstens E. **2011**. Virus taxonomy: ninth report of the International Committee on Taxonomy of Viruses.

Kolombia YA, Oviasuyi T, Dzola AK, Ale Gonh-Goh A, Atsu T, Oresanya A, Ogunsanya P, Alabi T, Kumar PL. **2021**. First Report of Banana Bunchy Top Virus in Banana (Musa spp.) and Its Eradication in Togo. *Plant Disease*.

Koonin E V., Dolja V V., Krupovic M, Kuhn JH. **2021**. Viruses Defined by the Position of the Virosphere within the Replicator Space. *Microbiology and Molecular Biology Reviews*.

Krupovic M. **2013**. Networks of evolutionary interactions underlying the polyphyletic origin of ssDNA viruses. *Current Opinion in Virology* **3**: 578–586.

Krupovic M, Varsani A, Kazlauskas D, Breitbart M, Delwart E, Rosario K, Yutin N, Wolf YI, Harrach B, Zerbini FM, *et al.* 2020. Cressdnaviricota : a Virus Phylum Unifying Seven Families of Rep-Encoding Viruses with Single-Stranded, Circular DNA Genomes. *Journal of*

Virology **94**.

Kumar PL, Hanna R, Alabi OJ, Soko MM, Oben TT, Vangu GHP, Naidu RA. **2011**. Banana bunchy top virus in sub-Saharan Africa: Investigations on virus distribution and diversity. *Virus Research* **159**: 171–182.

Kumar J, Kumar J, Singh SP, Tuli R. **2014**. Association of Satellites with a Mastrevirus in Natural Infection: Complexity of Wheat Dwarf India Virus Disease. *Journal of Virology* **88**: 7093–7104.

Kumar PL, Selvarajan R, Iskra-Caruana ML, Chabannes M, Hanna R. **2015**. Biology, etiology, and control of virus diseases of banana and plantain. In: Advances in Virus Research. Academic Press Inc., 229–269.

Lageix S, Catrice O, Deragon J-M, Gronenborn B, Pélissier T, Ramírez BC. **2007**. The Nanovirus-Encoded Clink Protein Affects Plant Cell Cycle Regulation through Interaction with the Retinoblastoma-Related Protein. *Journal of Virology* **81**: 4177–4185.

Lassoudière A. **2007**. Le bananier et sa culture.

Laufs J, Traut W, Heyraud F, Matzeit V, Rogers SG, Schell J, Gronenborn B. **1995**. In vitro cleavage and joining at the viral origin of replication by the replication initiator protein of tomato yellow leaf curl virus. *Proceedings of the National Academy of Sciences of the United States of America* **92**: 3879–3883.

Lescot T. **2015**. La diversité génétique des bananiers. *Fruitrop (Ed. Française)*.

Lescot T. **2020**. Diversité génétique de la banane. *Fruitrop (Ed. Française)*.

Lin LS, Meyer RJ. **1987**. Dna synthesis is initiated at two positions within the origin of replication of plasmid R1162. *Nucleic Acids Research* **15**: 8319–8331.

Londoño A, Riego-Ruiz L, Argüello-Astorga GR. **2010**. DNA-binding specificity determinants of replication proteins encoded by eukaryotic ssDNA viruses are adjacent to widely separated RCR conserved motifs. *Archives of Virology* **155**: 1033–1046.

Lowe S, Browne M, Boudjelas S, Poorter M De. **2020**. 100 of the World's Worst Invasive Alien Species: A Selection From The Global Invasive Species Database. In: Encyclopedia of Biological Invasions. 715–716.

Magee CJP. **1927**. Banana bunchy top disease of the banana. *CSIResearch Bull*.

Magee CJP. **1940**. Transmission Studies on the Banana Bunchy-top Virus. *Journal of the Australian Institute of Agricultural Science* **6**.

Magee CJ. **1953**. Some aspects of the Bunchy Top Disease of Banana and other Musa spp. *Journal and Proceedings of the Royal Society of New, South Wales. Vol. XXV. 1891. 348p* **87**: 3–18.

Mandal B. **2010**. Advances in small isometric multicomponent ssDNA viruses infecting plants. *Indian Journal of Virology* **21**: 18–30.

Mandal B, Mandal S, Pun KB, Varma A. **2004**. First Report of the Association of a Nanovirus with Foorkey Disease of Large Cardamom in India. *Plant Disease* **88**: 428–428.

Mandal B, Shilpi S, Barman AR, Mandal S, Varma A. **2013**. Nine novel DNA components associated with the foorkey disease of large cardamom: Evidence of a distinct babuvirus species in Nanoviridae. *Virus Research* **178**: 297–305.

Di Mattia J, Vernerey M-S, Yvon M, Pirolles E, Villegas M, Gaafar Y, Ziebell H, Michalakis Y, Zeddam J-L, Blanc S. **2020**. Route of a Multipartite Nanovirus across the Body of Its Aphid Vector. *Journal of Virology* **94**.

Mpoki S, Mahuku G, Mbanzibwa DR, Mkamilo G, Mark D, Mosha H, Pallangyyo BD, Fihavango M, Oresanya A, Ogunsanya P, *et al.* 2021. First report of banana bunchy top virus in banana and plantain (Musa spp.) in Tanzania. *Plant Disease*.

Nawaz-ul-Rehman MS, Nahid N, Mansoor S, Briddon RW, Fauquet CM. **2010**. Post-transcriptional gene silencing suppressor activity of two non-pathogenic alphasatellites associated with a begomovirus. *Virology* **405**: 300–308.

Nelson SC. **2004**. Banana bunchy top : detailed signs and symptoms. *Knowledge Creation Diffusion Utilization*: 1–22.

Niu S, Wang B, Guo X, Yu J, Wang X, Xu K, Zhai Y, Wang J, Liu Z. **2009**. Identification of two RNA silencing suppressors from banana bunchy top virus. *Archives of Virology* **154**: 1775–1783.

Perrier X, Bakry F, Carreel F, Jenny C, Horry JP, Lebot V, Hippolyte I. **2009**. Combining biological approaches to shed light on the evolution of edible bananas. *Ethnobotany Research and Applications* **7**: 199–216.

Perrier X, De Langhe E, Donohue M, Lentfer C, Vrydaghs L, Bakry F, Carreel F, Hippolyte I, Horry JP, Jenny C, *et al.* 2011. Multidisciplinary perspectives on (Musa spp.) domestication. *Proceedings of the National Academy of Sciences of the United States of America* **108**: 11311–11318.

Pinili, M.S., I. Nagshima, T.O. Dizon KTN. **2013**. Cross-Transmission and New Alternate Hosts of Banana bunchy top virus. *Tropical Agriculture and Development* **57**: 1–7.

Ploetz RC. **2006**. Fusarium wilt of banana is caused by several pathogens referred to as Fusarium oxysporum f. sp. cubense. In: Phytopathology. The American Phytopathological Society, 653–656.

Ploetz RC. **2015**. Management of Fusarium wilt of banana: A review with special reference to tropical race 4. *Crop Protection* **73**: 7–15.

Ploetz RC, Kema GHJ, Ma LJ. **2015**. Impact of Diseases on Export and Smallholder Production of Banana. *Annual Review of Phytopathology* **53**: 269–288.

Ploetz RC, Kepler K, Daniells J, Nelson SC. **2007**. *Banana and plantain—an overview with emphasis on Pacific island cultivars Musaceae (banana family)*.

Pooggin MM. **2013**. How can plant DNA viruses evade siRNA-directed DNA methylation and silencing? *International Journal of Molecular Sciences* **14**: 15233–15259.

Pooggin MM. **2018**. Small RNA-omics for plant virus identification, virome reconstruction, and antiviral defense characterization. *Frontiers in Microbiology* **9**: 2779.

Preiss W, Jeske H. **2003**. Multitasking in Replication Is Common among Geminiviruses. *Journal of Virology* **77**: 2972–2980.

Price NS. **1995**. The origin and development of banana and plantain cultivation. In: Bananas and Plantains. Springer Netherlands, 1–13.

Qazi J. **2016**. Banana bunchy top virus and the bunchy top disease. *Journal of General Plant Pathology* **82**: 2–11.

Rahayuniati RF, Subandiyah S, Hartono S, Somowiyarjo S, Kurniawan REK, Prakoso AB, Crew K, Vance ME, Ray JD, Thomas JE. **2021**. Recent distribution and diversity analysis on banana bunchy top virus of banana and alternative host in Indonesia. *Tropical Plant Pathology* **46**: 506–517.

Rao XQ, Wu ZL, Zhou L, Sun J, Li HP. **2017**. Genetic diversity of banana bunchy top virus isolates from China. *Acta Virologica* **61**: 217–222.

Safari Murhububa I, Tougeron K, Bragard C, Fauconnier ML, Bisimwa Basengere E, Walangululu Masamba J, Hance T. **2021**. Banana Tree Infected with Banana Bunchy Top Virus Attracts Pentalonia nigronervosa Aphids Through Increased Volatile Organic Compounds Emission. *Journal of Chemical Ecology* **47**: 755–767.

Selvarajan R, Balasubramanian V. **2014**. Host-virus interactions in banana-infecting viruses. In: Plant Virus-Host Interaction: Molecular Approaches and Viral Evolution. Academic Press, 57–78.

Sharman M, Thomas JE, Skabo S, Holton TA. **2008**. Abacá bunchy top virus, a new member of the genus Babuvirus (family Nanoviridae). *Archives of Virology* **153**: 135–147.

Sicard A, Michalakis Y, Gutiérrez S, Blanc S. **2016**. The Strange Lifestyle of Multipartite Viruses. *PLoS Pathogens* **12**: e1005819.

Sicard A, Pirolles E, Gallet R, Vernerey MS, Yvon M, Urbino C, Peterschmitt M, Gutierrez S, Michalakis Y, Blanc S. **2019**. A multicellular way of life for a multipartite virus. *eLife* **8**.

Simmonds NW. **1962**. The Evolution of the Bananas. *The Evolution of the Bananas.*

Simmonds NW, Shepherd K. **1955**. The taxonomy and origins of the cultivated bananas. *Journal of the Linnean Society of London, Botany* **55**: 302–312.

Soomro MH, Khalid S, Aslam M. **1992**. Outbreak of banana bunchy top virus in Sindh, Pakistan. *FAO Plant Protection Bulletin* **40/3**: 95–100.

Stainton D, Kraberger S, Walters M, Wiltshire EJ, Rosario K, Halafihi M, Lolohea S, Katoa I, Faitua TH, Aholelei W, *et al.* 2012. Evidence of inter-component recombination, intra-component recombination and reassortment in banana bunchy top virus. *Journal of General Virology* **93**: 1103–1119.

Stainton D, Martin DP, Muhire BM, Lolohea S, Halafihi M, Lepoint P, Blomme G, Crew KS, Sharman M, Kraberger S, *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.

Stover RH. **1962**. *Fusarial wilt (panama disease) of bananas and other Musa species.*

Su HJ, Wu RY, Tsao LY. **1993**. Ecology of banana bunchy top virus disease.

Suparman S, Gunawan B, Pujiastuti Y, Arsi A, Cameron RR. **2017**. Alternative Hosts of Banana Aphid Pentalonia nigronervosa Coq. (Hemiptera: Aphididae), the Vector Transmitting Banana Bunchy Top Virus. *Journal of Advanced Agricultural Technologies* **4**: 354–359.

tenOever BR. **2016**. The Evolution of Antiviral Defense Systems. *Cell Host and Microbe* **19**: 142–149.

Thomas JE, Dietzgen RG. **1991**. Purification, characterization and serological detection of virus-like particles associated with banana bunchy top disease in Australia. *Journal of General Virology* **72**: 217–224.

Thomas JE, Geering ADW, Dahal G, Lockhart BEL, Thottappilly G. **2003**. Banana and Plantain. In: Virus and Virus-like Diseases of Major Crops in Developing Countries. Springer, Dordrecht, 477–496.

Thomas JE, Iskra Caruana ML. **2000**. Bunchy top. In: Diseases of Banana, Abaca and Enset. 241–253.

Thömmes P, Osman TAM, Hayes RJ, Buck KW. **1993**. TGMV replication protein AL1 preferentially binds to single-stranded DNA from the common region. *FEBS Letters* **319**: 95–99.

Timchenko T, Katul L, Aronson M, Vega-Arreguín JC, Ramirez BC, Vetten HJ, Gronenborn B. **2006**. Infectivity of nanovirus DNAs: Induction of disease by cloned genome components of Faba bean necrotic yellows virus. *Journal of General Virology* **87**: 1735–1743.

Varsani A, Martin DP, Randles JW, Vetten HJ, Thomas JE, Fiallo-Olivé E, Navas-Castillo J, Lett JM, Zerbini FM, Roumagnac P, *et al.* 2021. Taxonomy update for the family Alphasatellitidae: new subfamily, genera, and species. *Archives of Virology* **1**: 1–9.

Wanitchakorn R, Hafner GJ, Harding RM, Dale JL. **2000a**. Functional analysis of proteins encoded by banana bunchy top virus DNA-4 to -6. *Journal of General Virology* **81**: 299–306.

Wanitchakorn R, Harding RM, Dale JL. **1997**. Banana bunchy top virus DNA-3 encodes the viral coat protein. *Archives of Virology* **142**: 1673–1680.

Wanitchakorn R, Harding RM, Dale JL. **2000b**. Sequence variability in the coat protein gene

of two groups of banana bunchy top isolates. *Archives of Virology* **145**: 593–602.

Watanabe S, Bressan A. **2013**. Tropism, compartmentalization and retention of banana bunchy top virus (Nanoviridae) in the aphid vector Pentalonia nigronervosa. *Journal of General Virology* **94**: 209–219.

Watanabe S, Greenwell AM, Bressan A. **2013**. Localization, concentration, and transmission efficiency of banana bunchy top virus in four asexual lineages of Pentalonia aphids. *Viruses* **5**: 758–776.

Wu RY, Su HJ. **1990**. Purification and Characterization of Banana Bunchy Top Virus. *Journal of Phytopathology* **128**: 153–160.

Yoichi H, Hiroshi S, Tohru K. **1988**. Two single-strand DNA initiation signals located in the oriV region of plasmid RSF1010. *Gene* **68**: 221–228.

Yu NT, Zhang YL, Feng TC, Wang JH, Kulye M, Yang WJ, Lin ZS, Xiong Z, Liu ZX. **2012**. Cloning and sequence analysis of two banana bunchy top virus genomes in Hainan. *Virus Genes* **44**: 488–494.

Zakaria L. **2021**. Diversity of colletotrichum species associated with anthracnose disease in tropical fruit crops - A review. *Agriculture (Switzerland)* **11**: 297.

Zhuang Jun TE, Zhi-Xin L. **2005**. Cloning and sequencing of DNA components of Banana bunchy top virus Hainan isolate. *Chinese Journal of Agricultural Biotechnology* **2**: 91–97.

Supplementary data

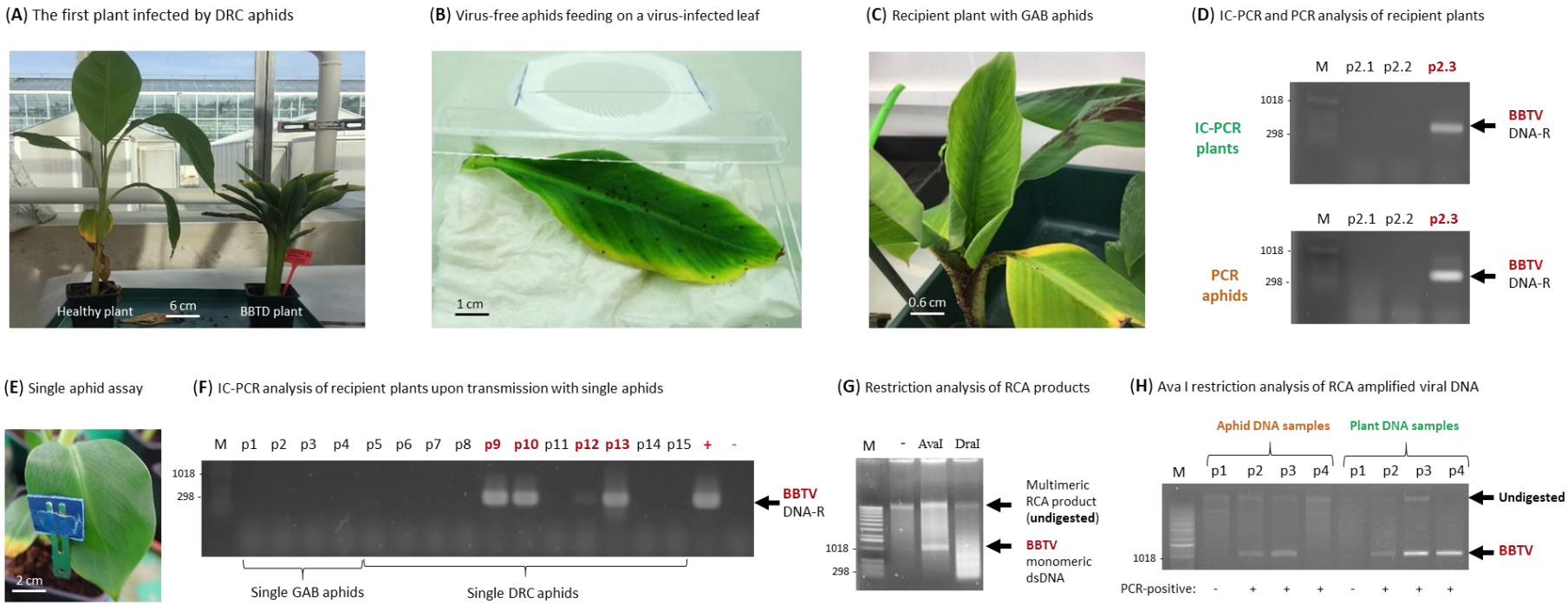
This part contains the supplementary data accompanying the research articles in Chapters 1, 2 and 3. It contain also one supplementary data for the introduction



Supplementary data S1: comparison of BBTD symptoms to healthy banana plants.

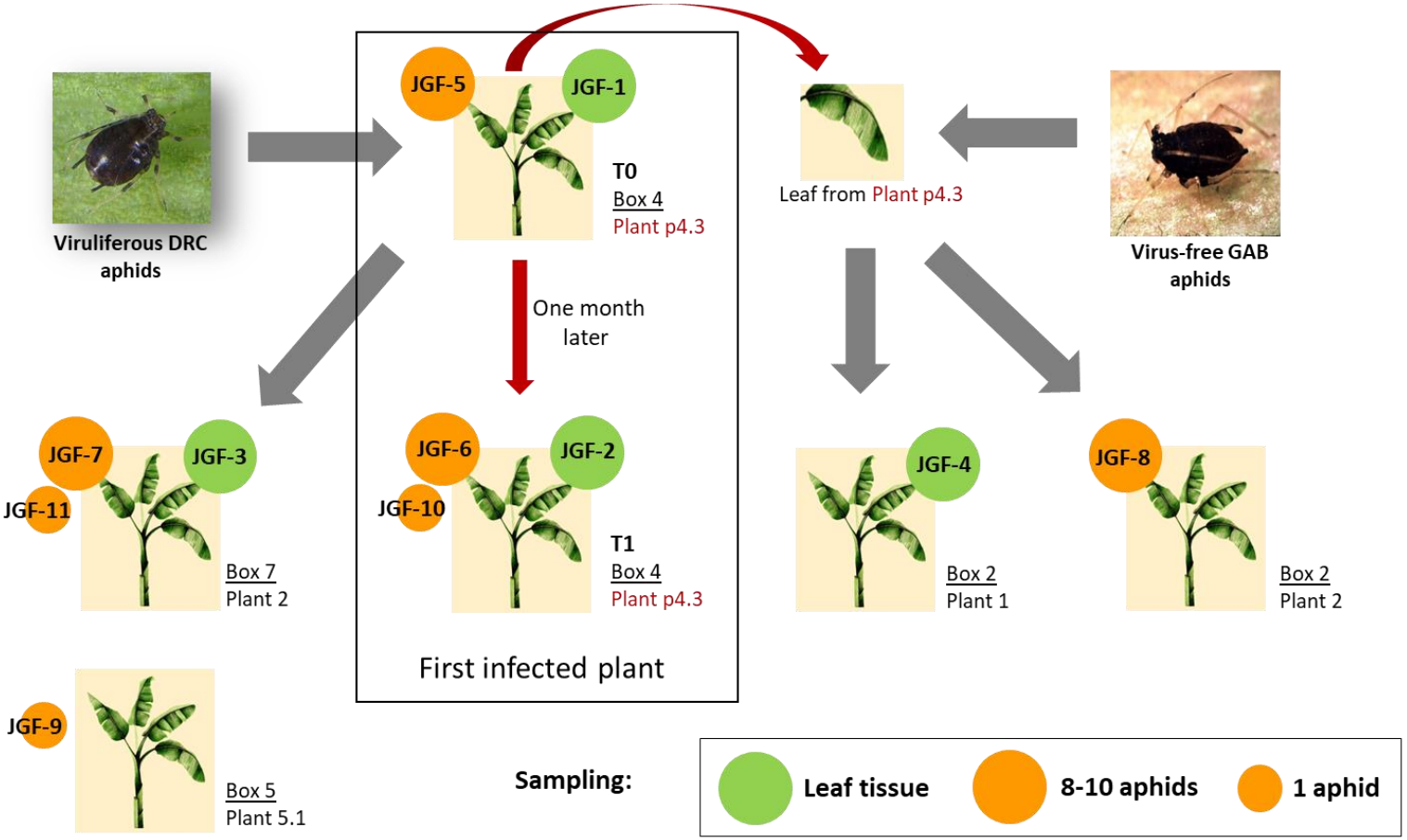
Left: healthy banana plants, cultivar Cavendish triploid AAA. Right: BBTV infected banana plants, cultivar Cavendish triploid AAA. All bananas plants have the same ages.

Supplementary data for Chapter 1: article 1:



S1 Fig. Banana bunchy top disease (BBTD) transmission to *Musa acuminata* Cavendish plants by the banana aphids from DRC (field DRC aphids) and Gabon (GAB aphids), followed by molecular analysis of recipient plants and aphids.

(A) Pictures of healthy and BBTV-infected Cavendish banana plants, following disease transmission by the field aphids collected on a symptomatic banana plant in the DRC province Bas Congo (now Congo-Central) in December 2016. (B) Aphid feeding on a detached leaf of BBTV-infected plant for 24 hrs to acquire the virus. (C) The symptomatic recipient plant colonised by aphids, following disease transmission by GAB aphids. (D) Immuno-capture (IC)-PCR of leaf tissues and PCR analysis of total DNA from respectively GAB aphids-inoculated recipient plants and aphids from the corresponding recipient plants (p2.1, p2.2 and p2.3). Position of DNA-R PCR product is indicated by arrow. (E) Virus transmission with a single viruliferous aphid placed on a recipient plant leaf. (F) Immuno-capture (IC)-PCR analysis of recipient plants upon transmission with single GAB and DRC aphids at 4 weeks post-inoculation. (G) Restriction analysis of RCA amplified viral DNA from the BBTD plant using AvaI and DraI enzymes. Positions of the undigested multimeric RCA product and the monomeric AvaI digestion products are indicated by arrows. (H) AvaI restriction analysis of RCA amplified viral DNA from BBTV-infected recipient plants (p1, p2, p3 and p4) and aphids taken from these plants. Positions of undigested and digested RCA products are indicated. In panels D, F, G and H, "M" stands for 1 Kb DNA ladder and positions of its 1018 and 298 bp bands are indicated.



S2 Fig. Banana bunchy top disease transmission by viruliferous *P. nigronervosa* aphids from Democratic Republic of the Congo (DRC) and virus-free banana aphids from Gabon (GAB) (maintained in Belgium for >15 years) and the plant and aphid samples used for rolling circle amplification and Illumina sequencing of viral DNA.

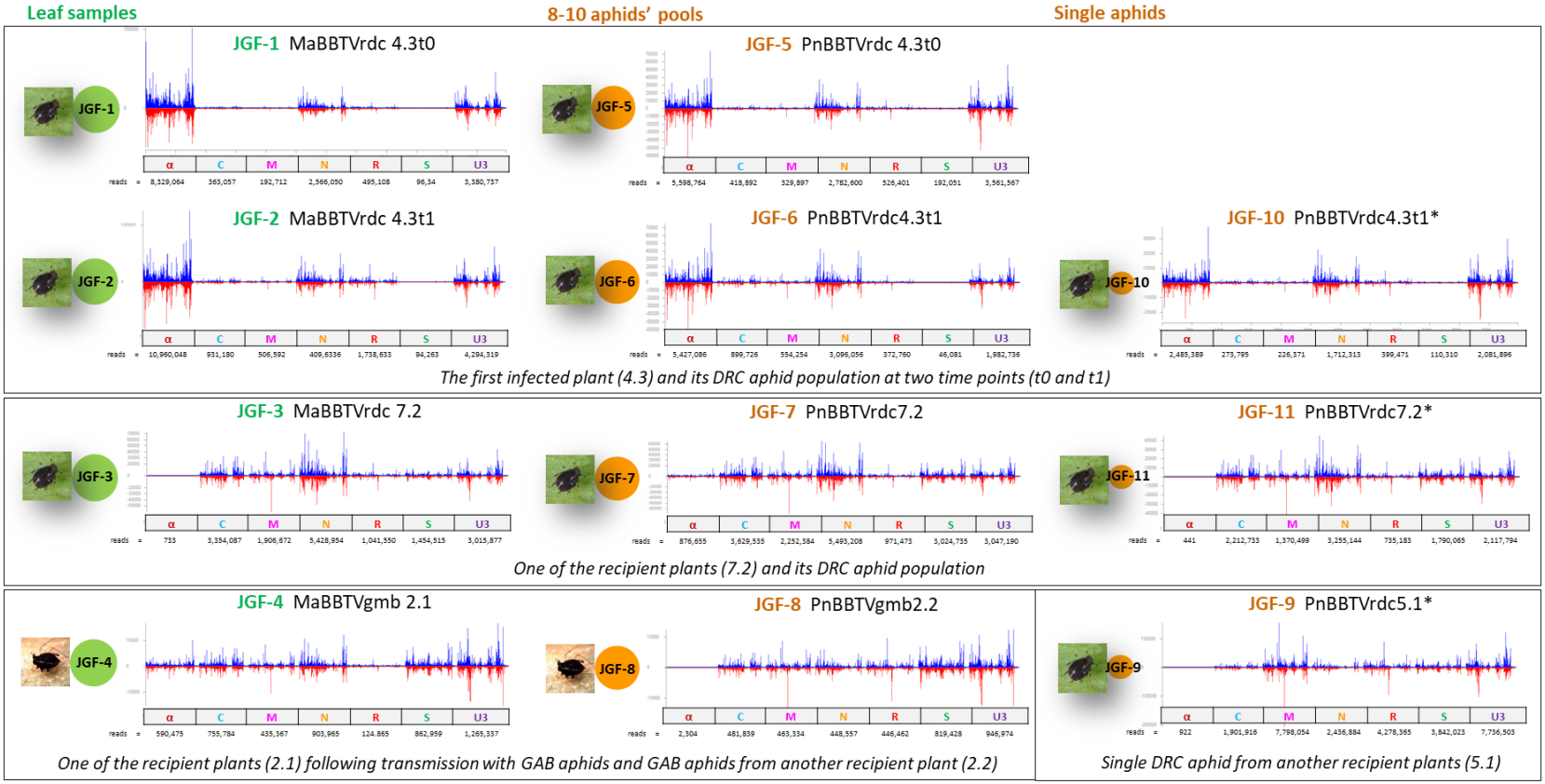
Scheme of the transmission experiments with the field DRC aphids and GAB aphids fed on a detached leaf the first infected plant (p4.3) is depicted with the collected samples of leaf tissues (JGF-1-4) and aphids (JGF-5-11) indicated with green and orange circles, respectively. T0 and T1 are two sampling time-points for the first infected plant p4.3.



Sample	α	C	M	N	R	S	U3
JGF-1 MaBBTVrdc 4.3t0	single	single	v1 (70 %) v2 (30 %)	single	single	v1 (70 %) v2 (30 %)	v1 (80 %) v2 (20 %)
JGF-2 MaBBTVrdc 4.3t1	single	single	v1 (60 %) v2 (40 %)	single	single	v1 (70 %) v2 (30 %)	v1 (70 %) v2 (30 %)
JGF-3 MaBBTVrdc 7.2	? (730 reads)	single	v1 (100%)	single	single	v1 (100%)	v2 (60 %) v1 (40 %)
JGF-4 MaBBTVgmb 2.1	single	single	v2 (75 %) v1 (25 %)	single	single	v2 (50 %) v1 (50 %)	v3 (100 %)
JGF-5 PnBBTVrdc 4.3t0	single	single	v1 (70 %) v2 (30 %)	single	single	v1 (60 %) v2 (40 %)	v1 (70 %) v2 (30 %)
JGF-6 PnBBTVrdc4.3t1	single	single	v1 (60 %) v2 (40 %)	single	single	v1 (70 %) v2 (30 %)	v1 (70 %) v2 (30 %)
JGF-7 PnBBTVrdc7.2	single	single	v1 (100%)	single	single	v1 (100%)	v2 (60 %) v1 (40 %)
JGF-8 PnBBTVgmb2.2	single	single	v2 (85 %) v1 (15 %)	single	single	v2 (50 %) v1 (50 %)	v3 (100 %)
JGF-9 PnBBTVrdc5.1*	? (922 reads)	single	v1 (100%)	single	single	v1 (100%)	v1 (50 %) v4 (50 %)
JGF-10 PnBBTVrdc4.3t1*	single	single	v1 (50 %) v2 (50 %)	single	single	v1 (50 %) v2 (50 %)	v1 (70 %) v2 (30 %)
JGF-11 PnBBTVrdc7.2*	? (441 reads)	single	v1 (100%)	single	single	v1 (100%)	v2 (60 %) v1 (40 %)

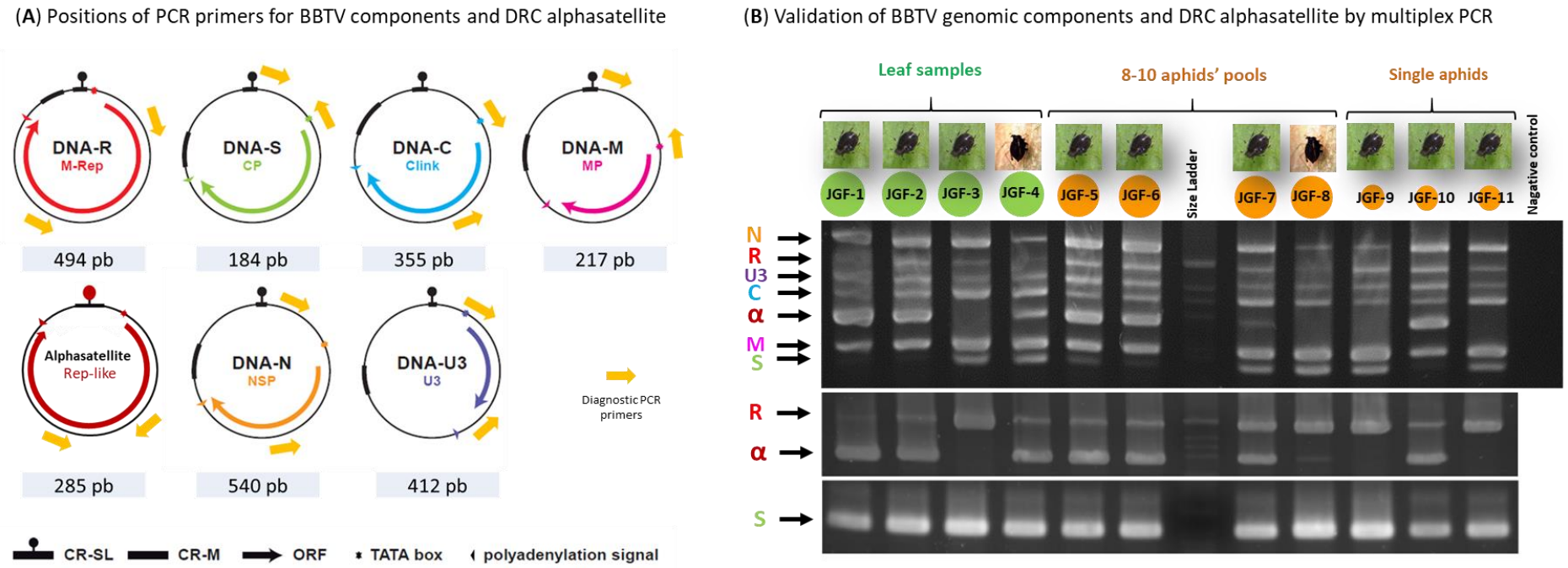
S3 Fig. Genetic variants of BBTV components and alphasatellite identified in Cavendish plants and aphids taken from these plants by rolling circle amplification and Illumina sequencing of viral DNA.

For each of the 11 samples (JGF-1-11, see Supplementary Figure S2) of BBTD-infected plant leaf tissues and viruliferous aphids (indicated with green and orange circles, respectively), the Illumina sequencing reads were mapped on the de novo reconstructed sequences of six BBTV components (C, M, N, R, S, U3) and alphasatellite (α) and the mapped reads were analysed using MISIS-2 (Seguin et al. 2016) to identify single-nucleotide polymorphism (SNP) positions and calculate percentage (%) of each SNP variant (v1 and v2 for DNAs M and S, and v1, v2, v3 and v4 for DNA-U3) at each SNP position. Single variants of virome components (without SNPs exceeding 10%) are highlighted in green. Arcs with arrowheads indicate from which plants the aphid samples were collected.



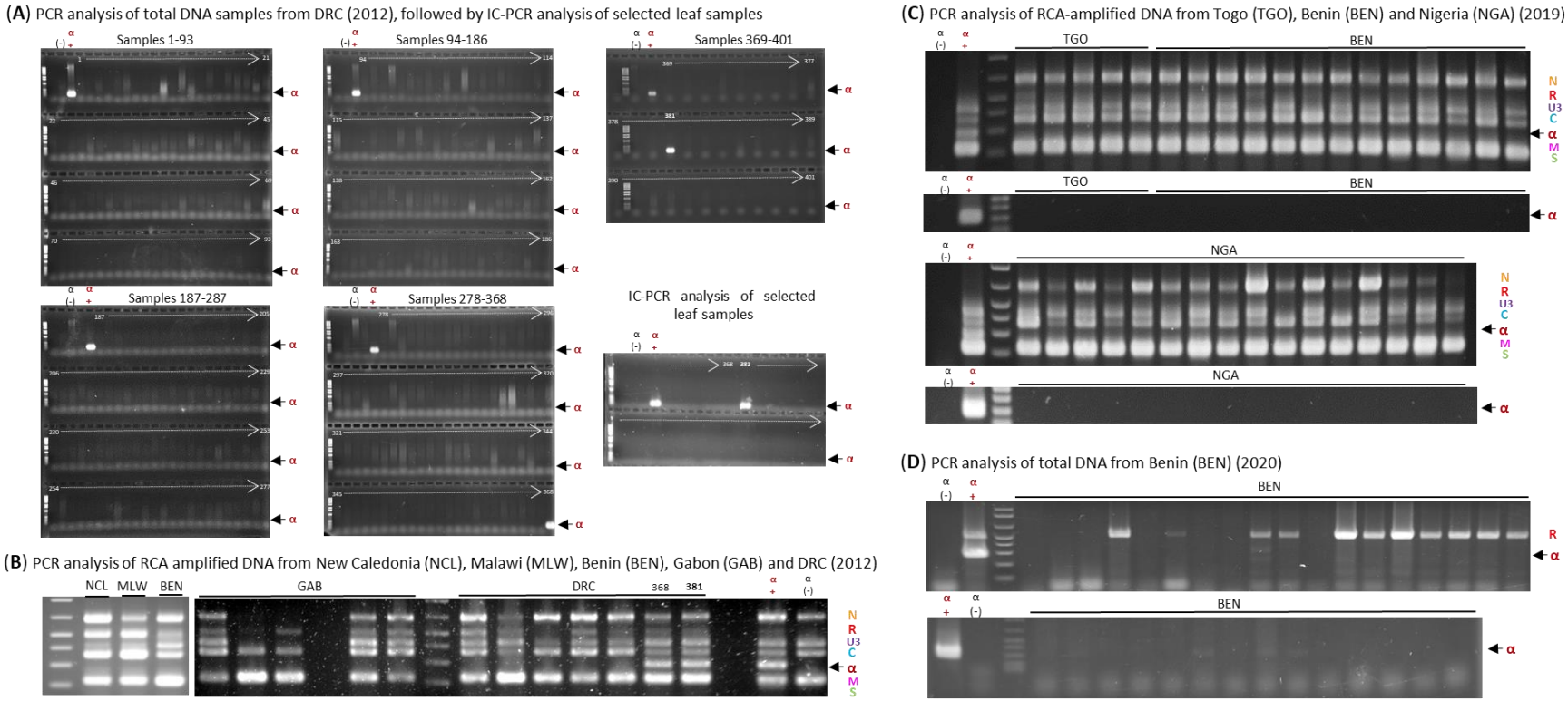
S4 Fig. Maps and relative abundance of Illumina sequencing reads representing RCA-amplified DNA of BBTV genome components and alphasatellite identified in BBTD-infected Cavendish plants and aphids taken from these plants.

For each of the 11 samples (JGF-1-11, see Supplementary Figure S2) of BBTD-infected plant leaf tissues and viruliferous aphids (indicated with green and orange circles, respectively), the Illumina 125 nt reads were mapped on concatenated sequences of alphasatellite (α) and six BBTV components (C, M, N, R, S, U3) and the mapped reads were visualized using MISIS-2 (Seguin et al. 2016) and counted. Histograms plot the numbers of viral 125 nt forward and reverse reads at each nucleotide position of the concatenated viral genome: blue bars above the axis represent forward reads ending at each respective position red bars below the axis represent reverse reads ending at each respective position. Numbers of reads mapped to each viral genome component are given below the histograms.



S5 Fig. Multiplex, duplex and single PCR validation of the BBTV genome components and DRC alphasatellite reconstructed by Illumina sequencing of RCA-amplified viral DNA from BBTD-infected Cavendish plants and aphids taken from these plants.

(A) Schematic representation of circular BBTV genome components (C, M, N, R, S, U3) and alphasatellite (α) with position of diagnostic PCR primers (yellow arrows). (B) Multiplex, duplex and single PCR analysis of the 11 samples (JGF-1-11, see Supplementary Figure S2) of BBTD-infected plant leaf tissues and viruliferous aphids (indicated with green and orange circles, respectively). PCR products were separated on 2% agarose gel. Positions of each BBTV component and alphasatellite are indicated by arrows.



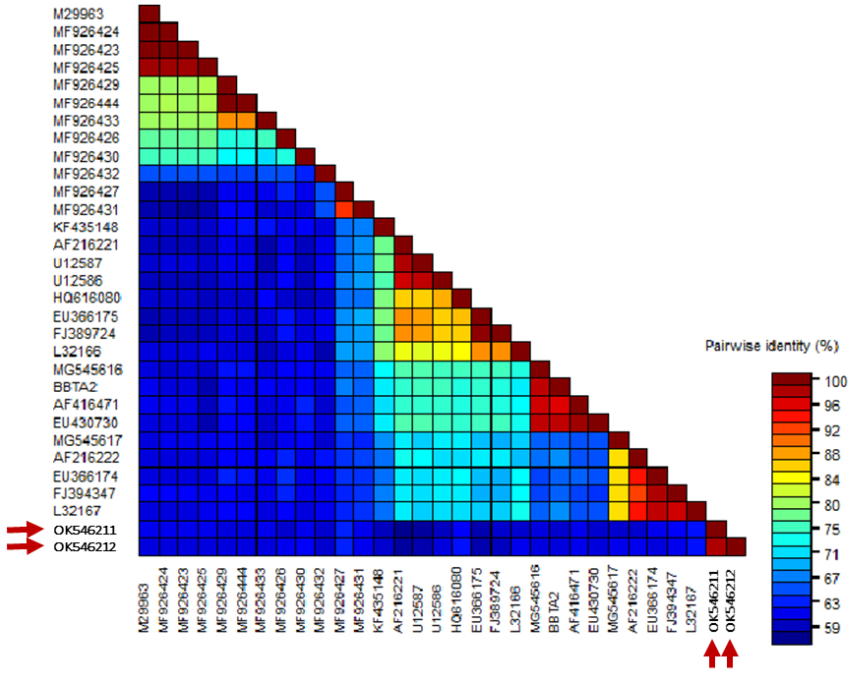
S6 Fig. PCR and immuno-capture (IC)-PCR analyses with alphasatellite-specific primers of banana plant samples from PIO (Sub-Saharan Africa and New Caledonia).

(A) PCR analysis of total DNA extracted from BBTV-infected plants from DRC (2012) using DRC-2016 alphasatellite-specific primers (S1 Table), followed by IC-PCR analysis of selected leaf samples. Position of the alphasatellite (α)-specific PCR product on each gel is indicated by an arrow. (B) Multiplex PCR analysis of RCA products of viral DNA from selected BBTV-infected plant samples from New Caledonia (NCL), Malawi (MLW), Benin (BEN), Gabon (GAB) and DRC using the primers for each BBTV component and DRC alphasatellite (S5 Fig) with positions of respective PCR products indicated by arrows. (C) Multiplex PCR analysis of RCA-amplified DNA of BBTV-infected plants from Togo (TGO), Benin (BEN) and Nigeria (NGA) (2019) using the primers for each BBTV component and DRC alphasatellite (S5 Fig) with positions of respective PCR products indicated by arrows, followed by single PCR analysis with degenerate PCR primers specific for DRC alphasatellite and all members of the genus *Fabenesatellite* (S1 Table). (D) Duplex and single PCR analysis of total DNA of BBTD-infected plants from Benin (BEN) (2020), using DRC- and Fabenesatellite-specific degenerate primers. Position of alphasatellite (α)-specific PCR product on each gel is indicated by an arrow. DRC alphasatellite-infected and healthy Cavendish plants were used as respectively positive "α (+)" and negative "α (-)" controls. Note that the PCR analysis of DNA samples from DRC revealed two alphasatellite-positive samples (368 and 381), while the follow-up IC-PCR of leaves confirmed the presence of encapsidated alphasatellite only in sample 381 (panel A), possibly due to cross-contamination during DNA extraction. Illumina sequencing of RCA amplified viral DNA (pre-analysed by multiplex PCR, panel B) revealed that the DNA samples 368 and 381 contain identical sequences of the alphasatellite and six BBTV components.

S7 Fig. Pairwise sequence comparison of DRC alphasatellite isolates DRC-2016 (OK546211) and DRC-2012 (OK546212) with other alphasatellites.

The nucleotide sequences of DRC alphasatellite isolates were compared with those of all alphasatellites from the subfamilies *Petromoalphasatellitinae* (A), *Nanoalphasatellitinae* (B) and *Geminialphasatellitinae* (C) available at the NCBI Genbank in September 2021 using Sequence Demarcation Tool (SDT) v1.2 with Muscles (excluding indels) (Muhire et al. 2014) and their pairwise identities (in %) were plotted as heatmap diagrams and shown in Tables. Two isolates of DRC alphasatellite are indicated with red arrows.

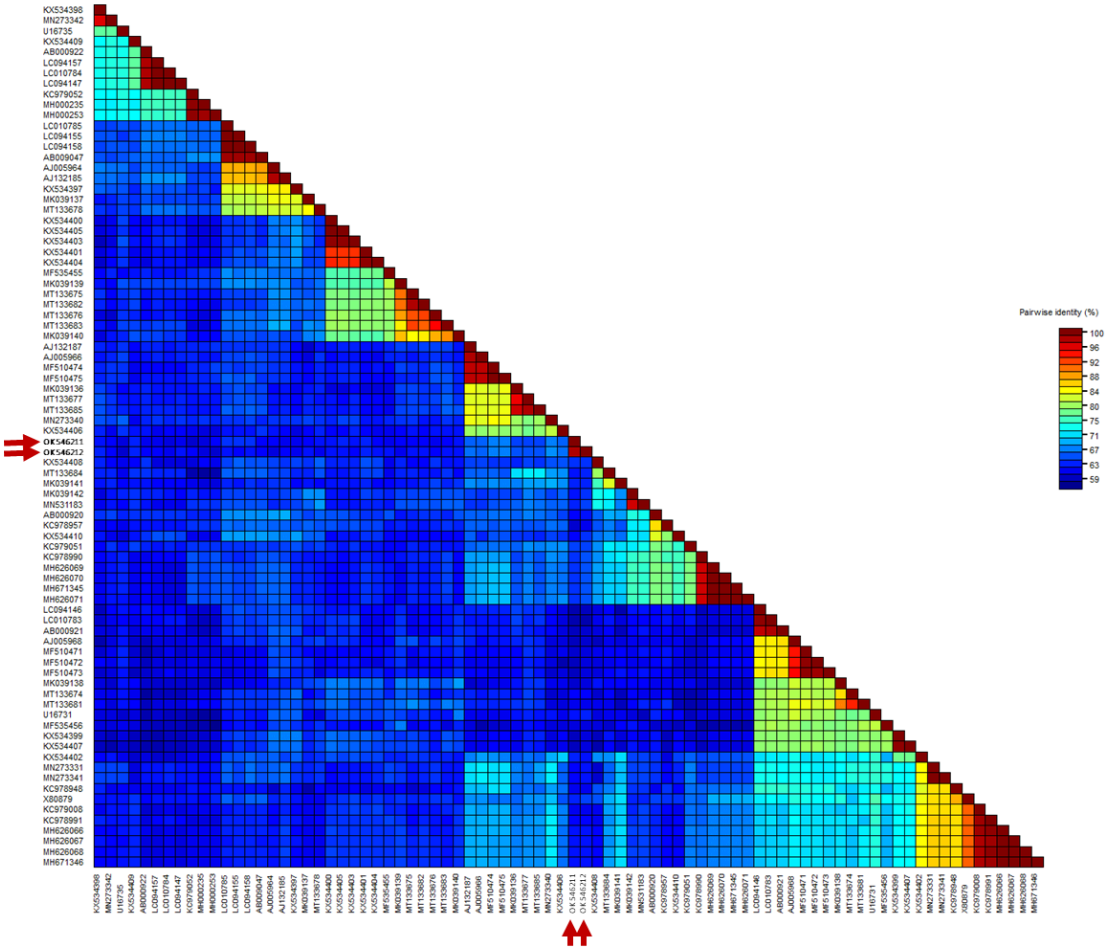
(A) DRC alphasatellite vs *Petromoalphasatellitinae*



First Sequence	Second Sequenc	Identity Score	%
>OK546211	>OK546212	0,9808044	98,1
>M29963	>OK546211	0,6062917	60,6
>MF926424	>OK546211	0,607245	60,7
>MF926423	>OK546211	0,6081983	60,8
>MF926425	>OK546211	0,6051379	60,5
>MF926429	>OK546211	0,6021297	60,2
>MF926444	>OK546211	0,6011617	60,1
>MF926433	>OK546211	0,6052889	60,5
>MF926426	>OK546211	0,5960785	59,6
>MF926430	>OK546211	0,608365	60,8
>MF926432	>OK546211	0,5919153	59,2
>MF926427	>OK546211	0,6282306	62,8
>MF926431	>OK546211	0,6101861	61,0
>KF435148	>OK546211	0,5901961	59,0
>AF216221	>OK546211	0,5666667	56,7
>U12587	>OK546211	0,5718446	57,2
>U12586	>OK546211	0,5852216	58,5
>HQ616080	>OK546211	0,6168699	61,7
>EU366175	>OK546211	0,5829195	58,3
>FJ389724	>OK546211	0,5912263	59,1
>L32166	>OK546211	0,5916334	59,2
>MG545616	>OK546211	0,5994036	59,9
>BBTA2	>OK546211	0,5974155	59,7
>AF416471	>OK546211	0,606	60,6
>EU430730	>OK546211	0,5954725	59,5
>MG545617	>OK546211	0,5918368	59,2
>AF216222	>OK546211	0,5974283	59,7
>EU366174	>OK546211	0,6072508	60,7
>FJ394347	>OK546211	0,6141249	61,4
>L32167	>OK546211	0,622449	62,2

First Sequence	Second Sequenc	Identity Score	%
>M29963	>OK546212	0,5977121	59,8
>MF926424	>OK546212	0,5986654	59,9
>MF926423	>OK546212	0,5977121	59,8
>MF926425	>OK546212	0,5996187	60,0
>MF926429	>OK546212	0,6021093	60,2
>MF926444	>OK546212	0,6011505	60,1
>MF926433	>OK546212	0,6036822	60,4
>MF926426	>OK546212	0,6045142	60,5
>MF926430	>OK546212	0,611379	61,1
>MF926432	>OK546212	0,5917387	59,2
>MF926427	>OK546212	0,6257485	62,6
>MF926431	>OK546212	0,6056752	60,6
>KF435148	>OK546212	0,5970297	59,7
>AF216221	>OK546212	0,5735294	57,4
>U12587	>OK546212	0,582762	58,3
>U12586	>OK546212	0,5998004	60,0
>HQ616080	>OK546212	0,6223132	62,2
>EU366175	>OK546212	0,5837513	58,4
>FJ389724	>OK546212	0,5903614	59,0
>L32166	>OK546212	0,6053719	60,5
>MG545616	>OK546212	0,603	60,3
>BBTA2	>OK546212	0,6037924	60,4
>AF416471	>OK546212	0,6087399	60,9
>EU430730	>OK546212	0,602	60,2
>MG545617	>OK546212	0,605315	60,5
>AF216222	>OK546212	0,6059063	60,6
>EU366174	>OK546212	0,6028226	60,3
>FJ394347	>OK546212	0,6096311	61,0
>L32167	>OK546212	0,6193416	61,9
>OK546211	>OK546212	0,9808044	98,1

(B) DRC alphasatellite vs *Nanoalphasatellitinae*



First Sequence	Second Sequence	Identity Score	%
>DIES462112	>DIES462112	0,9808044	98,1
>DIES462111	>KX534408	0,635405 7	63,5
>DIES462111	>MT133684	0,6182396	61,8
>DIES462111	>MK039141	0,629118	62,9
>DIES462111	>MK039142	0,6244541	62,4
>DIES462111	>MN531183	0,6310044	63,1
>DIES462111	>AB000920	0,6130064	61,3
>DIES462111	>KC978957	0,5964546	59,6
>DIES462111	>KX534410	0,6139785	61,4
>DIES462111	>KC979051	0,6427819	64,3
>DIES462111	>KC978990	0,6343042	63,4
>DIES462111	>MH626069	0,6346983	63,5
>DIES462111	>MH626070	0,6346983	63,5
>DIES462111	>MH671345	0,6346983	63,5
>DIES462111	>MH626071	0,6357759	63,6
>DIES462111	>LC094146	0,5835095	58,4
>DIES462111	>LC010783	0,5819328	58,2
>DIES462111	>AB000921	0,5891719	58,9
>DIES462111	>AJ005968	0,5856531	58,6
>DIES462111	>MF510471	0,590958	59,1
>DIES462111	>MF510472	0,588172	58,8
>DIES462111	>MF510473	0,5869099	58,7
>DIES462111	>MK039138	0,5935484	59,4
>DIES462111	>MT133674	0,6163934	61,6
>DIES462111	>MT133681	0,6086487	60,9
>DIES462111	>U16731	0,6237732	62,4
>DIES462111	>MF535456	0,6010363	60,1
>DIES462111	>KX534399	0,5897436	59,0
>DIES462111	>KX534407	0,5867944	58,7
>DIES462111	>KX534402	0,6260504	62,6
>DIES462111	>MN273331	0,6137787	61,4
>DIES462111	>MN273341	0,6137787	61,4
>DIES462111	>KC978948	0,604888	60,5
>DIES462111	>X80879	0,6302966	63,0
>DIES462111	>KC979008	0,6108786	61,1
>DIES462111	>KC978991	0,6145833	61,5
>DIES462111	>MH626066	0,6208333	62,1
>DIES462111	>MH626067	0,6239583	62,4
>DIES462111	>MH626068	0,6239583	62,4
>DIES462111	>MH671346	0,6229166	62,3
>KX534398	>DIES462111	0,621421	62,1
>MN273342	>DIES462111	0,6132771	61,3
>U16735	>DIES462111	0,5987526	59,9
>KX534409	>DIES462111	0,6293104	62,9
>AB000922	>DIES462111	0,6066946	60,7
>LC094157	>DIES462111	0,6056485	60,6
>LC010784	>DIES462111	0,5995829	60,0
>LC094147	>DIES462111	0,6108786	61,1
>KC979052	>DIES462111	0,5968254	59,7
>MH000235	>DIES462111	0,6027689	60,3
>MH000253	>DIES462111	0,6151351	61,5
>LC010785	>DIES462111	0,6344969	63,4
>LC094155	>DIES462111	0,6382323	63,8
>LC094158	>DIES462111	0,6355216	63,6
>AB009047	>DIES462111	0,617558	61,8
>AJ005964	>DIES462111	0,6165803	61,7
>AJ132185	>DIES462111	0,6103093	61,0
>KX534397	>DIES462111	0,6072508	60,7
>MK039137	>DIES462111	0,6148069	61,5
>MT133678	>DIES462111	0,6243272	62,4
>KX534400	>DIES462111	0,6207265	62,1
>KX534405	>DIES462111	0,6255319	62,6
>KX534403	>DIES462111	0,6193001	61,9
>KX534401	>DIES462111	0,6255274	62,6
>KX534404	>DIES462111	0,6161935	61,6
>MF535455	>DIES462111	0,6107106	61,1
>MK039139	>DIES462111	0,6144958	61,4
>MT133675	>DIES462111	0,6111696	61,1
>MT133682	>DIES462111	0,6115789	61,2
>MT133676	>DIES462111	0,6060606	60,6
>MT133683	>DIES462111	0,6119874	61,2
>MK039140	>DIES462111	0,6218221	62,2
>AJ132187	>DIES462111	0,6556225	65,6
>AJ005966	>DIES462111	0,6545821	65,5
>MF510474	>DIES462111	0,6697062	67,0
>MF510475	>DIES462111	0,668693	66,9
>MK039136	>DIES462111	0,6404495	64,0
>MT133677	>DIES462111	0,6445352	64,5
>MT133685	>DIES462111	0,6424924	64,2
>MN273340	>DIES462111	0,6720812	67,2
>KX534406	>DIES462111	0,6569417	65,7

Fabenesatellite

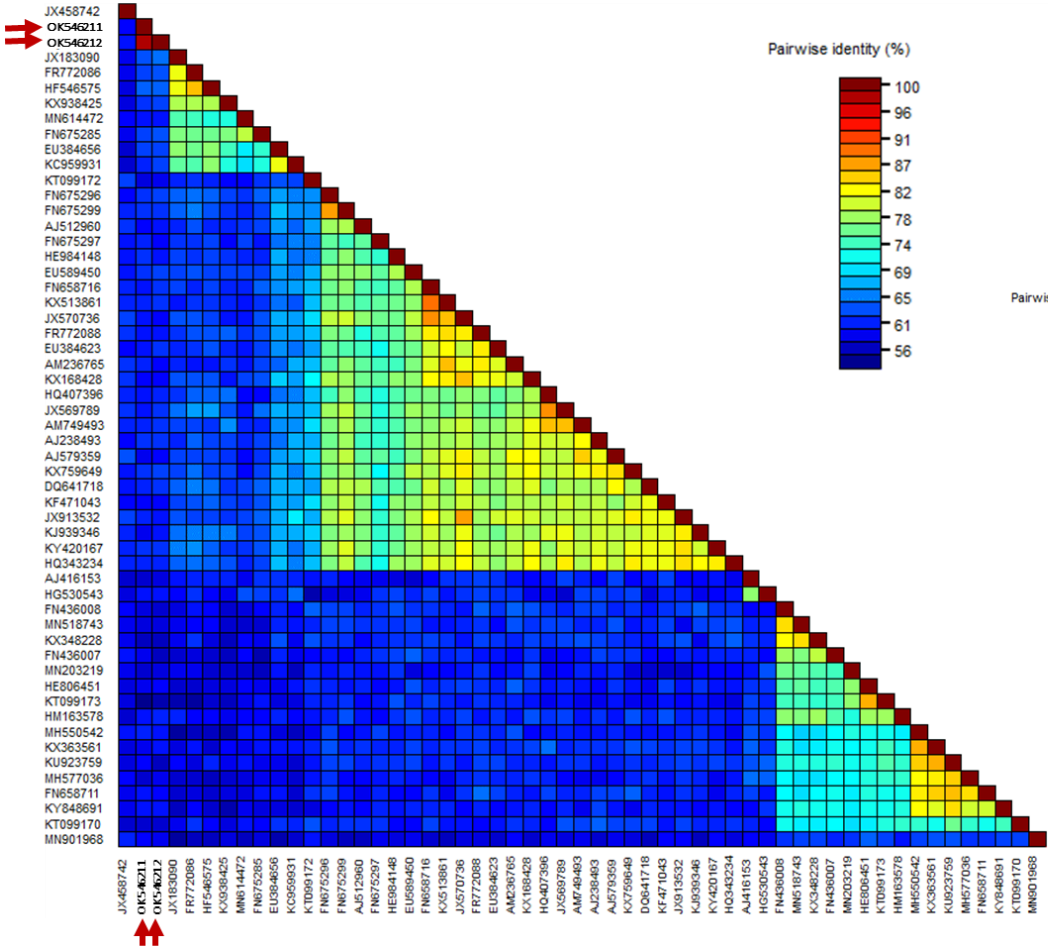
Fabenesatellite

>DIES462112	>KX534408	0,6348195	63,5
>DIES462111	>MT133684	0,6280814	62,8
>DIES462111	>MK039141	0,6280814	62,8
>DIES462111	>MK039142	0,6354626	63,5
>DIES462111	>MN531183	0,644839	64,5
>DIES462111	>AB000920	0,6192017	61,9
>DIES462111	>KC978957	0,6078224	60,8
>DIES462111	>KX534410	0,6335878	63,4
>DIES462111	>KC979051	0,642328	64,2
>DIES462111	>KC978990	0,6484716	64,8
>DIES462111	>MH626069	0,6474429	64,7
>DIES462111	>MH626070	0,6474429	64,7
>DIES462111	>MH671345	0,6474429	64,7
>DIES462111	>MH626071	0,648531	64,9
>DIES462111	>LC094146	0,5858369	58,6
>DIES462111	>LC010783	0,5774947	57,7
>DIES462111	>AB000921	0,5933477	59,3
>DIES462111	>AJ005968	0,5932755	59,3
>DIES462111	>MF510471	0,6032787	60,3
>DIES462111	>MF510472	0,6037118	60,4
>DIES462111	>MF510473	0,6002179	60,0
>DIES462111	>MK039138	0,5989071	59,9
>DIES462111	>MT133674	0,6135105	61,4
>DIES462111	>MT133681	0,6105611	61,1
>DIES462111	>U16731	0,6273429	62,7
>DIES462111	>MF535456	0,6014799	60,1
>DIES462111	>KX534399	0,5982628	59,8
>DIES462111	>KX534407	0,5982628	59,8
>DIES462111	>KX534402	0,6244681	62,4
>DIES462111	>MN273331	0,6240602	62,4
>DIES462111	>MN273341	0,6240602	62,4
>DIES462111	>KC978948	0,606405	60,6
>DIES462111	>X80879	0,6430868	64,3
>DIES462111	>KC979008	0,6092925	60,9
>DIES462111	>KC978991	0,6162791	61,6
>DIES462111	>MH626066	0,6226215	62,3
>DIES462111	>MH626067	0,6205074	62,1
>DIES462111	>MH626068	0,6205074	62,1
>DIES462111	>MH671346	0,6194503	61,9
>KX534398	>DIES462112	0,6252654	62,5
>MN273342	>DIES462112	0,6162791	61,6
>U16735	>DIES462112	0,5914316	59,1
>KX534409	>DIES462112	0,6293104	62,9
>AB000922	>DIES462112	0,6114044	61,1
>LC094157	>DIES462112	0,6166843	61,7
>LC010784	>DIES462112	0,608421	60,8
>LC094147	>DIES462112	0,6198522	62,0
>KC979052	>DIES462112	0,613846	61,5
>MH000235	>DIES462112	0,6034115	60,3
>MH000253	>DIES462112	0,5955414	59,6
>LC010785	>DIES462112	0,6200544	62,1
>LC094155	>DIES462112	0,616172	61,6
>LC094158	>DIES462112	0,6216769	62,2
>AB009047	>DIES462112	0,6146789	61,5
>AJ005964	>DIES462112	0,6006192	60,1
>AJ132185	>DIES462112	0,6118627	61,2
>KX534397	>DIES462112	0,6087845	60,9
>MK039137	>DIES462112	0,6182213	61,8
>MT133678	>DIES462112	0,6092077	60,9
>KX534400	>DIES462112	0,6205788	62,1
>KX534405	>DIES462112	0,6215352	62,2
>KX534403	>DIES462112	0,6170213	61,7
>KX534401	>DIES462112	0,6315218	63,2
>KX534404	>DIES462112	0,6208017	62,1
>MF535455	>DIES462112	0,6209424	62,1
>MK039139	>DIES462112	0,6086956	60,9
>MT133675	>DIES462112	0,6118818	61,8
>MT133682	>DIES462112	0,621656	62,2
>MT133676	>DIES462112	0,6136606	61,4
>MT133683	>DIES462112	0,6227224	62,3
>MK039140	>DIES462112	0,6196581	62,0
>AJ132187	>DIES462112	0,659919	66,0
>AJ005966	>DIES462112	0,6568528	65,7
>MF510474	>DIES462112	0,6703854	67,0
>MF510475	>DIES462112	0,6713996	67,1
>MK039136	>DIES462112	0,6456612	64,6
>MT133677	>DIES462112	0,6518595	65,2
>MT133685	>DIES462112	0,6487603	64,9
>MN273340	>DIES462112	0,6755914	67,7
>KX534406	>DIES462112	0,6602434	66,0
>DIES462111	>DIES462112	0,9808044	98,1

Fabenesatellite

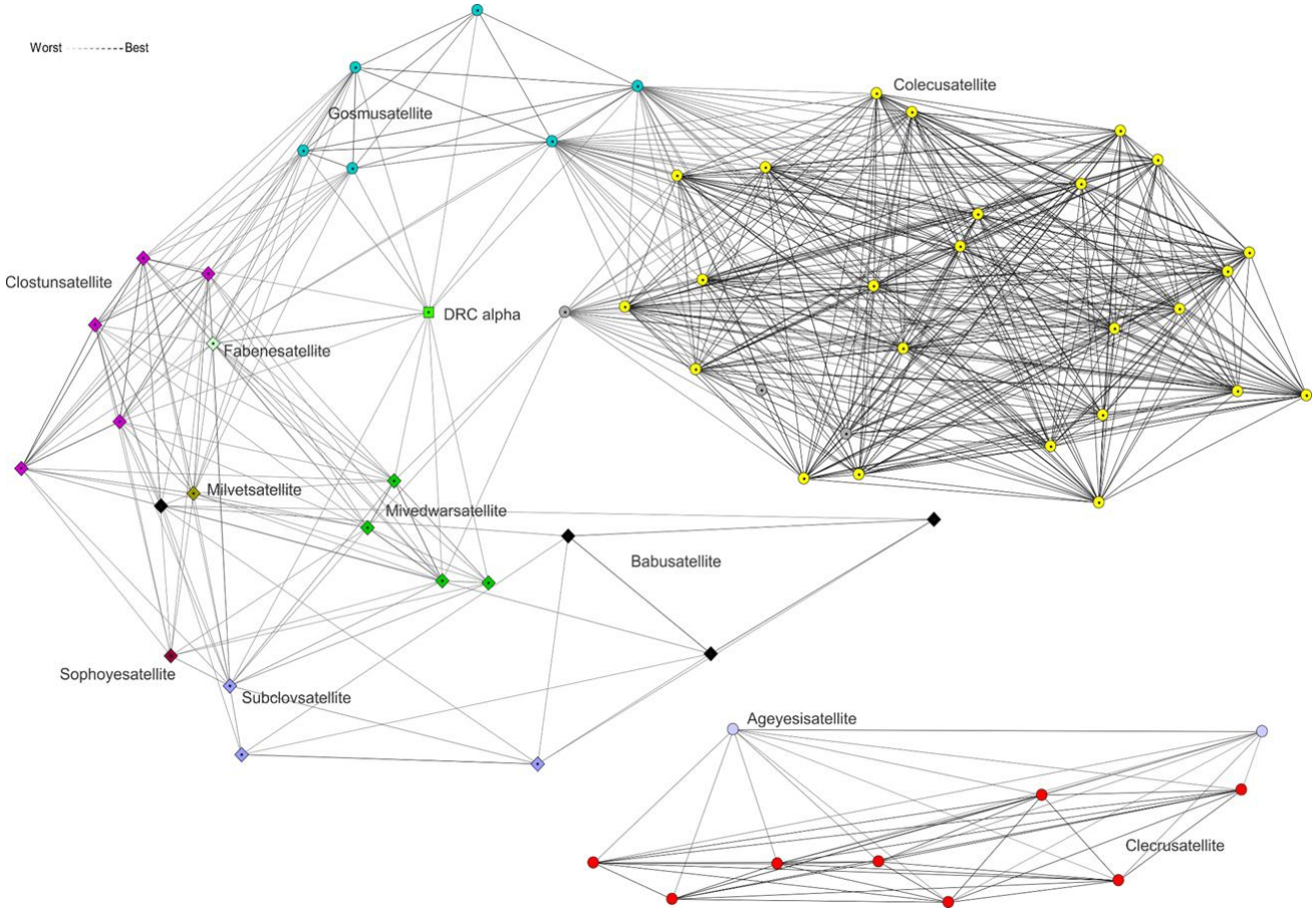
Fabenesatellite

(C) DRC alphasatellite vs *Geminialphasatellitinae*



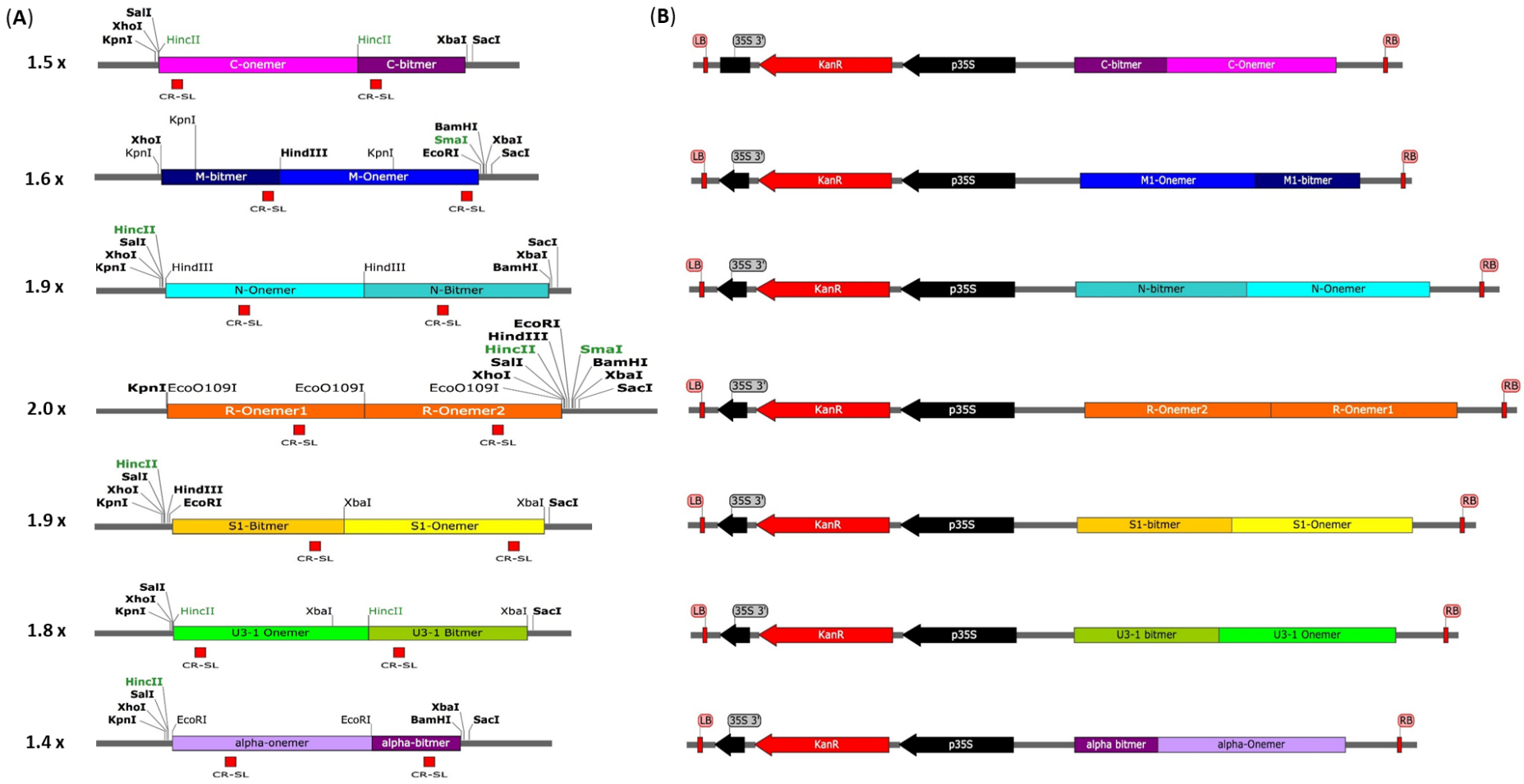
First Sequence	Second Sequence	Identity Score	%
>OK546211	**>OK546212**	0,9808044	98,08044
>OK546211	>JX183090	0,6316271	63,16271
>OK546211	>FR772086	0,6223648	62,23648
>OK546211	>HF546575	0,6363636	63,63636
>OK546211	>KX938425	0,6195256	61,95256
>OK546211	>MN614472	0,601476	60,1476
>OK546211	>FN675285	0,6203704	62,03704
>OK546211	>EU384656	0,6270066	62,70066
>OK546211	>KC959931	0,6113657	61,13657
>OK546211	>KT099172	0,5786948	57,86948
>OK546211	>FN675296	0,6190926	61,90926
>OK546211	>FN675299	0,614435	61,4435
>OK546211	>AJ512960	0,5943396	59,43396
>OK546211	>FN675297	0,6094339	60,94339
>OK546211	>HE984148	0,6146527	61,46527
>OK546211	>EU589450	0,6185958	61,85958
>OK546211	>FN658716	0,6037209	60,37209
>OK546211	>KX513861	0,6067415	60,67415
>OK546211	>JX570736	0,6043233	60,43233
>OK546211	>FR772088	0,6009259	60,09259
>OK546211	>EU384623	0,6027778	60,27778
>OK546211	>AM236765	0,6071094	60,71094
>OK546211	>KX168428	0,5917603	59,17603
>OK546211	>HQ407396	0,6027907	60,27907
>OK546211	>JX569789	0,6050186	60,50186
>OK546211	>AM749493	0,6074767	60,74767
>OK546211	>AJ238493	0,6199616	61,99616
>OK546211	>AJ579359	0,5901943	59,01943
>OK546211	>KX759649	0,6168401	61,68401
>OK546211	>DQ641718	0,606403	60,6403
>OK546211	>KF471043	0,5983225	59,83225
>OK546211	>JX913532	0,6093156	60,93156
>OK546211	>KJ939346	0,5847145	58,47145
>OK546211	>KY420167	0,6133584	61,33584
>OK546211	>HQ343234	0,6098239	60,98239
>OK546211	>AJ416153	0,571161	57,1161
>OK546211	>HG530543	0,6007569	60,07569
>OK546211	>FN436008	0,5798479	57,98479
>OK546211	>MN518743	0,5775862	57,75862
>OK546211	>KX348228	0,5667627	56,67627
>OK546211	>FN436007	0,5877712	58,77712
>OK546211	>MN203219	0,5744076	57,44076
>OK546211	>HE806451	0,5907336	59,07336
>OK546211	>KT099173	0,5442561	54,42561
>OK546211	>HM163578	0,6047198	60,47198
>OK546211	>MH550542	0,6075581	60,75581
>OK546211	>KX363561	0,5883477	58,83477
>OK546211	>KU923759	0,5977482	59,77482
>OK546211	>MH577036	0,573501	57,3501
>OK546211	>FN658711	0,5990431	59,90431
>OK546211	>KY848691	0,603235	60,3235
>OK546211	>KT099170	0,5734733	57,34733
>OK546211	>MN901968	0,5923445	59,23445
>JX458742	**>OK546211**	0,5925581	59,25581

First Sequence	Second Sequence	Identity Score	%
>OK546212	>JX183090	0,6436567	64,36567
>OK546212	>FR772086	0,6315789	63,15789
>OK546212	>HF546575	0,6418605	64,18605
>OK546212	>KX938425	0,624198	62,4198
>OK546212	>MN614472	0,6053604	60,53604
>OK546212	>FN675285	0,6283847	62,83847
>OK546212	>EU384656	0,6226591	62,26591
>OK546212	>KC959931	0,6211982	62,11982
>OK546212	>KT099172	0,5896686	58,96686
>OK546212	>FN675296	0,6203966	62,03966
>OK546212	>FN675299	0,620155	62,0155
>OK546212	>AJ512960	0,6005639	60,05639
>OK546212	>FN675297	0,595884	59,5884
>OK546212	>HE984148	0,6026365	60,26365
>OK546212	>EU589450	0,6238095	62,38095
>OK546212	>FN658716	0,6145251	61,45251
>OK546212	>KX513861	0,6170412	61,70412
>OK546212	>JX570736	0,597561	59,7561
>OK546212	>FR772088	0,5998143	59,98143
>OK546212	>EU384623	0,6158878	61,58878
>OK546212	>AM236765	0,5981482	59,81482
>OK546212	>KX168428	0,5918944	59,18944
>OK546212	>HQ407396	0,610438	61,0438
>OK546212	>JX569789	0,6155285	61,55285
>OK546212	>AM749493	0,6156752	61,56752
>OK546212	>AJ238493	0,6137866	61,37866
>OK546212	>AJ579359	0,592215	59,2215
>OK546212	>KX759649	0,623216	62,3216
>OK546212	>DQ641718	0,6151659	61,51659
>OK546212	>KF471043	0,5981221	59,81221
>OK546212	>JX913532	0,6049383	60,49383
>OK546212	>KJ939346	0,6013011	60,13011
>OK546212	>KY420167	0,6141509	61,41509
>OK546212	>HQ343234	0,6110591	61,10591
>OK546212	>AJ416153	0,5785513	57,85513
>OK546212	>HG530543	0,6017192	60,17192
>OK546212	>FN436008	0,5745489	57,45489
>OK546212	>MN518743	0,5815534	58,15534
>OK546212	>KX348228	0,5633397	56,33397
>OK546212	>FN436007	0,5628572	56,28572
>OK546212	>MN203219	0,5797517	57,97517
>OK546212	>HE806451	0,5957447	59,57447
>OK546212	>KT099173	0,5570342	55,70342
>OK546212	>HM163578	0,6092978	60,92978
>OK546212	>MH550542	0,6038835	60,38835
>OK546212	>KX363561	0,6028985	60,28985
>OK546212	>KU923759	0,5688623	56,88623
>OK546212	>MH577036	0,5860735	58,60735
>OK546212	>FN658711	0,5929119	59,29119
>OK546212	>KY848691	0,6050339	60,50339
>OK546212	>KT099170	0,5764023	57,64023
>OK546212	>MN901968	0,5846007	58,46007
>JX458742	**>OK546212**	0,6052142	60,52142
>OK546211	**>OK546212**	0,9808044	98,08044



S8 Fig. Clustering analysis of Rep proteins encoded by DRC-2016 alphasatellite and alphasatellites associated with ssDNA viruses of the families *Nanoviridae, Metaxiviridae* and *Geminiviridae*.

Rep proteins of the alphasatellites classified by Briddon et al. (2018) (see Table 1 of that paper for the NCBI accession numbers) were compared "all-against-all" and clustered using CLANS (Frickey and Lupas 2004). DRC alphasatellite (DRC alpha) is indicated with a bright green cube close to the middle and its evolutionary relatedness (connection) to other alphasatellites is shown with solid grey lines whose color intensity - from lightest to darkest - indicates the strength of connections from worse (no direct connection) to best. Each alphasatellite genus is color-coded and named. Grey circles indicate unassigned alphasatellites. Note that the genus *Babusatellite* comprising four species (indicated with black diamonds) was recently split into two genera: *Babusatellite* and *Muscarsatellite*.



S9 Fig. Construction of BBTV and alphasatellite infectious clones.

The complete sequences of BBTV components and alphasatellite from the DRC-2016 aphids (sample JGF-5, Supplementary Figure S2) were cloned as partial (DNAs C, M_v1, N, S_v1, U3_v1) or complete (DNA-R) dimers first in pBluescript SK(-) (A) and then in pCambia2300 (B). For each construct, positions of the duplicated common region stem loop (CR-SL) are indicated with red quadrates.

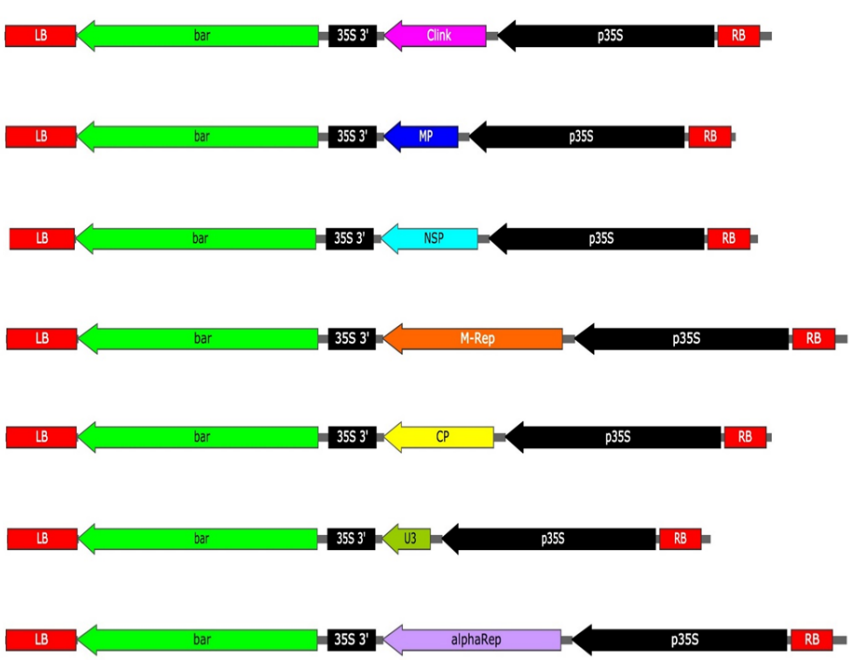
S10 Fig. Analysis of BBTV and alphasatellite proteins for silencing suppression and chlorosis induction activities in *Nicotiana benthamiana*.

(A-B) Subcloning of BBTV and alphasatellite ORFs from partial dimer constructs (A) (see Supplementary Figure S9) into the binary vector pB7WG2 under the control of cauliflower mosaic virus 35S promoter (35S) and terminator (35S 3') (B) by PCR amplification with viral ORF-specific primers carrying AttB1 and AttB2 recombination sites (Supplementary Table S1), followed by Gateway recombination of the PCR products into the vector. (C-D) Screening of BBTV and alphasatellite ORFs for silencing suppression and chlorosis induction following infiltration of the binary constructs in leaves of *N. benthamiana* GFP-transgenic (16c line) plants at 8 days post infiltration (dpi) under ultraviolet (UV) light (C) and day light (D). Pictures of representative leaves (from three infiltrated plants per construct, three leaves per plant) are shown. The experiment was repeated 3 times with similar results. Note that in addition to the BBTV Rep construct inducing strong chlorosis, the BBTV NSP construct induced weaker chlorosis, albeit it was barely visible in some of the infiltrated leaves.

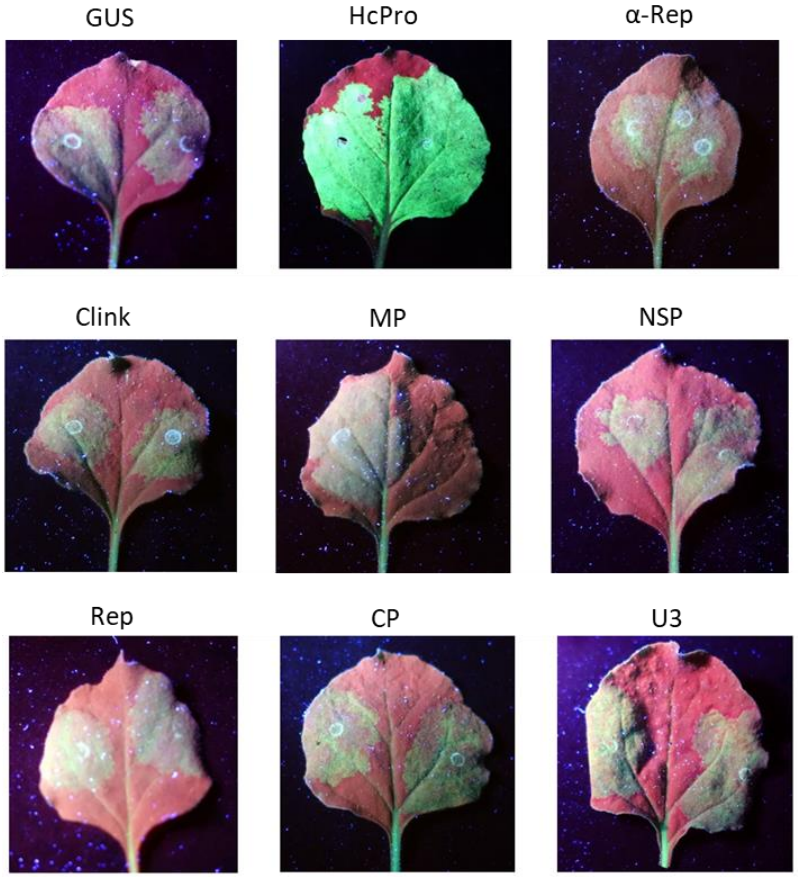
(**A**) Positions of viral ORFs and PCR primers carrying AttB1 and AttB2 recombination sites



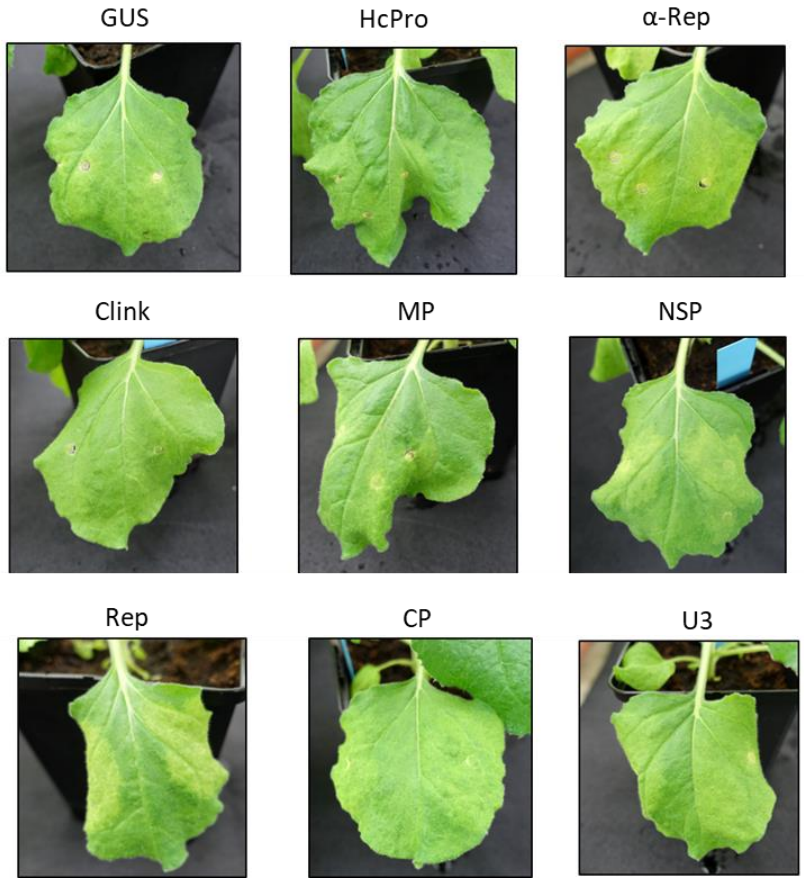
(**B**) Organization of viral ORF expression cassettes within T-DNA of the binary vector pB7WG2

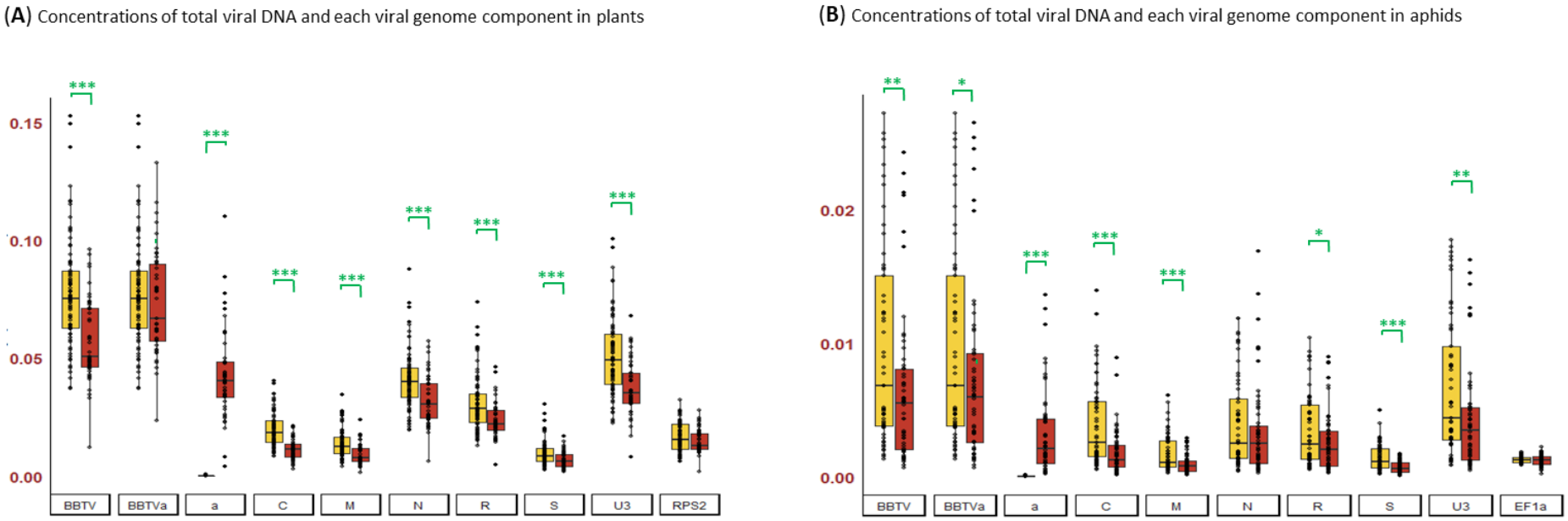


(**C**) Infiltrated leaves of *N. benthamiana* under UV light



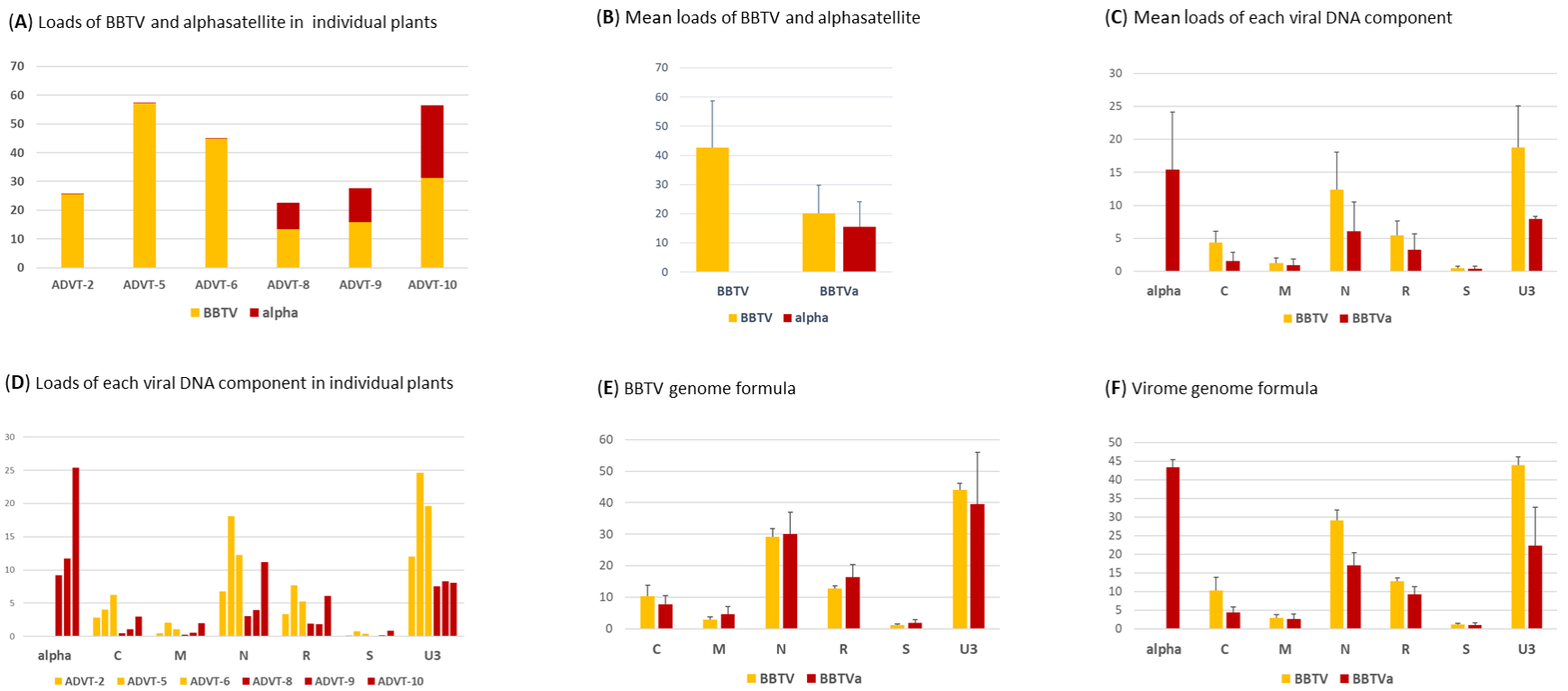
(**D**) Infiltrated leaves of *N. benthamiana* under day light





S11 Fig. Impact of DRC alphasatellite on viral DNA concentrations in plants and aphids.

Viral DNA concentrations for each BBTV component (C, M, N, R, S, U3) and alphasatellite (a) and for total helper virus without (BBTV) or with (BBTVa) alphasatellite were measured by quantitative PCR using equal amounts of total DNA in (A) plants without (n = 77) and with (n = 47) alphasatellite (all experiments) and (E) aphids without (n = 58) and with (n = 56) alphasatellite (all experiments). Concentrations of total host DNA in the same samples were measured by qPCR with primers specific for the banana and aphid housekeeping genes (RPS2 and eEF1a, respectively). * Kruskal-Wallis $P < 0.05$. *** Kruskal-Wallis $P < 0.005$.



S12 Fig. Viral DNA loads and formulas in BBTV-infected Cavendish banana plants with or without DRC alphasatellite selected for viral transcriptome and small RNA-ome profiling.

Viral DNA loads for each BBTV component (C, M, N, R, S, U3) and alphasatellite (alpha) as well as for total helper virus (BBTV) were measured by quantitative PCR (using the banana RPS2 gene as internal control for normalization) in 3 plants infected with BBTV alone (ADVT-2, ADVT-5, ADVT-6) and 3 plants co-infected with BBTV and alphasatellite (ADVT-8, ADVT-9, ADVT-10). (A) Loads of helper virus (BBTV, yellow) and alphasatellite (alpha, red) DNA in plants without and with alphasatellite. (B) Mean loads of helper virus (BBTV, yellow) and alphasatellite (alpha, red) DNA in plants without (BBTV) and with (BBTVa) alphasatellite. (C) Mean loads of each BBTV component and alphasatellite (alpha) DNA in plants with (red) and without (yellow) alphasatellite. (D) Loads of each BBTV component and alphasatellite (alpha) in plants without (yellow) and with (red) alphasatellite. (E) Helper virus genome formula in plants without (yellow, BBTV) and with (red, BBTVa) alphasatellite. (F) Virome genome formula in plants without (yellow, BBTV) and with (red, BBTVa) alphasatellite. Error bars represent standard deviations.

S13 Fig. Single nucleotide resolution maps of Illumina mRNA-seq reads representing viral mRNAs and other viral transcripts from BBTV-infected Cavendish banana plants with or without DRC alphasatellite and identification of poly(A) sites by analysis of the mapped reads.

For each of the two conditions, i.e. without (BBTV-alpha) and with (BBTV+alpha) alphasatellite, Illumina 75 nt reverse reads of the three biological replicates (leaf tissues of three plants) were combined and mapped simultaneously onto the reference sequences of six BBTV components (-/+ alphasatellite). Histograms plot the numbers of viral 75 nt sense and antisense reads at each nucleotide position of the 1018-to-1111 nt BBTV genome components (DNAs C, M, N, R, S, U3 - subpanels A1-F1) and 1105 nt alphasatellite (Alpha – subpanel G1): blue bars above the axis represent sense reads starting at each respective position, while red bars below the axis represent antisense reads ending at each respective position. The genome organizations of BBTV components and alphasatellite are shown schematically above the respective histograms, with the Pol II promoter (TATA-box and transcription start site, TSS) and terminator (polyA signal, PAS) elements indicated in pink, capped and polyadenylated mRNA shown as solid blue lines, viral protein-coding ORFs boxed and their nucleotide positions given. In each panel (A-G), subpanels 2 and 3 show the Pol II terminator region with the mapped poly(A) sites and the respective poly(A) signals at the upstream positions visualized using MISIS-2 (Seguin et al. 2016). Note that the mapped mRNA reads ending with oligo(A) tails generate A-SNPs in the sequence logo just downstream of the poly (A) site (i.e., pre-mRNA cleavage and polyadenylation site).

Figure S13 **A1**

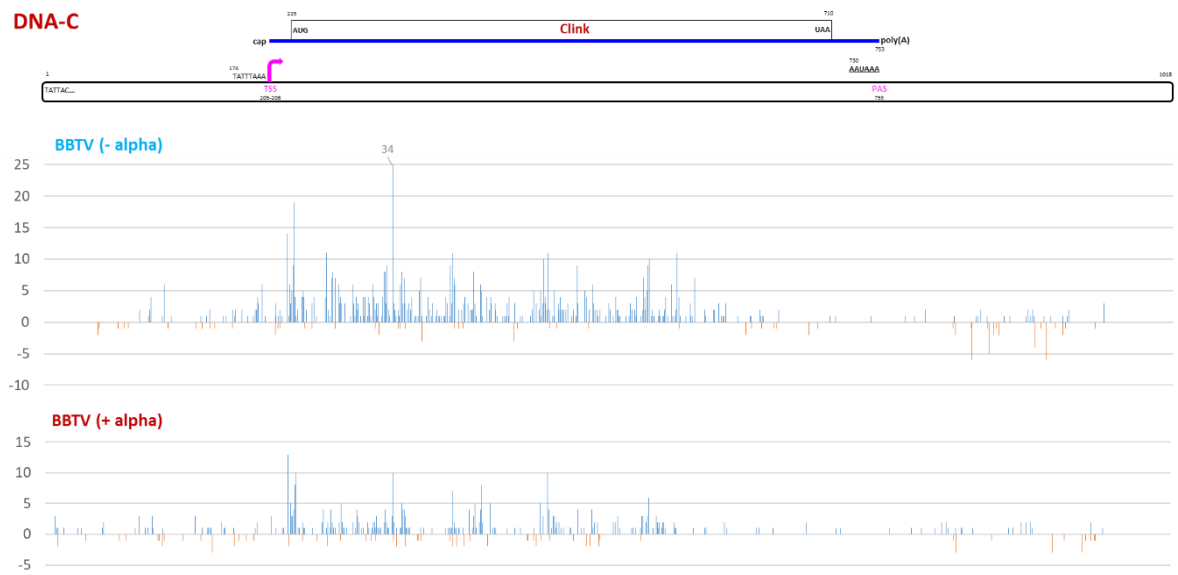


Figure S13 **A2**

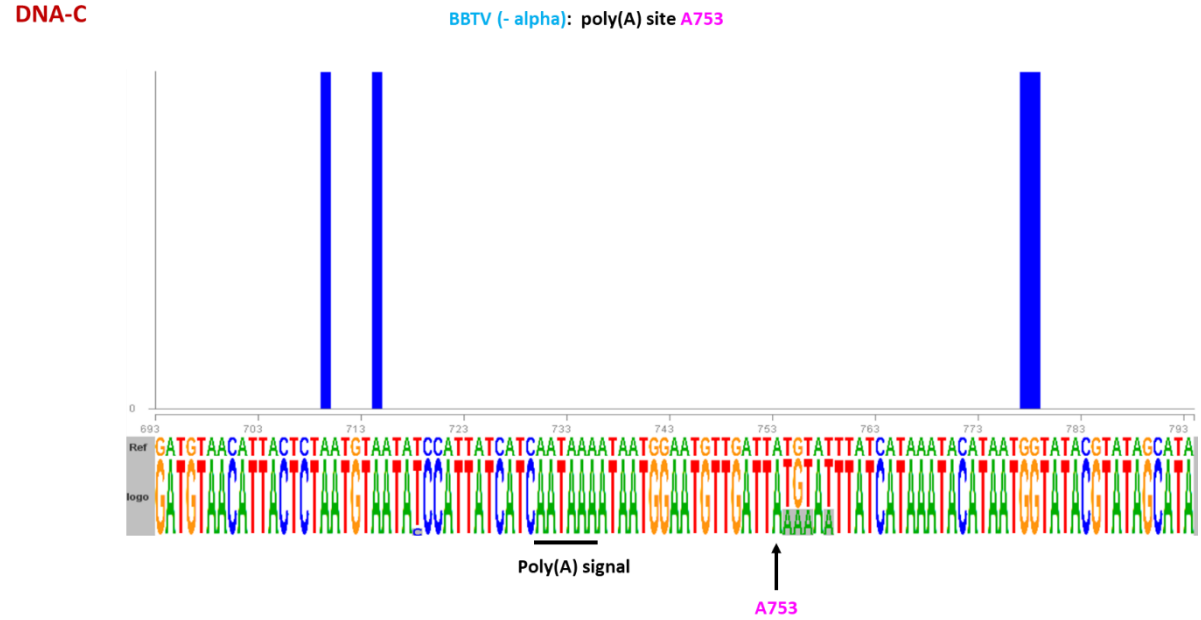


Figure S13 **B1**

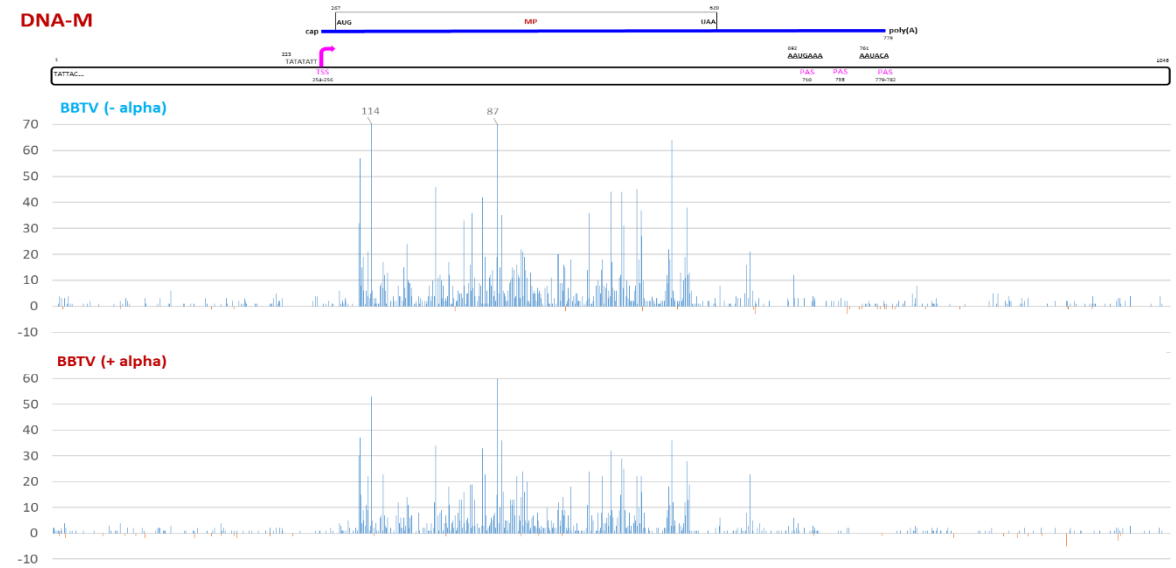


Figure S13 **B2**

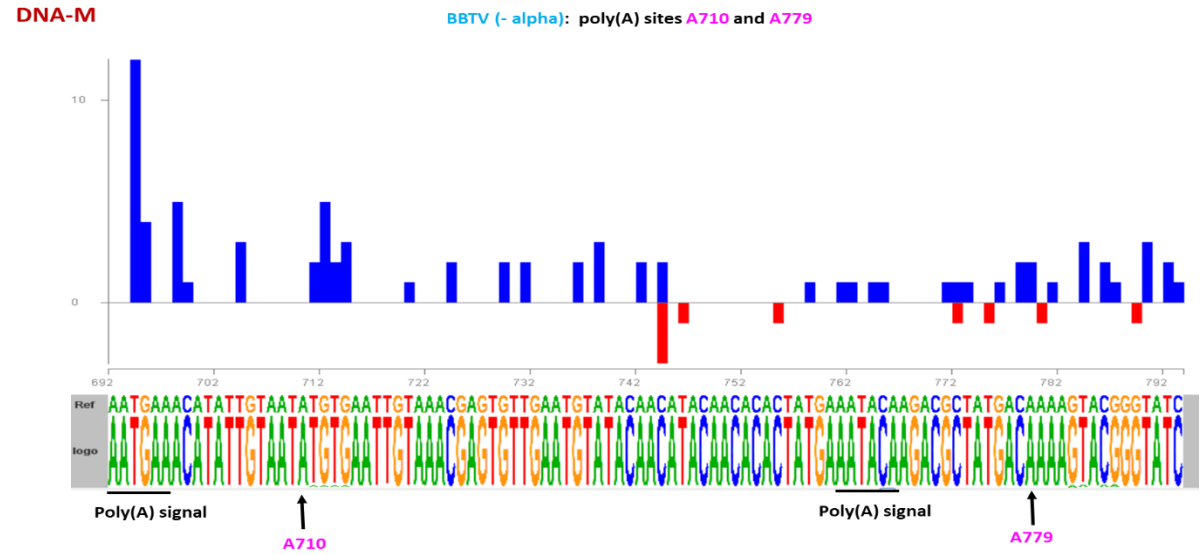


Figure S13 **B3**

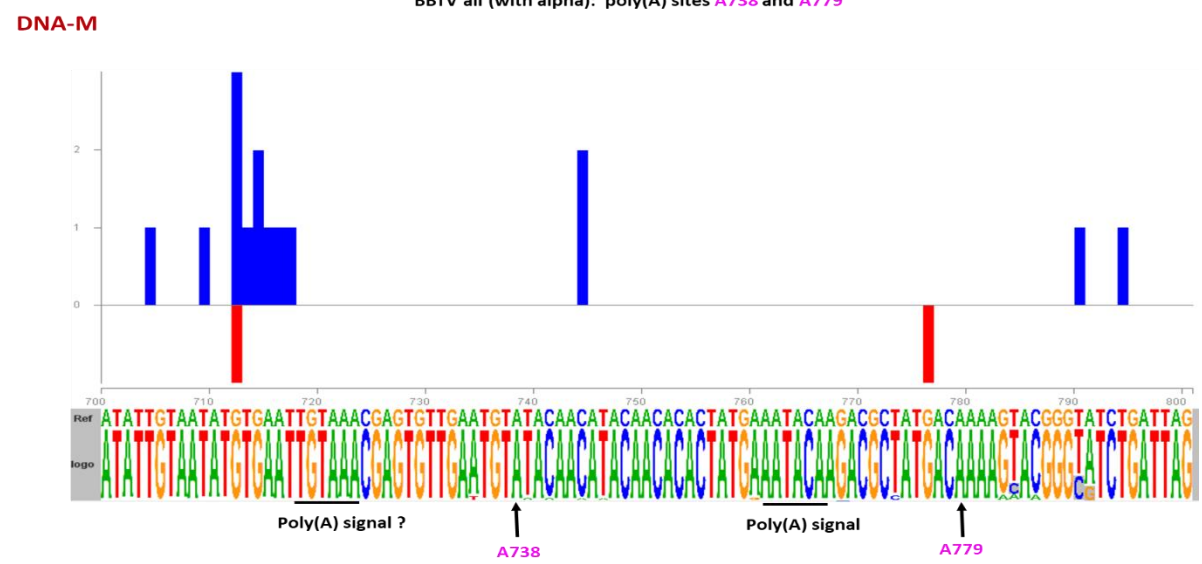


Figure S13 **C1**

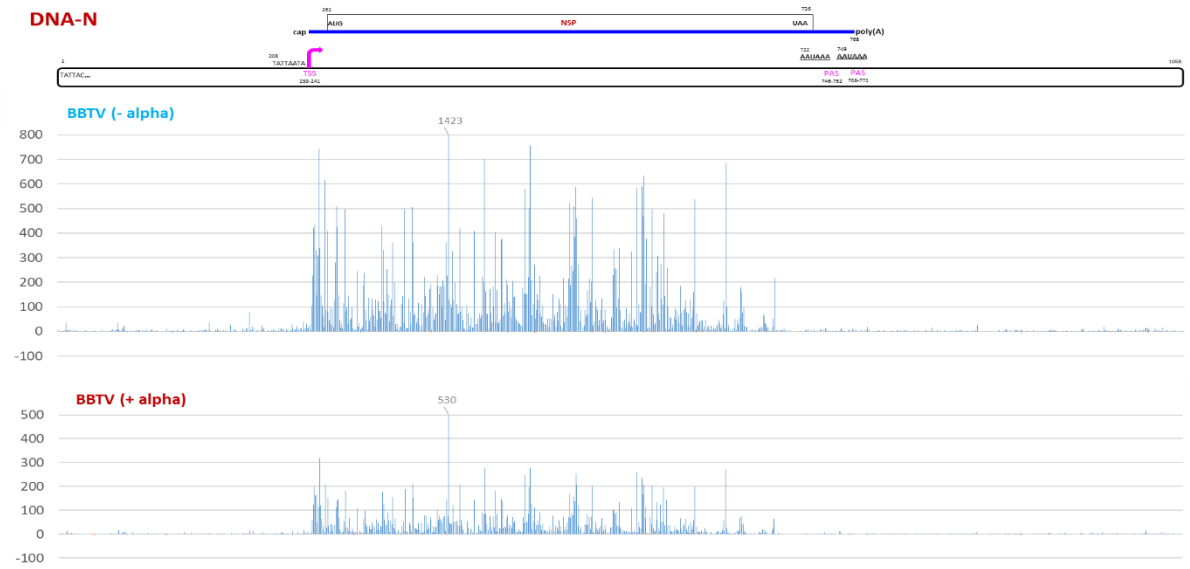


Figure S13 **B2**

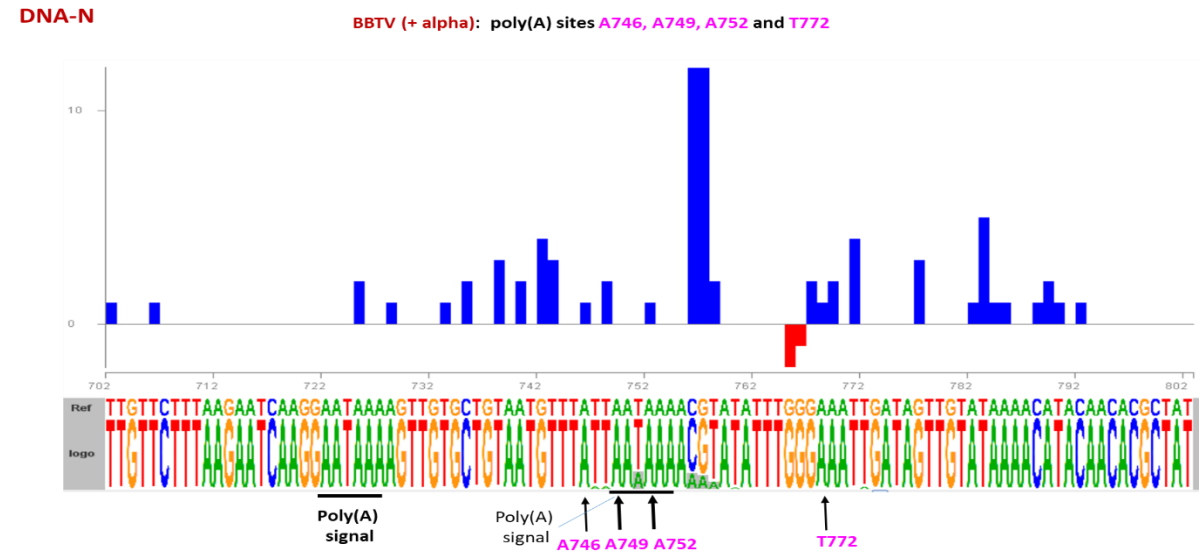


Figure S13 **B2**

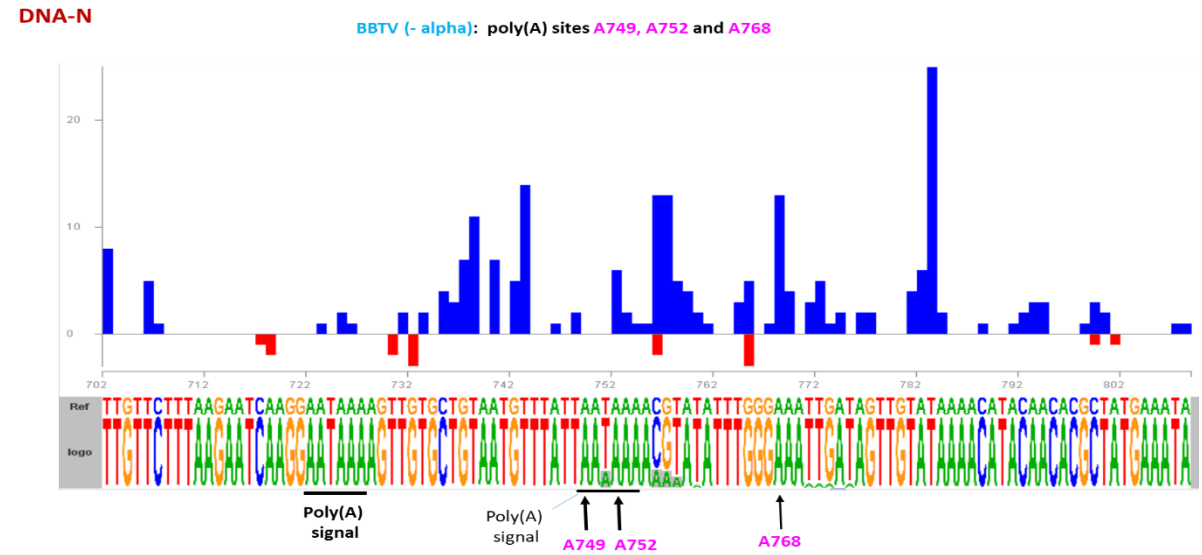


Figure S13 **D1**

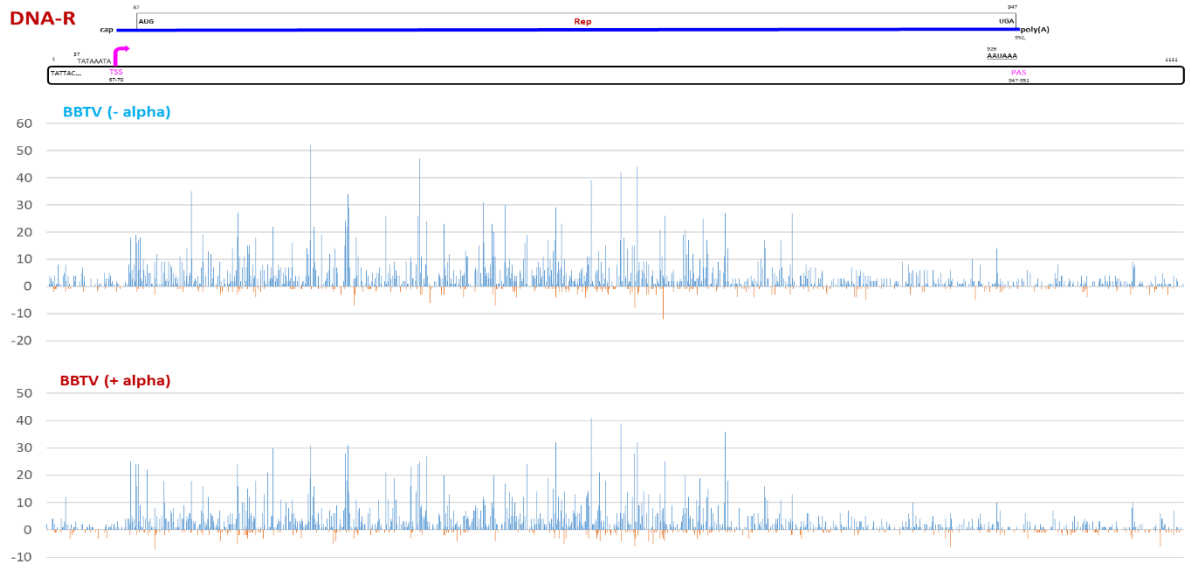


Figure S13 **D2**

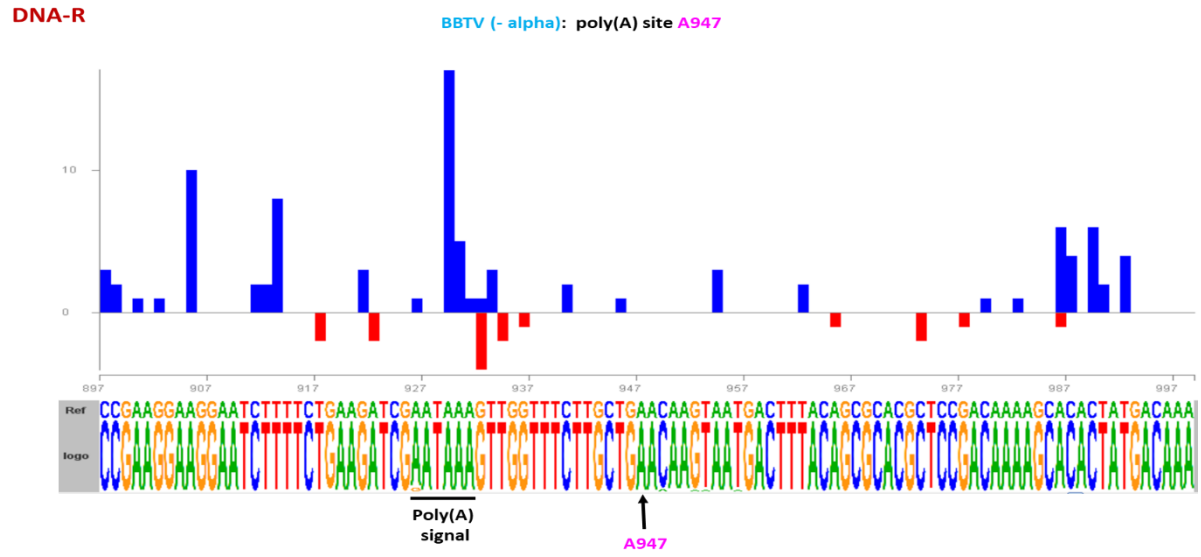


Figure S13 **D3**

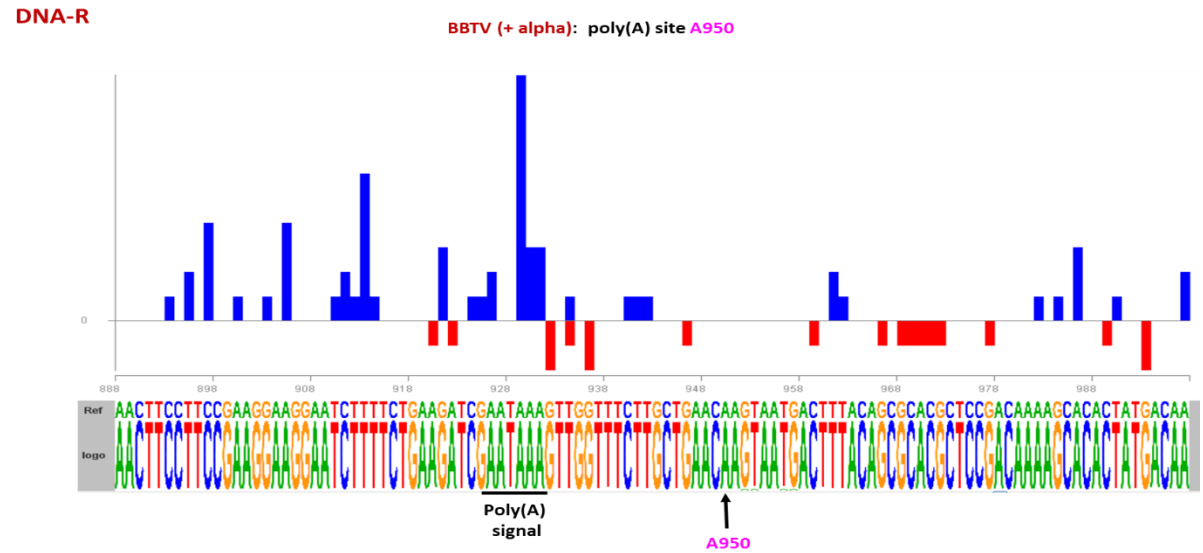


Figure S13 **E1**

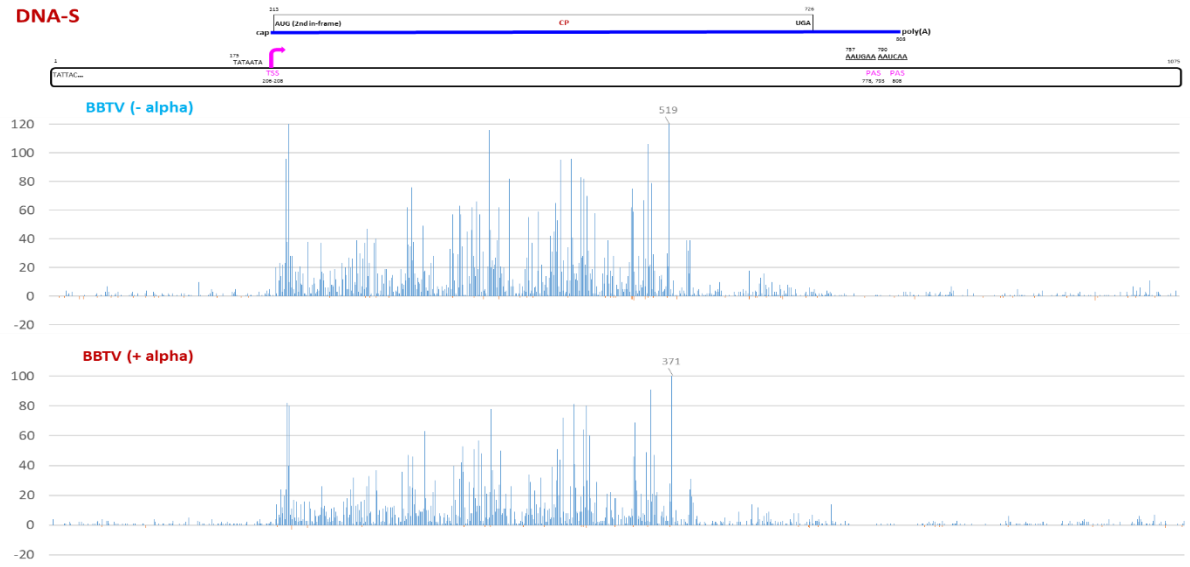


Figure S13 **E2**

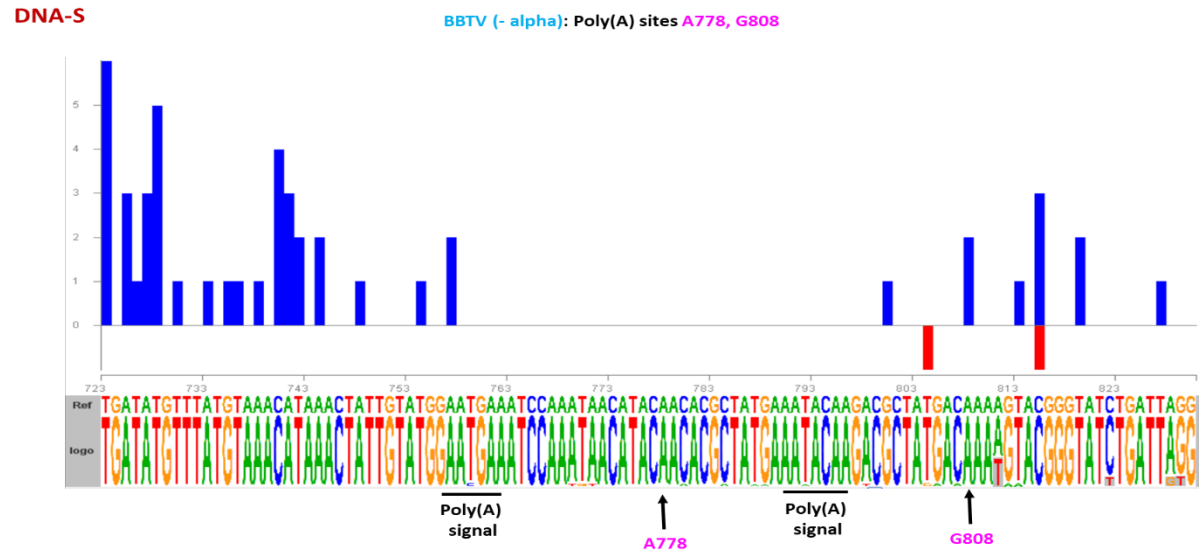


Figure S13 **E3**

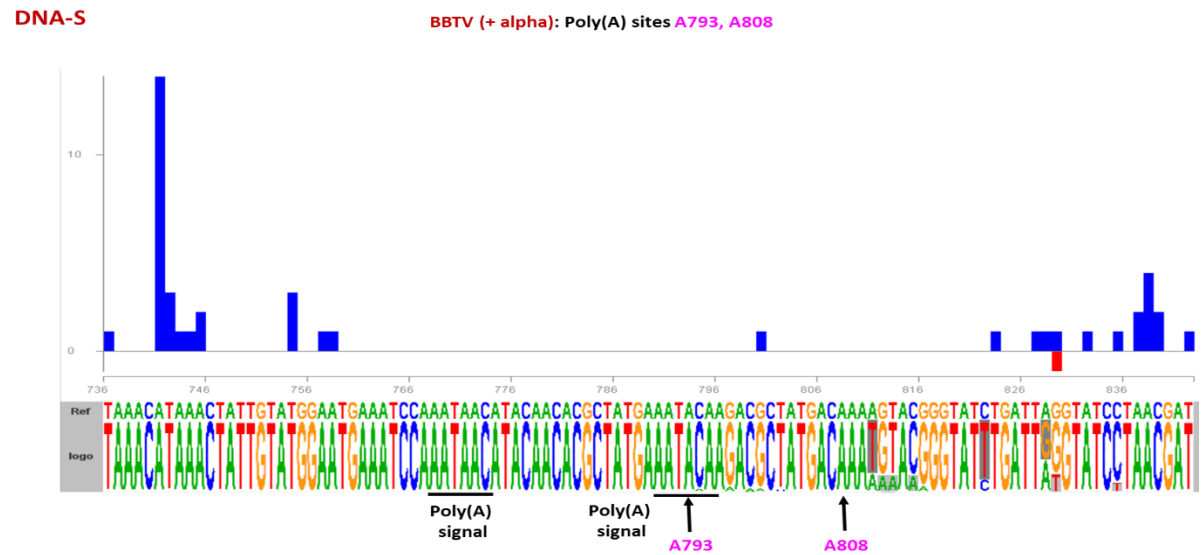


Figure S13 **F1**

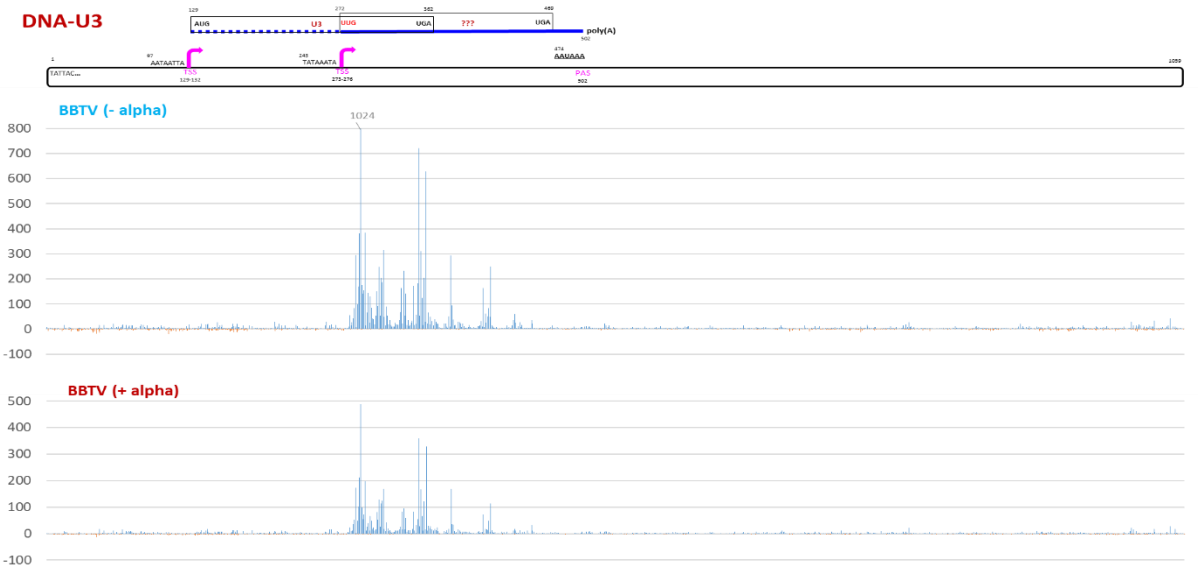


Figure S13 **F2**

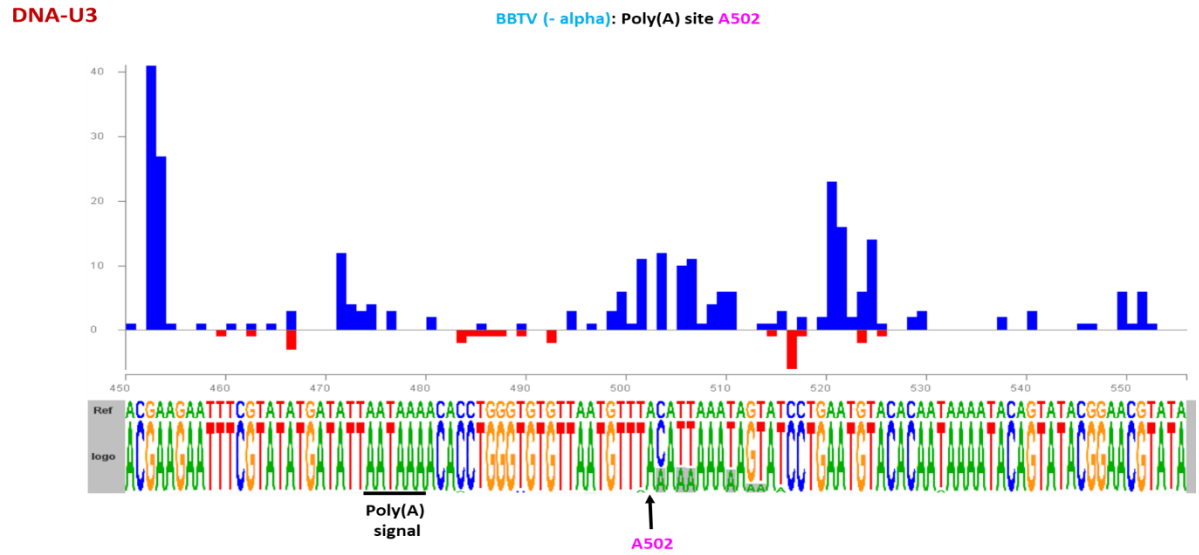


Figure S13 **F3**

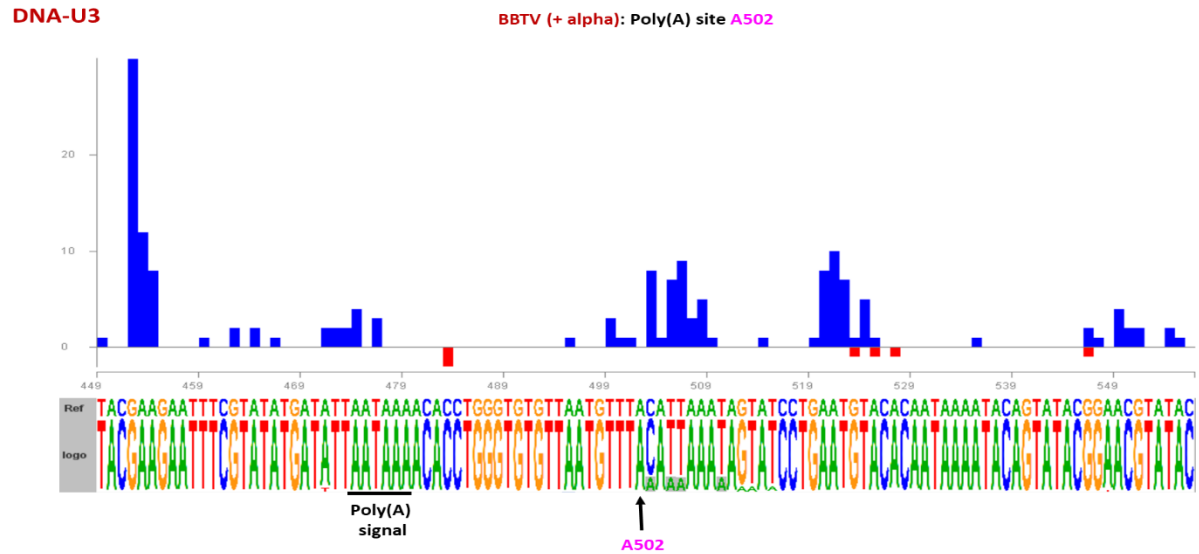


Figure S13 **G1**

Alpha

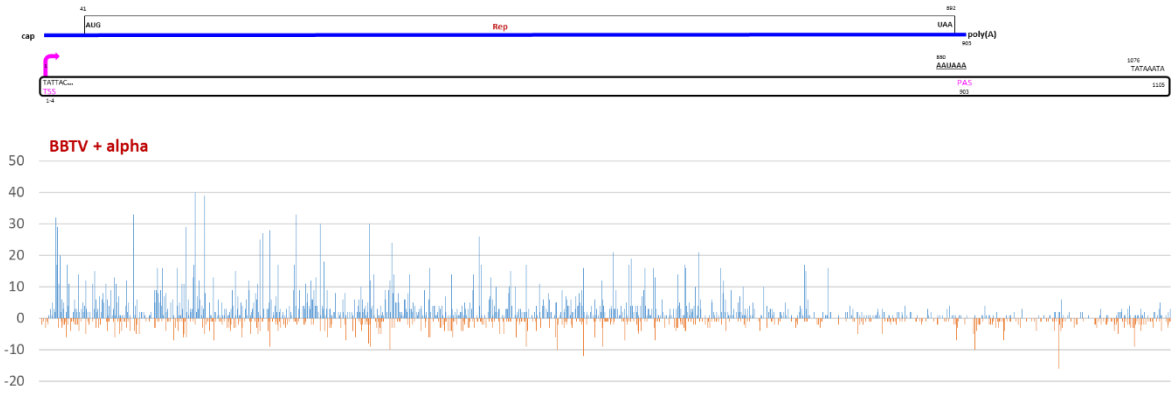
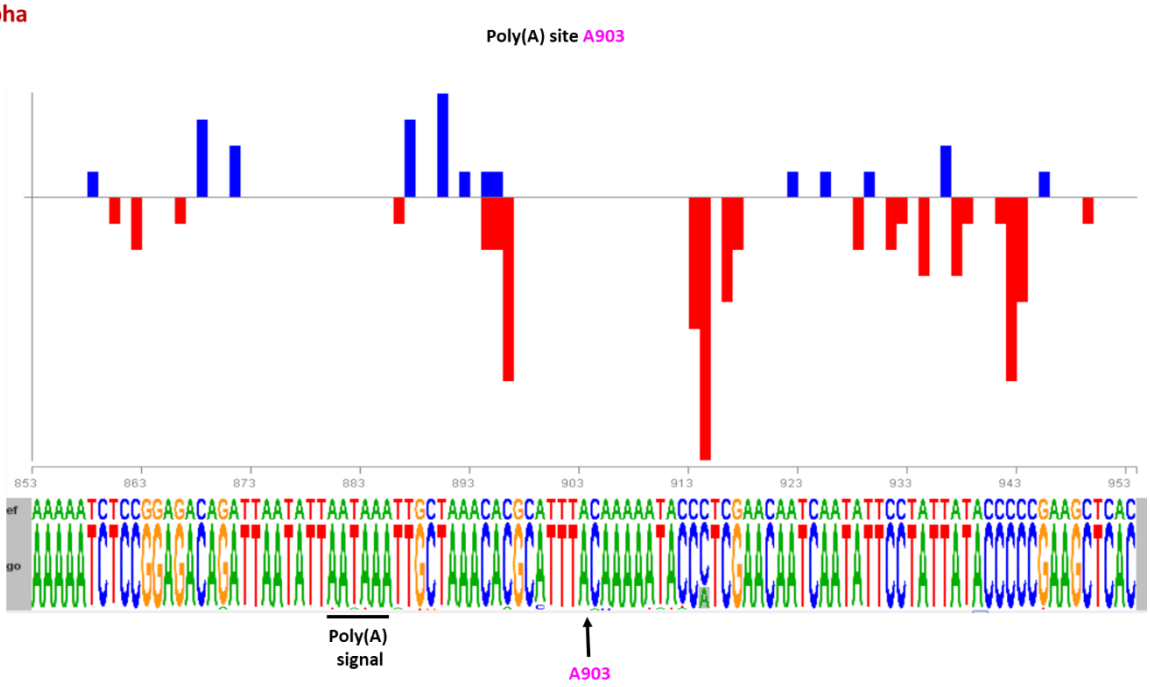
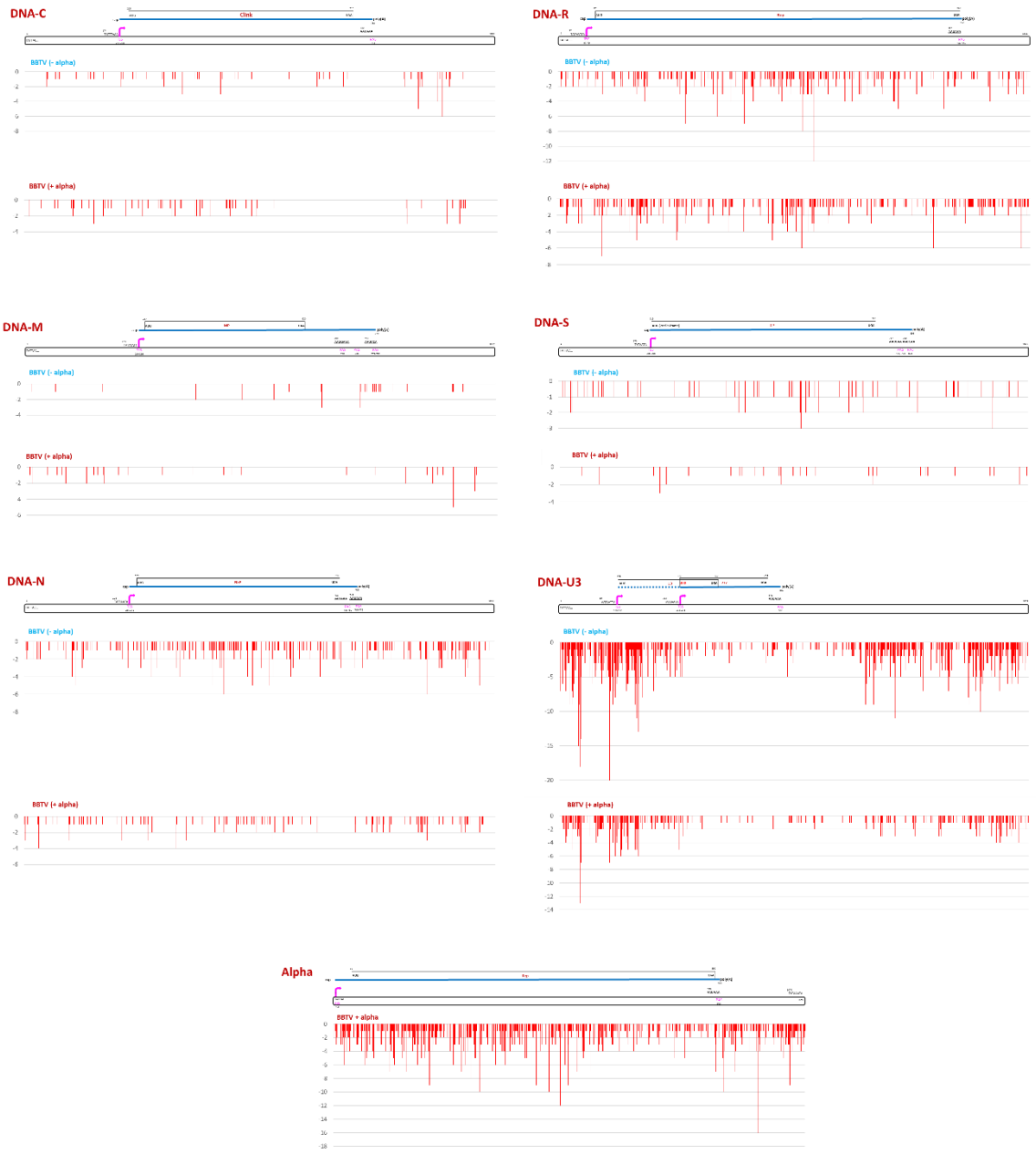


Figure S13 **G2**

Alpha





S14 Fig. Single nucleotide resolution maps of Illumina mRNA-seq reads representing viral antisense transcripts from BBTV-infected Cavendish banana plants with or without DRC alphasatellite.

For each of the two conditions, i.e. without (BBTV-alpha) and with (BBTV+alpha) alphasatellite, Illumina 75 nt reverse reads of the three biological replicates (leaf tissues of three plants) were combined and mapped simultaneously onto the reference sequences of six BBTV components (-/+ alphasatellite). Histograms plot the numbers of viral 75 nt antisense reads at each nucleotide position of the 1018-to-1111 nt BBTV genome components (DNAs C, M, N, R, S, U3) and 1105 nt alphasatellite (Alpha): red bars below the axis represent antisense reads ending at each respective position. The genome organizations of BBTV components and alphasatellite are shown schematically above the histograms, with the Pol II promoter (TATA-box and transcription start site, TSS) and terminator (polyA signal, PAS) elements indicated in pink, capped and polyadenylated mRNA shown as solid blue lines, viral protein-coding ORFs boxed and their nucleotide positions given.

S1 Table. Primers used for immune-capture (IC)-PCR, single, duplex and multiplex PCR and quantitative (q)PCR analyses.

Name original	Specific to / strand	Sequence (5' to 3')	Size of Amplicon
IC-PCR and duplex PCR for detection of DRC-2016 alphasatellite and BBTV			
BBTV a-sat_F	DRC alphasatellite forward	AATTCGGAGAAATTCTGCTGAAGG	187 bp
BBTV a-sat_R	DRC alphasatellite reverse	GGTTTTTTGCCACGAACGTA	
V_BBTV_R_F	DNA-R forward	AGATGAGAGGCTTCTTCCCAG	494 bp
V_BBTV_R_R	DNA-R reverse	AATATCCTCGTAATTATACAGTCTACA	
Diagnostic single PCR for detection of DRC-2016 and DRC-2012 alphasatellite			
Alpha_NcoI_F	DRC alphasatellite forward	GCCATGGATGAATCTCGGAA	285 bp
a-sat V_R	DRC alphasatellite reverse	CTTTCGTACACCTGGGGAAA	
Multiplex PCR for DRC-2016 alphasatellite and BBTV			
Alpha_NcoI_F	DRC alphasatellite forward	GCCATGGATGAATCTCGGAA	285 bp
a-sat V_R	DRC alphasatellite reverse	CTTTCGTACACCTGGGGAAA	
V_BBTV_C_F	DNA-C forward	CGAGATGCGAAAATGGAGGC	355 bp
V_BBTV_C_R	DNA-C reverse	TGATAAAAGAACACCATGCTGTTTA	
V_BBTV_M_F	DNA-M forward	GATATAAGGGACATAACGGGTTCAG	217 bp
V_BBTV_M_R	DNA-M reverse	ATGCCATTTAGCAGGGTCCTATT	
V_BBTV_N_F	DNA-N forward	GTCAGCAATGATTATAATGGGCTTT	540 bp
V_BBTV_N_R	DNA-N reverse	CTGTTACTAATACTAACTTCCATGTCT	
V_BBTV_R_F	DNA-R forward	AGATGAGAGGCTTCTTCCCAG	494 bp
V_BBTV_R_R	DNA-R reverse	AATATCCTCGTAATTATACAGTCTACA	
V_BBTV_S_F	DNA-S forward	GGGCTAATGGATTGTGGATATAG	184 bp
V_BBTV_S_R	DNA-S reverse	TTTGTTGTCTGAACATCTATCTAC	
V_BBTV_U3_F	DNA-U3 forward	CGGCCCAATAATTAAGAGAACG	412 bp
V_BBTV_U3_R	DNA-U3 reverse	TAAACATTAACACACCCAGG	
Degenerate primers for detection of fabenesatellites and DRC alphasatellite			
deg_Fabenesat_s	alphasatellite forward	CAYGACCAYMTHCAGGGAGT	296 bp
deg_Fabenesat_as	alphasatellite reverse	CGAGATKCDTCCATGGCGGA	
qPCR primers for DRC-2016 alphasatellite and BBTV			
DRC-alphasatellite	DRC alphasatellite forward	TCACGACCACCTACAGGGAG	85 bp
	DRC alphasatellite reverse	GTGTGGATTTCCCCCGATCA	
BBTV DNA-C	DNA-C forward	CGAGATGCGAAAATGGAGGC	96 bp
	DNA-C reverse	TGGCAGACGATTCCCAGAAC	
BBTV DNA-M	DNA-M forward	GGCTGTCATACCACAGGCAT	88 bp
	DNA-M reverse	CATAGGTCCGACGTTTCCTCG	
BBTV DNA-N	DNA-N forward	CTGAGATGCCTGTGGTATGAC	103 bp
	DNA-N reverse	CCGATAATCGACAATATGTGC	
BBTV DNA-R	DNA-R forward	CAACAATCCCACAACACTACC	103 bp
	DNA-R reverse	AACATAACCTTGCACATGACG	
BBTV DNA-S	DNA-S forward	CTTGGGAAATCAACCAGCCG	109 bp
	DNA-S reverse	CCTGCTTCAAGTTCCCCACT	
BBTV DNA-U3	DNA-U3 forward	ACGGACCGAAATACTCCTGC	88 bp
	DNA-U3 reverse	CAGCACAACCACCTTGACAG	
Banana RPS2	RPS2 forward	TAGGGATTCCGACGATTTGTTT	84 bp
	RPS2 reverse	TAGCGTCATCATTGGCTGGGA	
Aphid EF1α	EF1α forward	AGATTGGACAAACCCGTGAA	96 bp
	EF1α reverse	GCTGTATGGTGGTTCAGTAGAG	
Primers for cloning DRC-2016 alphasatellite and BBTV components			
RR1 C-FL HincII s	DNA-C forward	actgGTCAACATCAATAAAGAAGAAGGAATATTC	

RR2 C-FL HincII as	DNA-C reverse	actgGTTGACGTCTATCTTTCGCTGC
RR3 C-bit BglII s	DNA-C forward bitmer	actgAGATCTCTATTGTGATGAGGTTC
RR4 M-FL HindIII s	DNA-M forward	gtcaAAGCTTGGCAACCACCACTTTAG
RR5 M-FL HindIII as	DNA-M reverse	gtcaAAGCTTATCCGACACAAATATACTC
RR6 M-bit HincII as	DNA-M reverse bitmer	gtcaGTTAACCTGAGATGCCTGTGG
RR7 N-FL HindIII s	DNA-N forward	cagtAAGCTTCAGCGGAAATAATAGGAAC
RR8 N-FL HindIII as	DNA-N reverse	cagtAAGCTTCCGTTAAACACAGCTTCAA
RR9 N-bit ScaI s	DNA-N forward bitmer	cagtAGTACTCATAGTACCATTCGGTGG
RR10 R-FL Eco01 s	DNA-R forward	agtcAGGTCCCTTCGAGTTTGGTGCATTTA
RR11 R-FL Eco01 as	DNA-R reverse	agtcGGGACCTTCGATTCTTGTATCTTCC
RR13 S-FL XbaI s	DNA-S forward	ctagATCTAGAAGTCTATAAATACCAGTGTG
RR14 S-FL XbaI as	DNA-S reverse	ctagTCTAGATGGCTTGCAGGGCAAGC
RR15 S-bit SspI s	DNA-S forward bitmer	ctagAATATTGGTTCCTGAAAACACCGTC
RR16 U3-FL HincII s	DNA-U3 forward	agtcGTCAACATTATTCTGGCTTGCGC
RR17 U3-FL HincII as	DNA-U3 reverse	agtcGTTGACCGGTTACCTTGACCTTC
RR18 U3-bit XbaI s	DNA-U3 forward bitmer	agtcTCTAGAGAGGCGGTGAAGCAAAC
RR19 a-FL EcoRI s	DNA-alpha forward	gtcaGAATTCGGAGAAATTCTGCTGAAGG
RR20 a-FL EcoRI as	DNA-alpha reverse	gtcaGAATTCCCACGGACCTGCGATC
RR21 a-bit SspI s	DNA-alpha bitmer	gtcaAATATTCCTATTATACCCCCGAAGCTC

Primers for cloning ORFs of DRC-2016 alphasatellite and BBTV

RR50 attB1 a-Rep s	GGGGACAAGTTTGTACAAAAAAGCAGGCT**ATG**ACGTGTCATCAGCGAAATTG
RR51 attB2 a-Rep as	GGGGACCACTTTGTACAAGAAAGCTGGGT**TTA**GCAATTTATTATTATTAATCTGTCTCCGGAGATTTTTTC
RR52 attB1 Clink s	GGGGACAAGTTTGTACAAAAAAGCAGGCT**ATG**GAGTTCTGGGAATCGTCTG
RR53 attB2 Clink as	GGGGACCACTTTGTACAAGAAAGCTGGGT**TTA**GAGTAATGTTACATCATAGTCTGATATAACTATCTTCTGACAAT
RR54 attB1 MP s	GGGGACAAGTTTGTACAAAAAAGCAGGCT**ATG**GCATTAACAACAGAGCGGG
RR55 attB2 MP as	GGGGACCACTTTGTACAAGAAAGCTGGGT**TTA**AAACATAGGTCCGACGTTTCCTCGTC
RR56 attB1 NSP s	GGGGACAAGTTTGTACAAAAAAGCAGGCT**ATG**GATTGGGCGGAATCACAATTCAAGAC
RR57 attB2 NSP as	GGGGACCACTTTGTACAAGAAAGCTGGGT**TTA**TTCCTTGATTCTTAAAGAACAAACAACATAACACTTCAGA
RR58 attB1 Rep s	GGGGACAAGTTTGTACAAAAAAGCAGGCT**ATG**GCGCGATATGTGGTATGC
RR59 attB2 Rep as	GGGGACCACTTTGTACAAGAAAGCTGGGT**TCA**GCAAGAAACCAACTTTATTCGATCTTC
RR60 attB1 CP s	GGGGACAAGTTTGTACAAAAAAGCAGGCT**ATG**TTCAGACAACAAATGGCTAGGTAT
RR61 attB2 CP as	GGGGACCACTTTGTACAAGAAAGCTGGGT**TCA**AACATGATATGTAATTCTGTTCTGGTAGTTTATACTTACTCC
RR62 attB1 U3 s	GGGGACAAGTTTGTACAAAAAAGCAGGCT**ATG**ACCGAAGGTCAAGGTAACCGG
RR63 attB2 U3 as	GGGGACCACTTTGTACAAGAAAGCTGGGT**TCA**CCGCCTCTCTAGATGCACATC

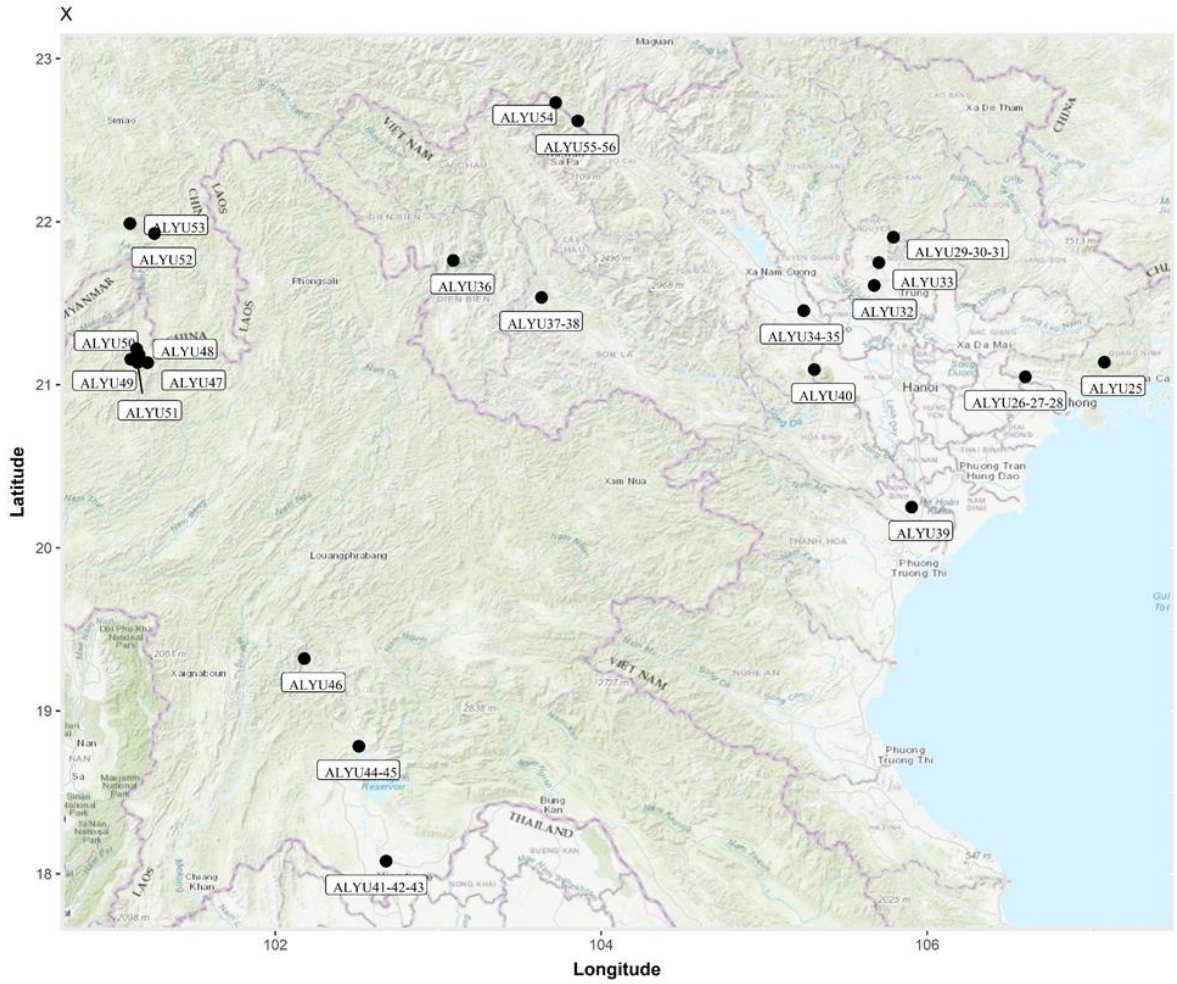
Supplementary data for Chapter 2, article 2:

Accession number	Alphasatellite original name	Species name	Genus	Sub-family	isolate name	Genome size (nt)	Sampling place (country, province, village)	sampling date (year, month)	Host plant	Reference (doi)
U02312	Banana bunchy top virus component 2	BBTA3	*Muscarsatellite*	*Petromoalphasatellitinae*	/	1095	Taiwan	1994	*Musa acuminata* Colla	(Yeh et al. 1994) - doi: 10.1006/viro.1994.1076
L32166	Banana bunchy top virus (BBTV DNA I) V1 and C1-C3 genes	BBTA3	*Muscarsatellite*	*Petromoalphasatellitinae*	/	1106	Taiwan, Southern	1994	*Musa sp.*	(Wu et al. 1994) - doi: 10.1094/Phyto-84-952
L32167 or NC_038892	Banana bunchy top alphasatellite 1 V2 (V2), V1 (V1), C2 (C2), and C1	BBTA1	*Babusatellite*	*Petromoalphasatellitinae*	/	1096	Taiwan, Southern	1994	*Musa sp.*	(Wu et al. 1994) - doi: 10.1094/Phyto-84-952
U12586	Banana bunchy top virus DNA III ORF V1 and ORF C1 genes	BBTA3	*Muscarsatellite*	*Petromoalphasatellitinae*	/	1111	Taiwan	1994	*Musa sp.*	Wu et al., unpublished Nucleotide sequences of DNA III and DNA IV associated with banana bunchy top virus and their relation to other closely related virus
U12587	Banana bunchy top virus DNA IV ORF V1 and ORF C1 genes	BBTA3	*Muscarsatellite*	*Petromoalphasatellitinae*	/	1127	Taiwan	1994	*Musa sp.*	Wu et al., unpublished Nucleotide sequences of DNA III and DNA IV associated with banana bunchy top virus and their relation to other closely related virus
AF216221	Banana bynchy top virus satellite S1	BBTA3	*Muscarsatellite*	*Petromoalphasatellitinae*	S1	1109	Taiwan	< 2001	*Musa sp.*	(Horser et al. 2001) - doi: 10.1007/s007050170192
AF216222	Banana bunchy top virus satellite S2	BBTA1	*Babusatellite*	*Petromoalphasatellitinae*	S2	1085	Taiwan	< 2001	*Musa sp.*	(Horser et al. 2001) - doi: 10.1007/s007050170192
AF416471 or NC_038953	Banana bunchy top virus satellite S3	BBTA2	*Muscarsatellite*	*Petromoalphasatellitinae*	S3	1100	Vietnam, Yen Bai	< 2002	*Musa sp.*	(Bell et al., 2002) - doi: 10.1007/s007050200019
EU366174	Banana bunchy top virus satellite S2 clone TW3S2 01	BBTA1	*Babusatellite*	*Petromoalphasatellitinae*	S2 clone TW3S2 01	1078	Taiwan	< 2007	*Musa sp.*	Liu et al., unpublished. Submitted (26-DEC-2007) Plant Pathology and Microbiology, National Taiwan University
EU366175	Banana bunchy top virus satellite Y clone TW3Y 01	BBTA3	*Muscarsatellite*	*Petromoalphasatellitinae*	Y clone TW3Y 01	1097	Taiwan	< 2007	*Musa sp.*	Liu et al., unpublished. Submitted (26-DEC-2007) Plant Pathology and Microbiology, National Taiwan University
FJ394347	Banana bunchy top virus satellite S2	BBTA1	*Babusatellite*	*Petromoalphasatellitinae*	S2 TW3	1077	Taiwan	< 2009	*Musa sp.*	(Fu et al. 2009) - doi: 10.1094/PHYTO-99-7-0812.
FJ389724	Banana bunchy top virus satellite Y	BBTA3	*Muscarsatellite*	*Petromoalphasatellitinae*	Y TW3	1098	Taiwan	< 2009	*Musa sp.*	(Fu et al. 2009) - doi: 10.1094/PHYTO-99-7-0812.
EU430730	Banana bunchy top virus satellite S4	BBTA2	*Muscarsatellite*	*Petromoalphasatellitinae*	S4	1103	Vietnam	< 2008	*Musa sp.*	Tsao et al., unpublished https://core.ac.uk/download/pdf/10885632.pdf
HQ616080 or NC_038955	Banana bunchy top alphasatellite 3 isolate Haikou 2	BBTA3	*Muscarsatellite*	*Petromoalphasatellitinae*	Haikou 2	1093	China, Hainan, Haikou	< 2012	*Musa sp.*	(Yu et al., 2012) - doi: 10.1007/s11262-012-0718-9
MG545616	DNA_satellite_4	BBTA2	*Muscarsatellite*	*Petromoalphasatellitinae*	4	1102	China, Hainan, Haikou	2013, October	*Musa acuminata* Cavendish	(Yu et al., 2019) - doi: 10.1186/s12864-019-5901-0
MG545617	DNA_satellite_NewS2	BBTA1	*Babusatellite*	*Petromoalphasatellitinae*	NewS2	1091	China, Hainan, Haikou	2013, October	*Musa acuminata* Cavendish	(Yu et al., 2019) - doi: 10.1186/s12864-019-5901-0
OK546212	DRC alphasatellite	BBTA4	*Banaphisatellite*	*Nanoalphasatellitinae*	DRC-2012	1101	Congo-DRC, Congo-Central, Boko	2012	*Musa* AAB	(Guyot et al., 2022) - PLoS Pathogens, in press
OK546211	DRC alphasatellite	BBTA4	*Banaphisatellite*	*Nanoalphasatellitinae*	DRC-2016	1105	Congo-DRC, Congo-Central, Kimpoko	2016, December	*Musa* AAB	(Guyot et al., 2022) - PLoS Pathogens, in press

S1 Table. BBTV alphasatellites previously identified in SEA and PIO.

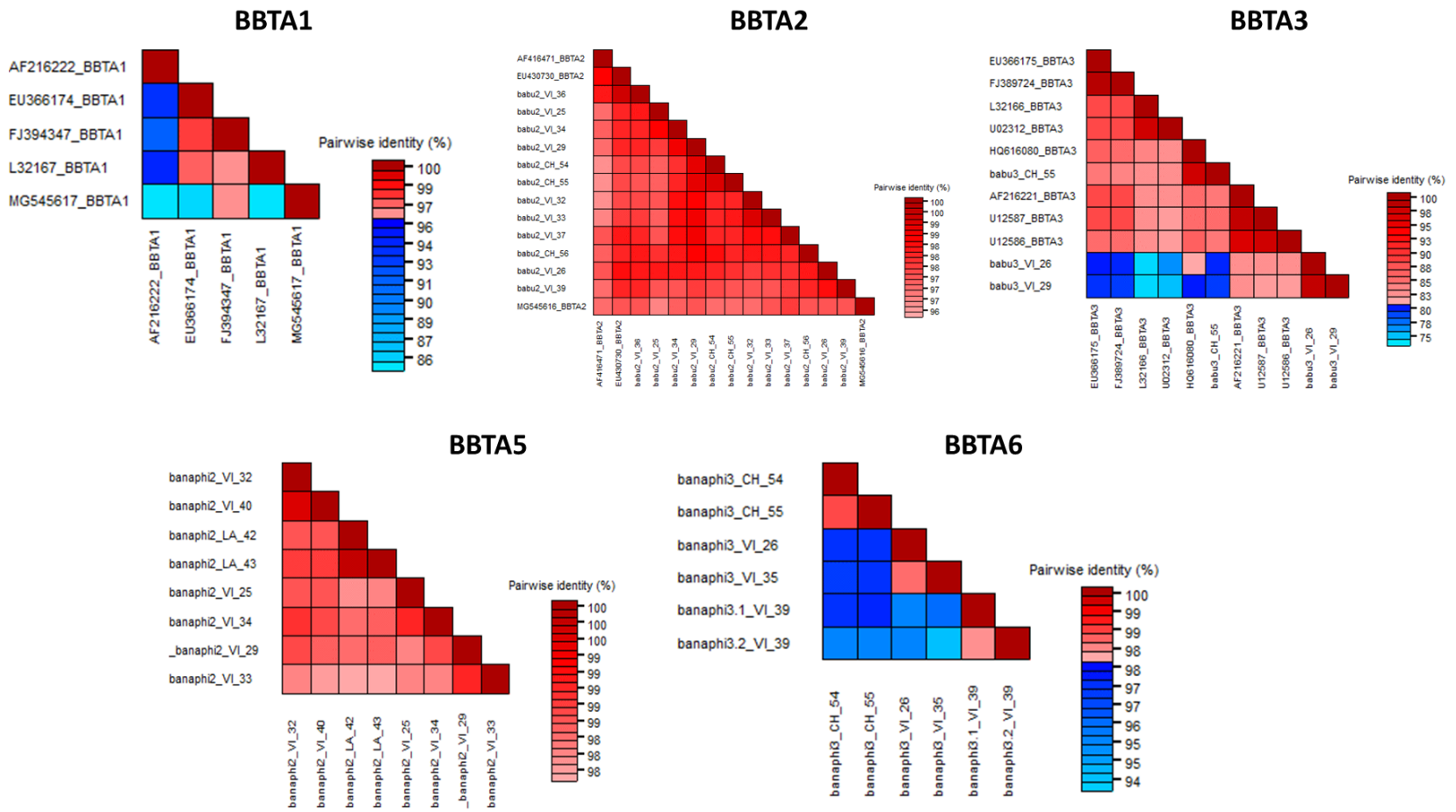
Sample ID	Country	Plant species/Genome	BBTV genome	No. of alpha-satellites	BBTA2	BBTA3	BBTA4	BBTA5	BBTA6	defective (d) molecules	others DNA viruses
ALYU-14	New Caledonia	*Musa sp.*	full								
ALYU-15	Malawi	*Musa sp.*	full								
ALYU-16	Benin	*Musa sp.*	full								
ALYU-17	Congo-DRC	*Musa* AAB Faux Corne	full								
ALYU-18	Congo-DRC	*Musa* AAA var. local	full								
ALYU-19	Congo-DRC	*Musa* AAA var. local	full								
ALYU-20	Congo-DRC	*Musa* AAB	full	1			1				
JGF-1	Congo-DRC	*Musa* AAB	full	1			1				
ALYU-22	Gabon	*Musa sp.*	full								
ALYU-23	Gabon	*Musa sp.*	no U3								
ALYU-24	Gabon	*Musa sp.*	full								
ALYU-25	Vietnam	*Musa itinerans*	full	2	1			1			Badnavirus new
ALYU-26	Vietnam	*Musa sp.*	full	3	1	1			1		
ALYU-27	Vietnam	*Commelina sp.*	full	3	1	1			1		
ALYU-28	Vietnam	*Ipomoea aquatica*	no								
ALYU-29	Vietnam	*Musa* AA sucrier Pisang mas ?	full	3	1	1		1		d-BBTA5, d-BBT2, d-DNA-R	
ALYU-30	Vietnam	*Phyllanthus sp.*	no								
ALYU-31	Vietnam	*Arachis sp.*	no								
ALYU-32	Vietnam	*Musa* sp. Sweet banana	full	2	1			1		d-BBT2	
ALYU-33	Vietnam	*Musa* AAB Chuoi Ngop	full	2	1			1			Badnavirus BSVNV
ALYU-34	Vietnam	*Musa sp.*	full	2	1			1			
ALYU-35	Vietnam	*Musa sp.*	full	1					1		
ALYU-36	Vietnam	*Musa sp.*	full	1	1						
ALYU-37	Vietnam	*Musa sp.*	full	1	1						
ALYU-38	Vietnam	*Bidens pilosa*	full	4	1	1		1	1		Cyclovirus, Microvirus
ALYU-39	Vietnam	*Musa sp.*	full	2	1				1		
ALYU-40	Vietnam	*Musa* AAA Red banana	full	1				1			
ALYU-41	Laos	*Chromolaena odorata*	no N	1				1			
ALYU-42	Laos	*Musa acuminata* Cavendish AAA	full	1				1			
ALYU-43	Laos	*Musa acuminata* Cavendish AAA	full	1				1			
ALYU-44	Laos	*Musa ornata*	full								
ALYU-45	Laos	*Musa sp.*	full								
ALYU-46	Laos	*Musa* ABB Klue Tiparot	no								
ALYU-47	Laos	*Musa yunnenensis*	full								
ALYU-48	Laos	*Musa sp.*	full								
ALYU-49	Laos	*Musa sp.*	full								
ALYU-50	Laos	*Musa* AA sucrier Kouay niew mung	full								
ALYU-51	Laos	*Musa* ABB Pisang Awak ?	full								
ALYU-52	China	*Musa acuminata* wild	no								
ALYU-53	China	*Musa yunnanensis*	full								
ALYU-54	China	*Musa acuminata* Cavendish AAA	full	2	1				1	d-BBTA6	
ALYU-55	China	*Musa acuminata* Cavendish AAA	full	3	1	1			1		
ALYU-56	China	*Musa acuminata* Cavendish AAA	full	1	1						

S2 Table. Virome components identified in wild and cultivated banana plants sampled in SEA (Vietnam, Laos and China) during surveys in 2018 and 2019 and in PIO in 2012-2020.



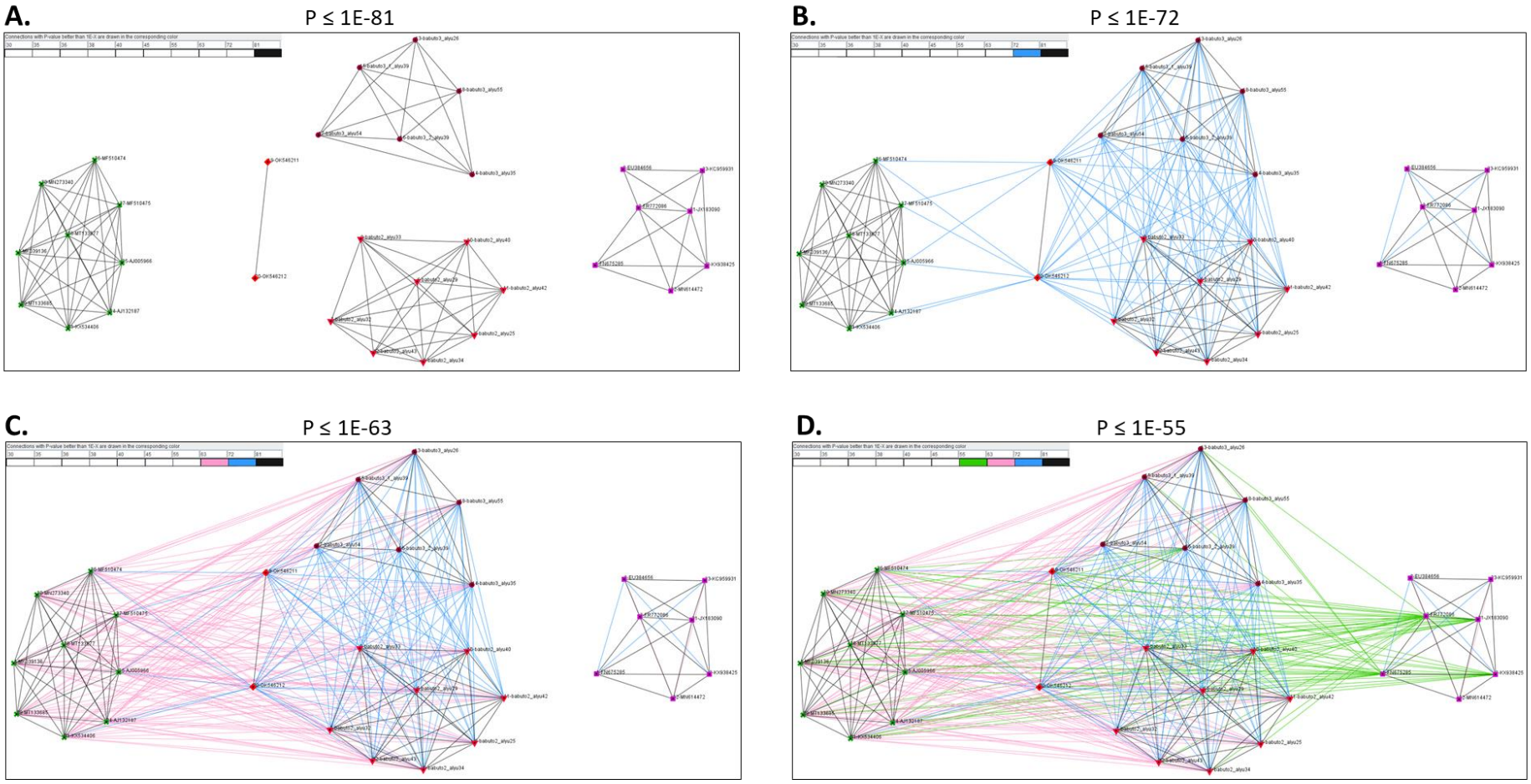
S1 Fig. Locations of banana leaf sampling places in Vietnam, Laos and China.

The map was created using R studio and package OpenStreepMap.

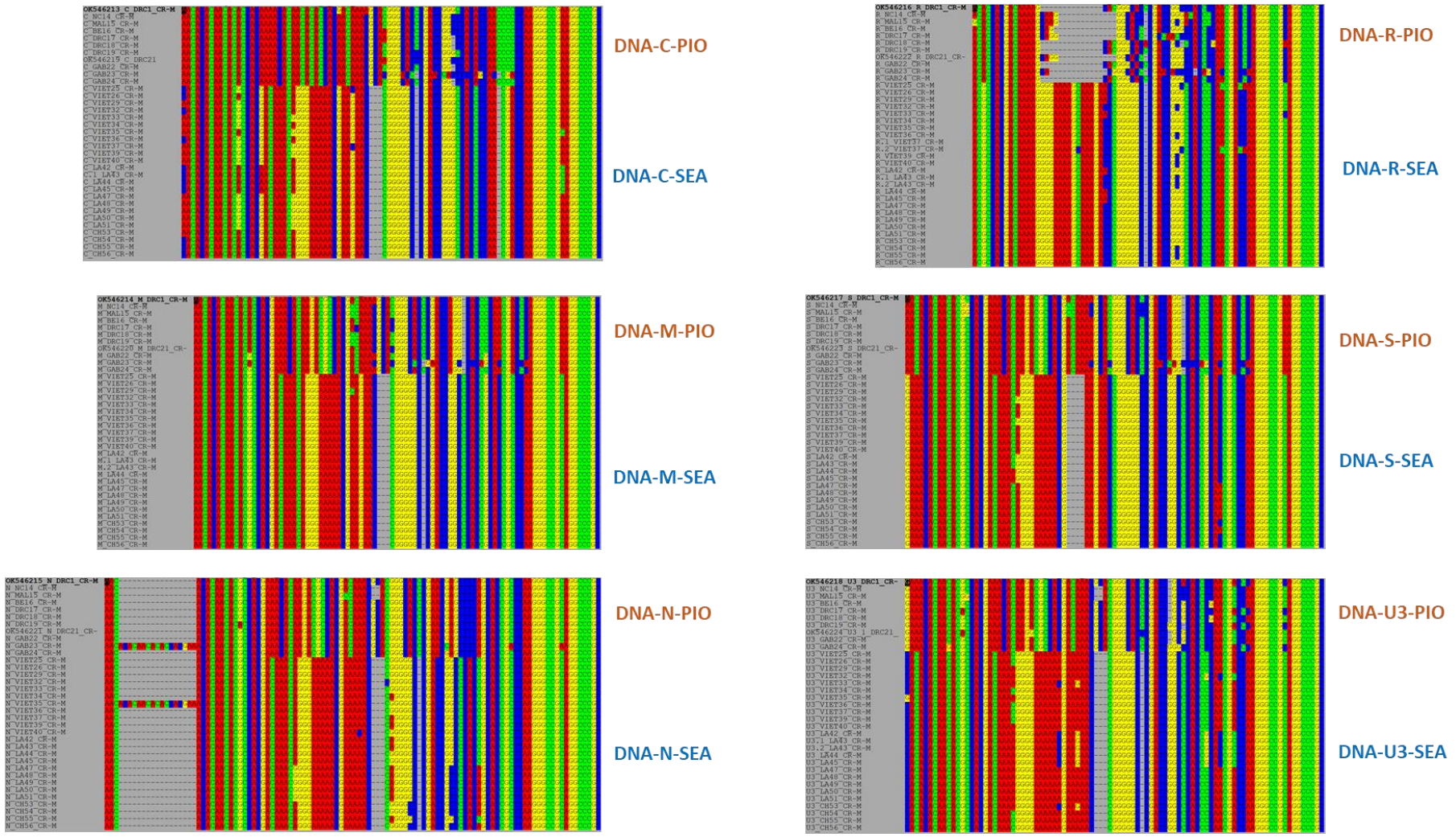


S2 Fig. Pairwise sequence comparison of BBTV alphasatellites using Sequence Demarcation Tool (SDT) v1.2

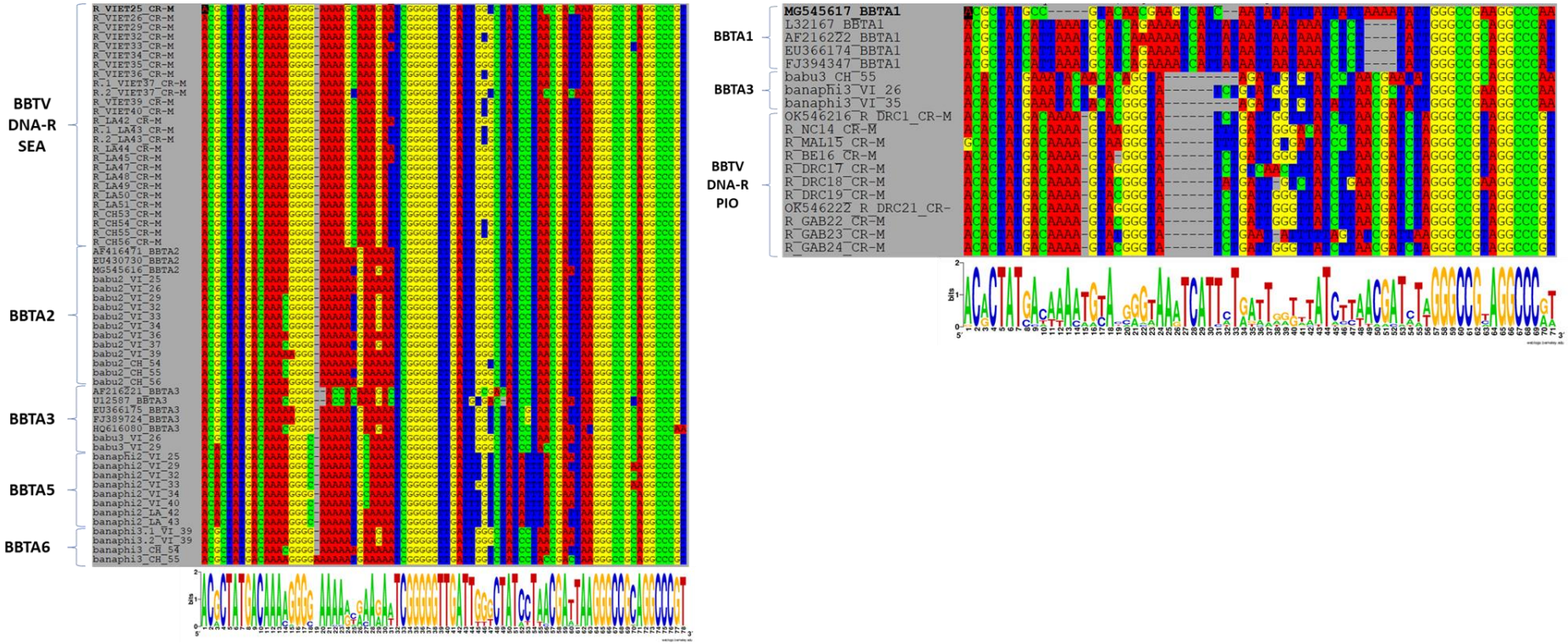
The nucleotide sequences of all isolates of BBTV alphasatellites available at the NCBI Genbank in September 2021 and those *de novo* reconstructed in this study were compared for each species (BBTA1-to-6) using Sequence Demarcation Tool (SDT) v1.2 with Muscles (excluding indels) (Muhire et al. 2014).



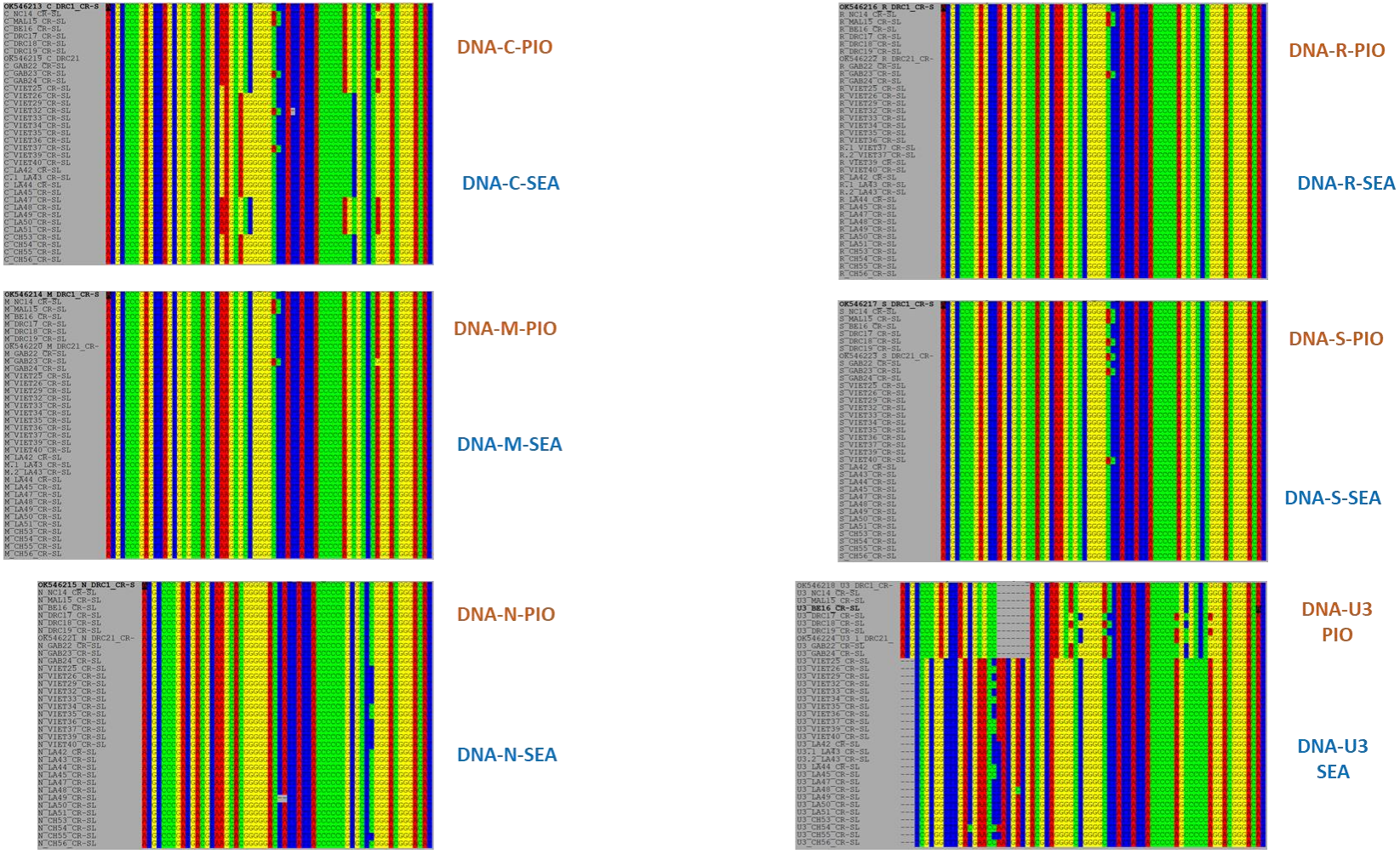
S3 Fig. Sequence similarity-based clustering analysis of Rep proteins encoded by alphasatellites from *Fabenesatellite*, *Banaphisatellite* and *Gosmusatellite* genera of the family *Alphasatellitidae*



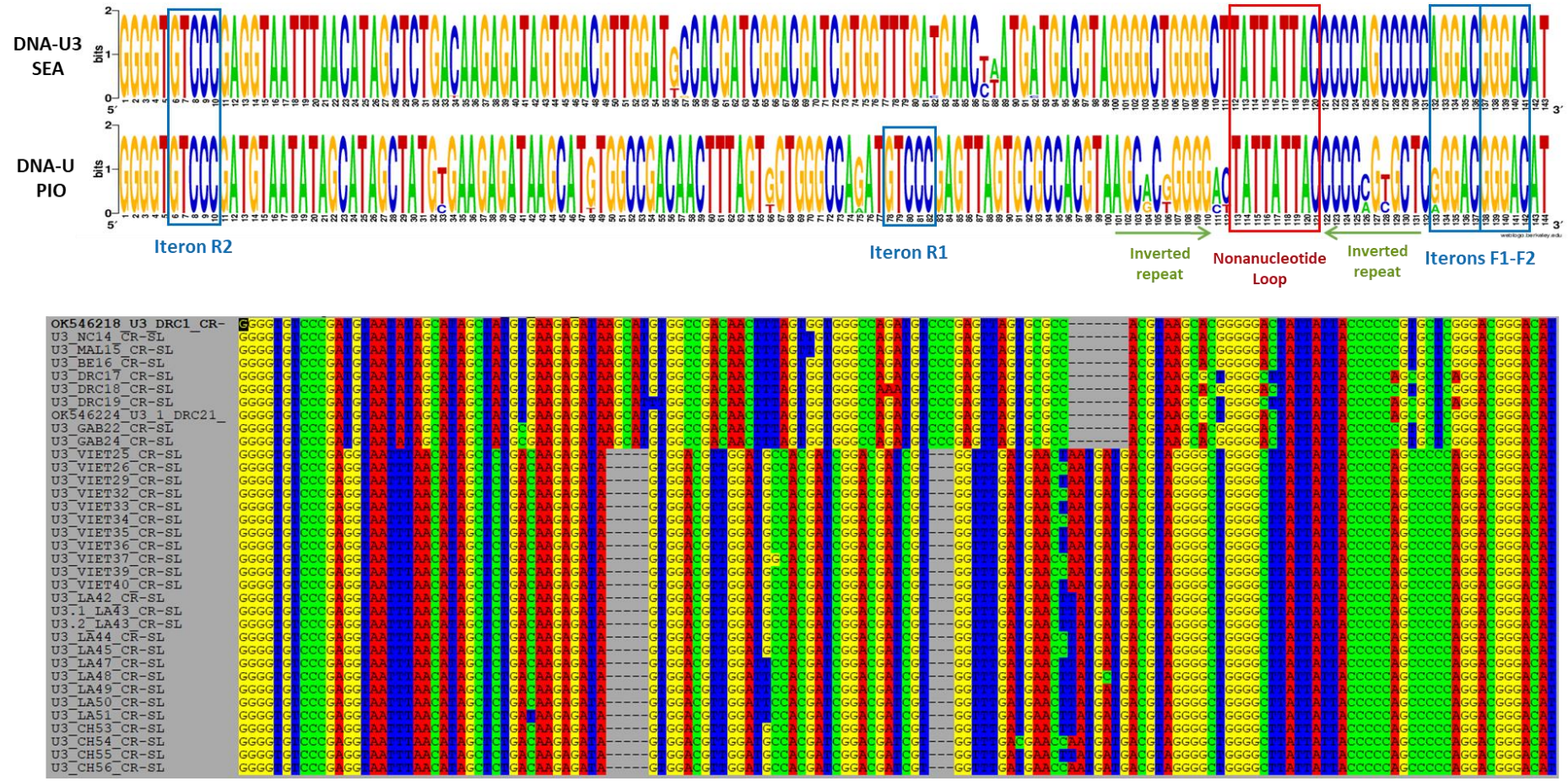
S4 Fig. Alignment of the common-region major (CR-M) sequences of the SEA and PIO isolates of BBTV



S5 Fig. Comparison of CR-M of and alphasatellites and BBTV DNA-R from SEA and PIO phylogenetic group



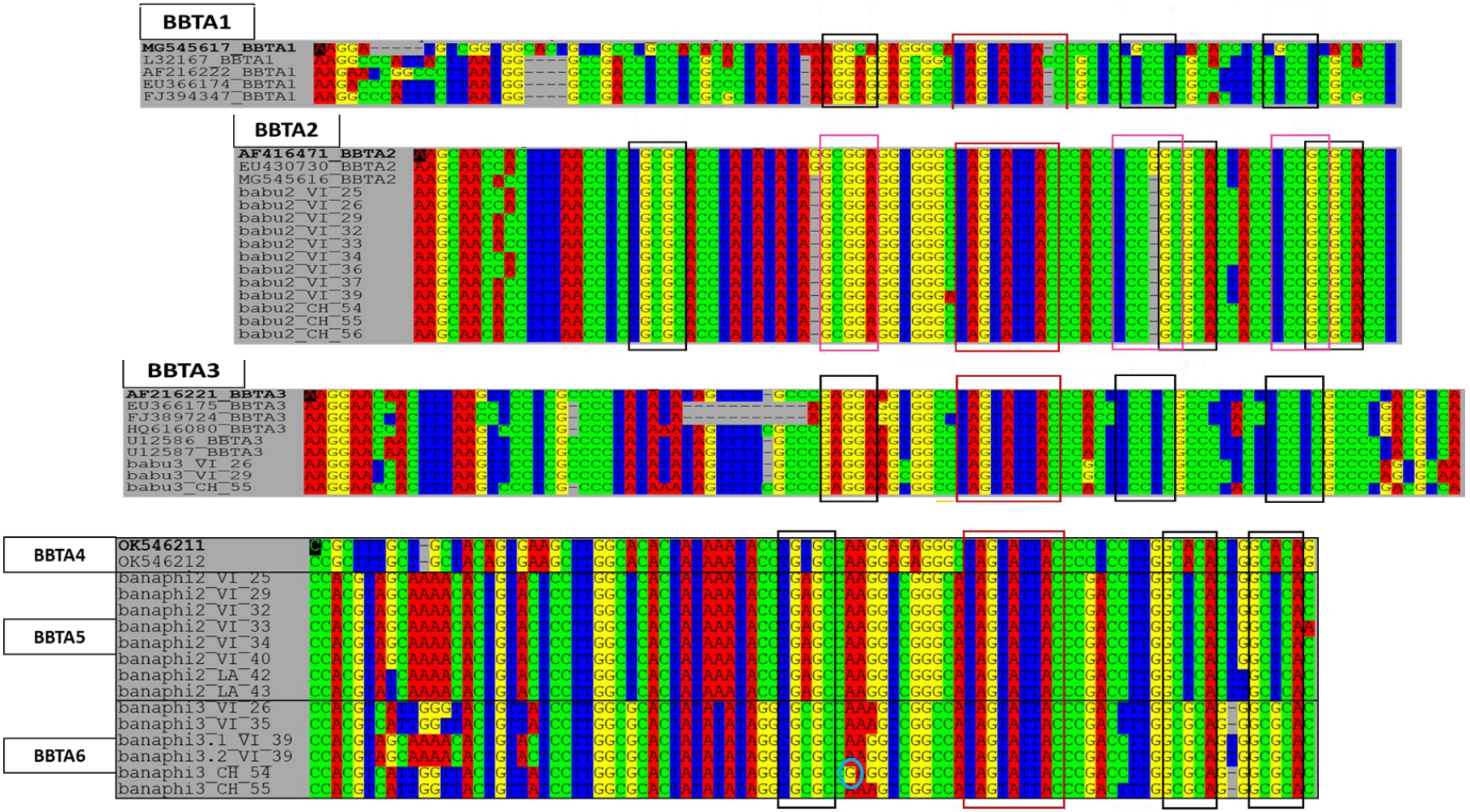
S6A Fig. Alignment of the common region stem-loop (CR-SL) sequences of the SEA and PIO isolates of BBTV



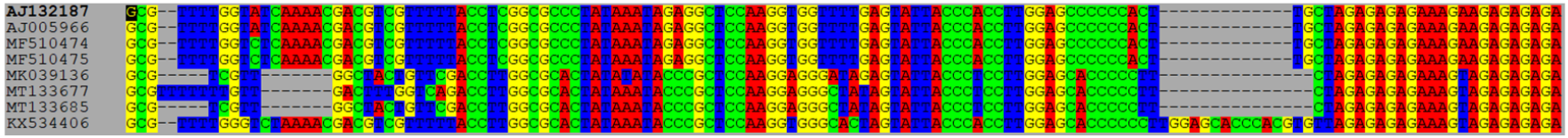
S6B Fig. Logo and Alignment of the common region stem-loop (CR-SL) sequences of the SEA and PIO isolates of BBTV DNA-U3

S7 Fig. Alignment of the common region stem-loop (CR-SL) sequences of BBTV alphasatellites and fabenesatellites

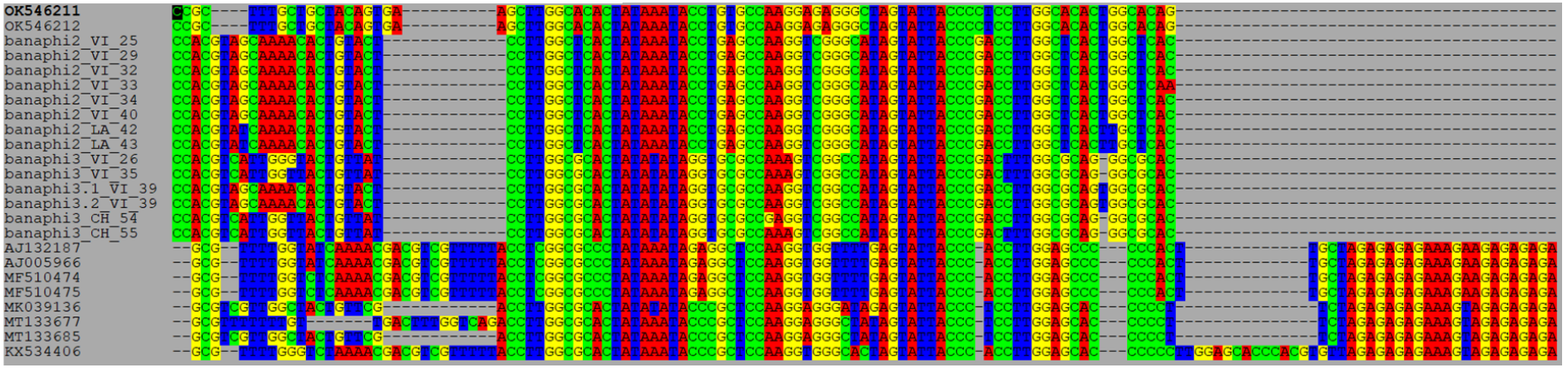
A. Babusatellites vs Muscarsatellites vs Banaphisatellites

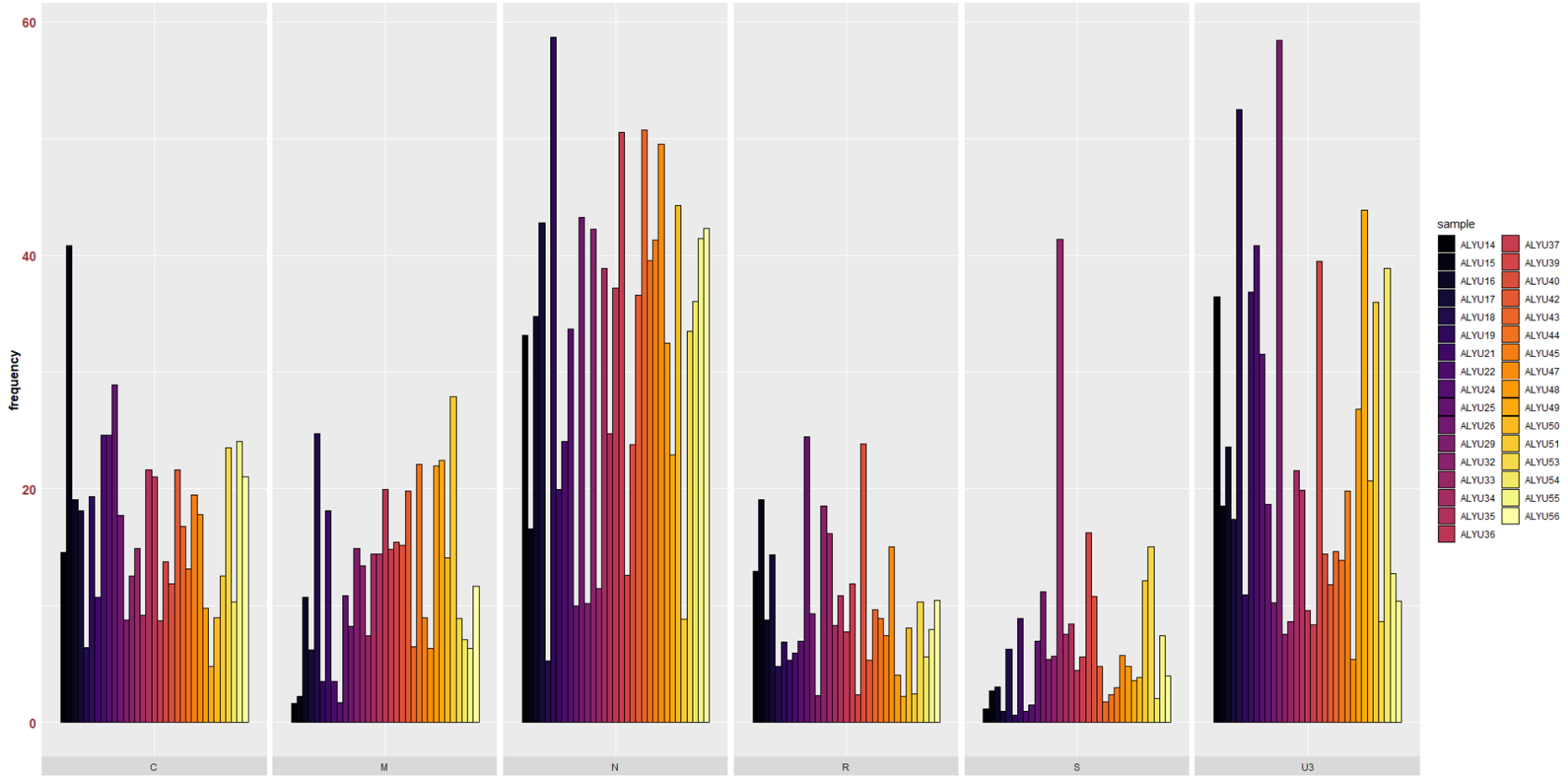


B. Fabenesatellites

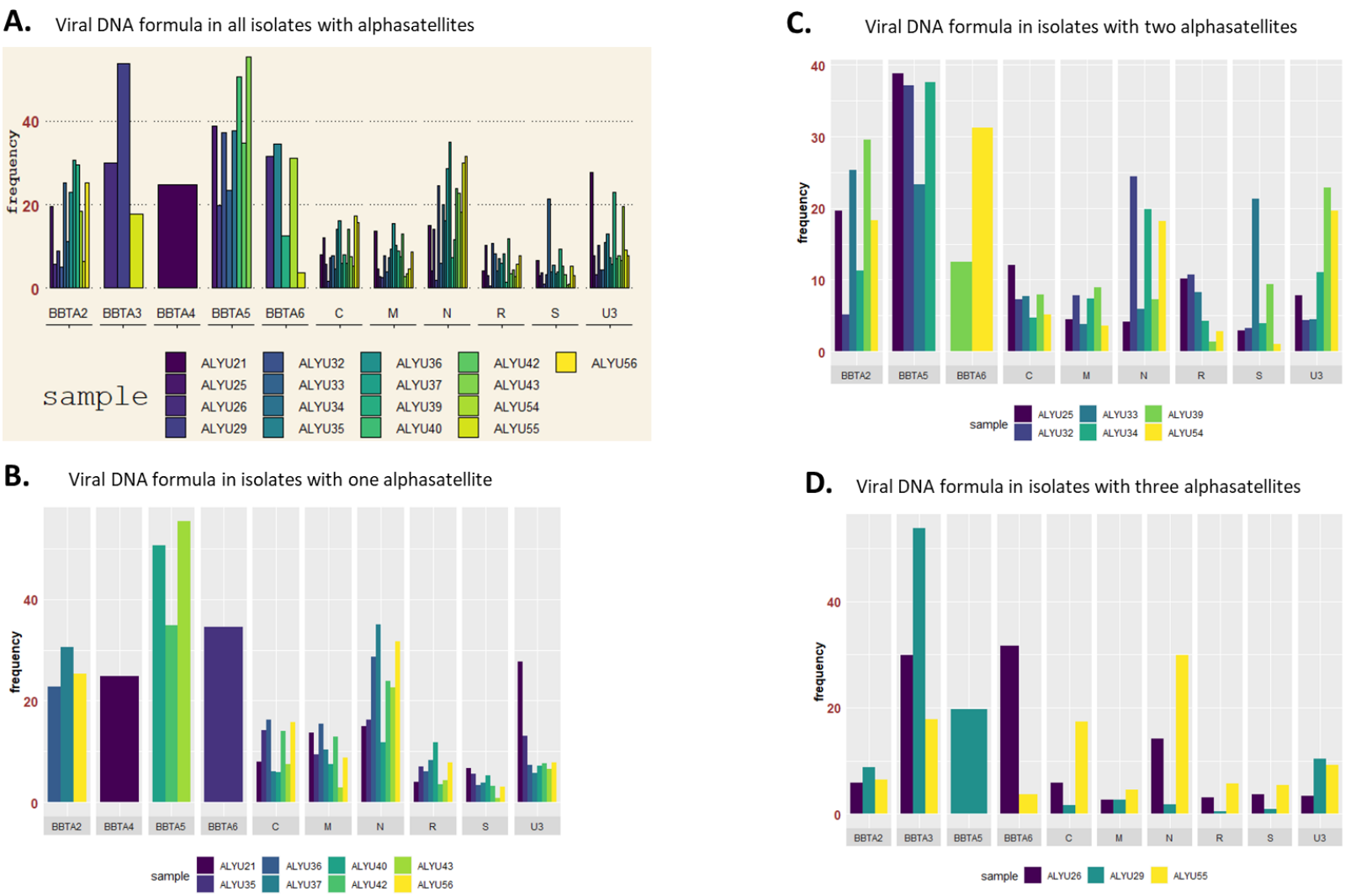


C. Banaphisatellites vs Fabenesatellites

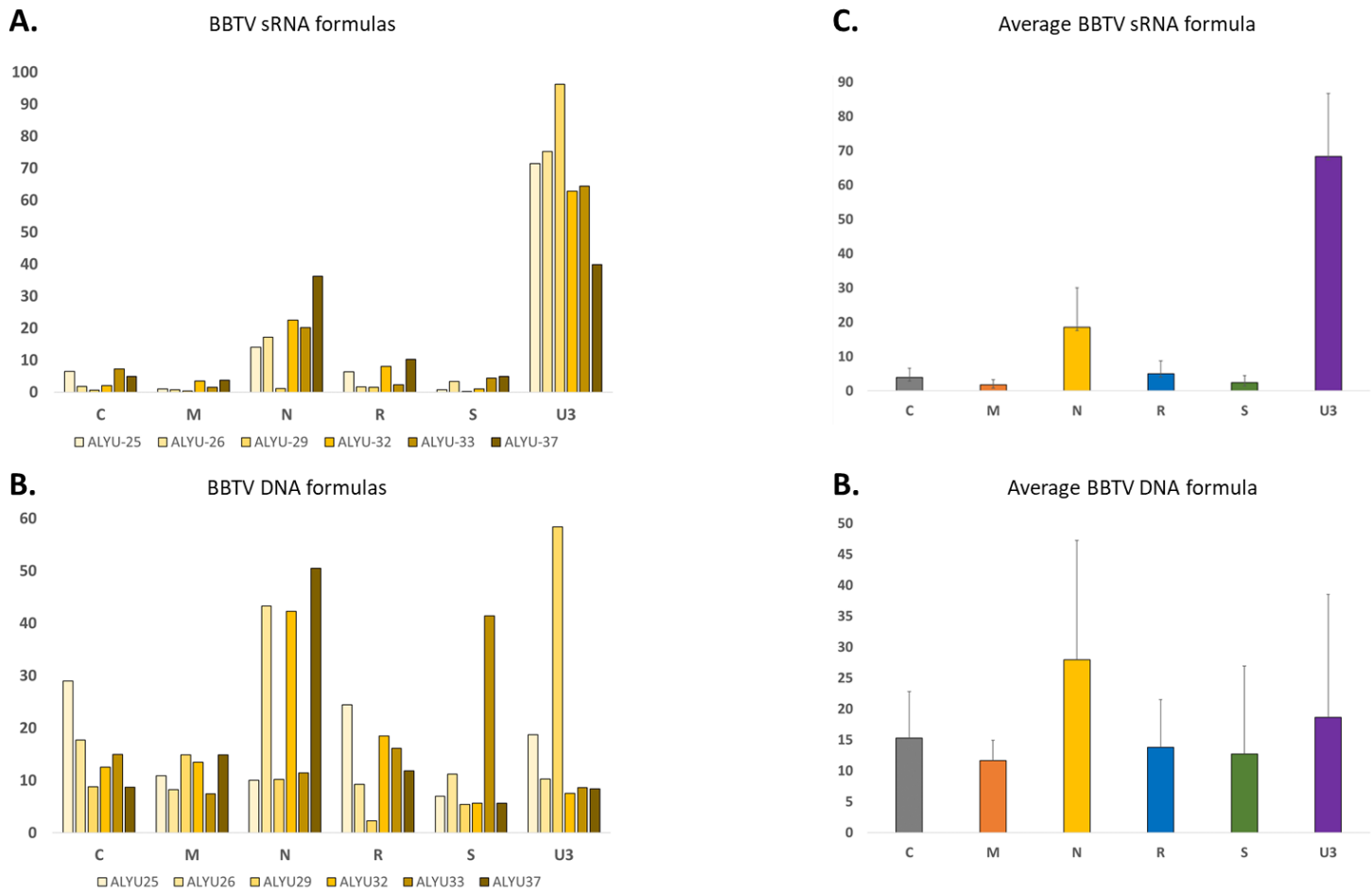




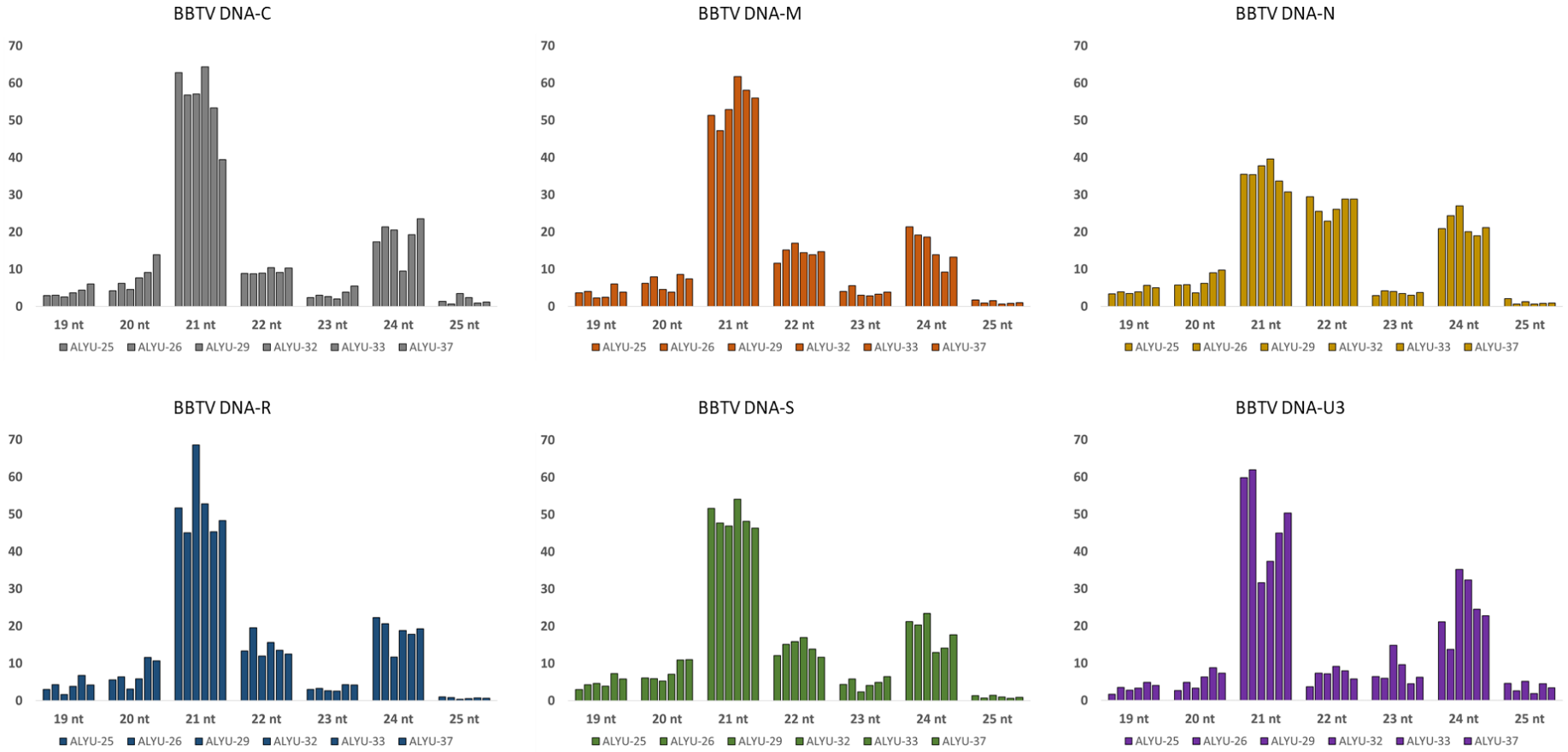
S8 Fig. Relative abundance of BBTV genome components in PIO (ALYU-14-24) and SEA (ALYU-25-56) isolates



S9 Fig. (A) Relative abundance of BBTV genome components and alphasatellite of all samples with alphasatellite(s), or with (B) one, (C) two, (D) three alphasatellite species.



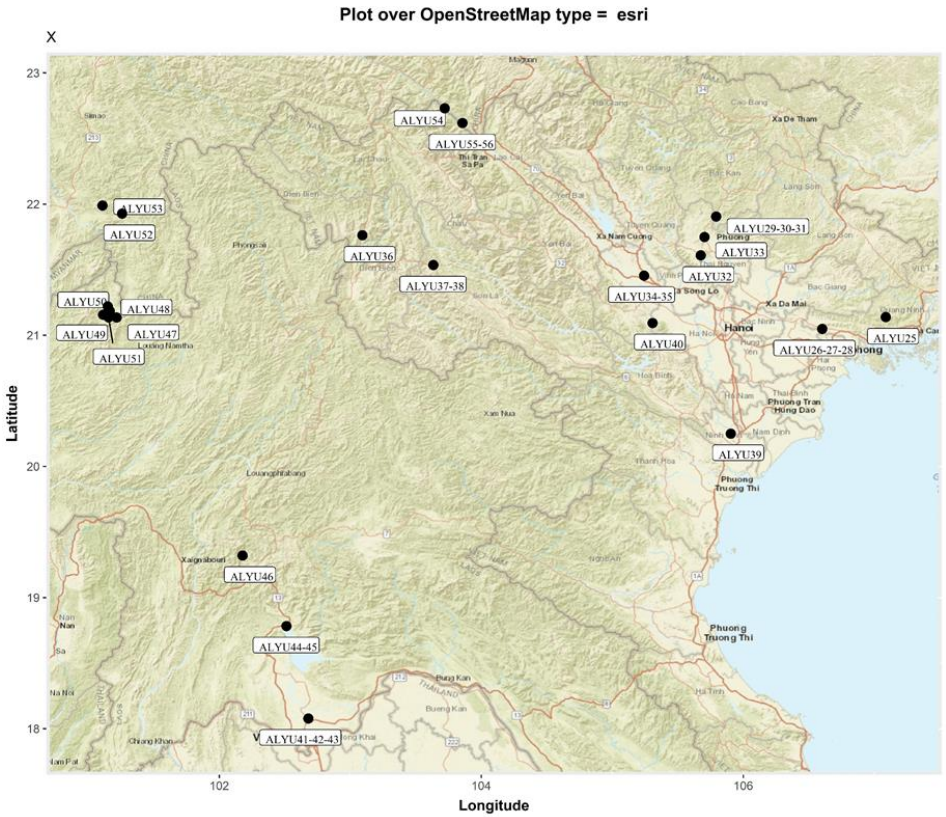
S10 Fig. Comparison of helper virus sRNA and DNA formulas



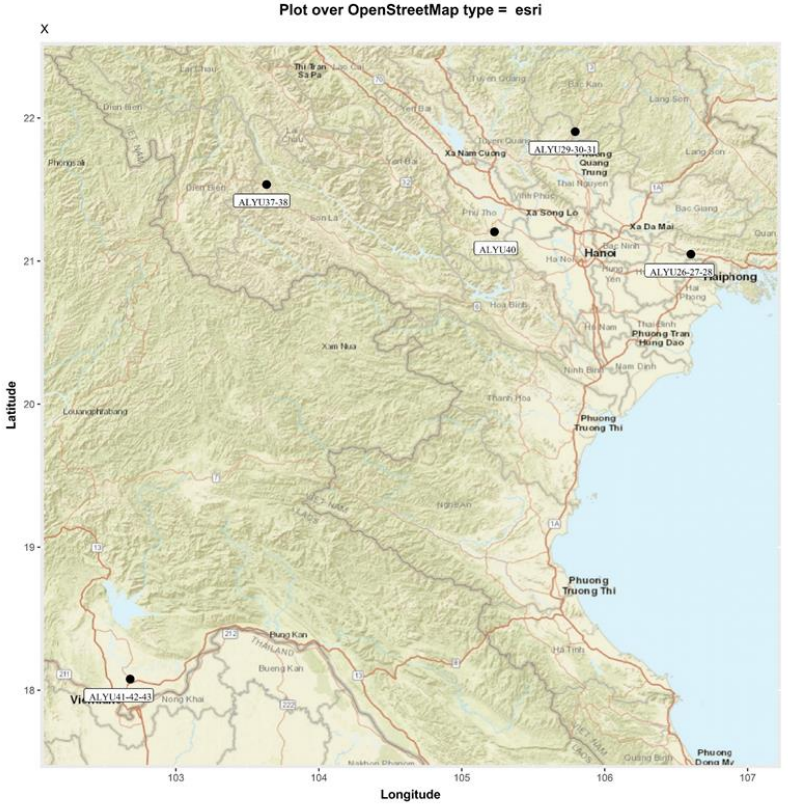
S11 Fig. Relative abundance of 19-, 20-, 21-, 22-, 23-, 24- and 25-nt classes of BBTV component sRNAs (size profile)

Supplementary data for Chapter 3, article 3 (short communication):

(A) All plant samples



(B) Samples of non-banana plants and their banana neighbours



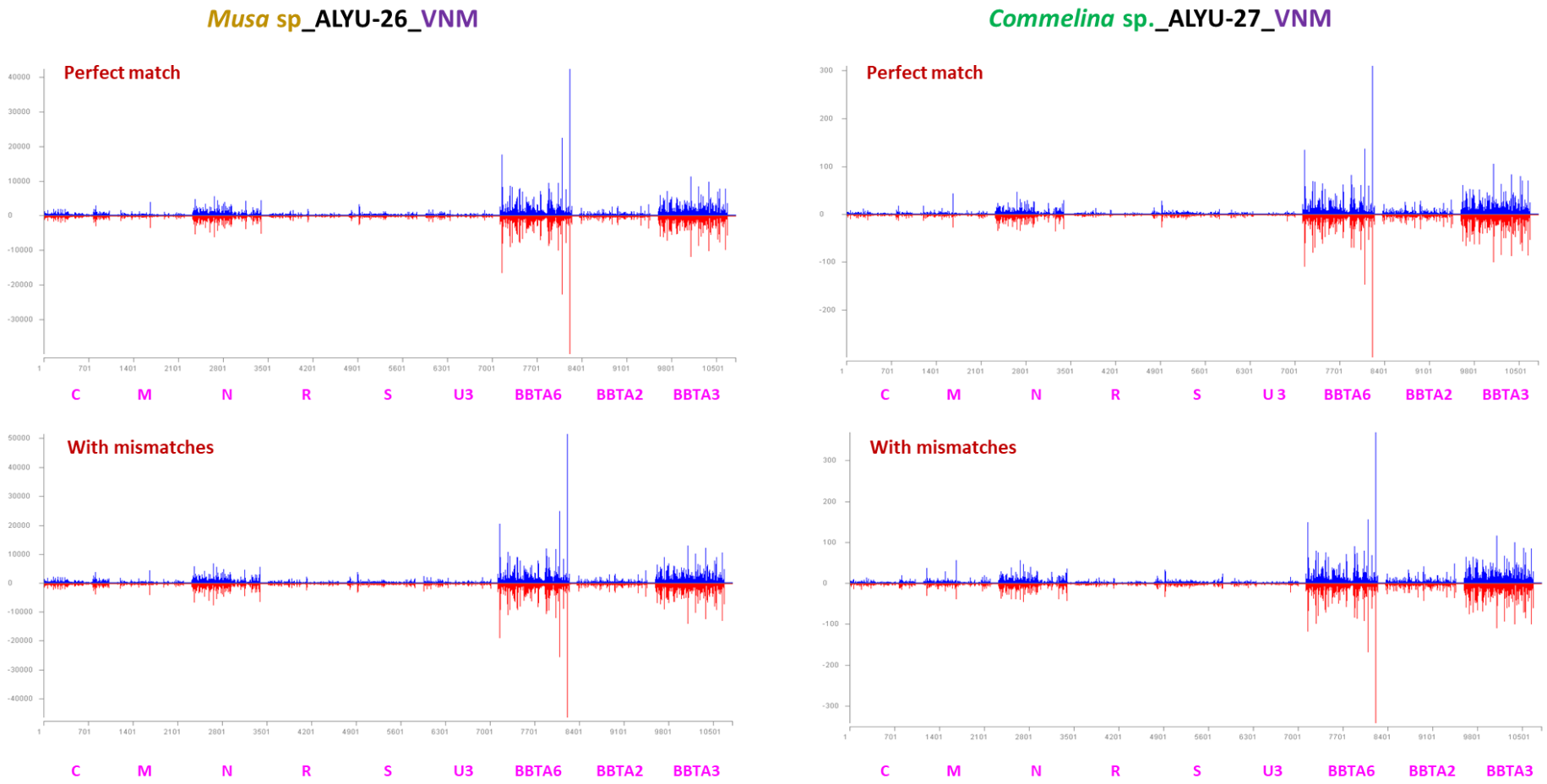
S1 Fig. Locations of banana and non-banana leaf sampling places in Vietnam, Laos and China



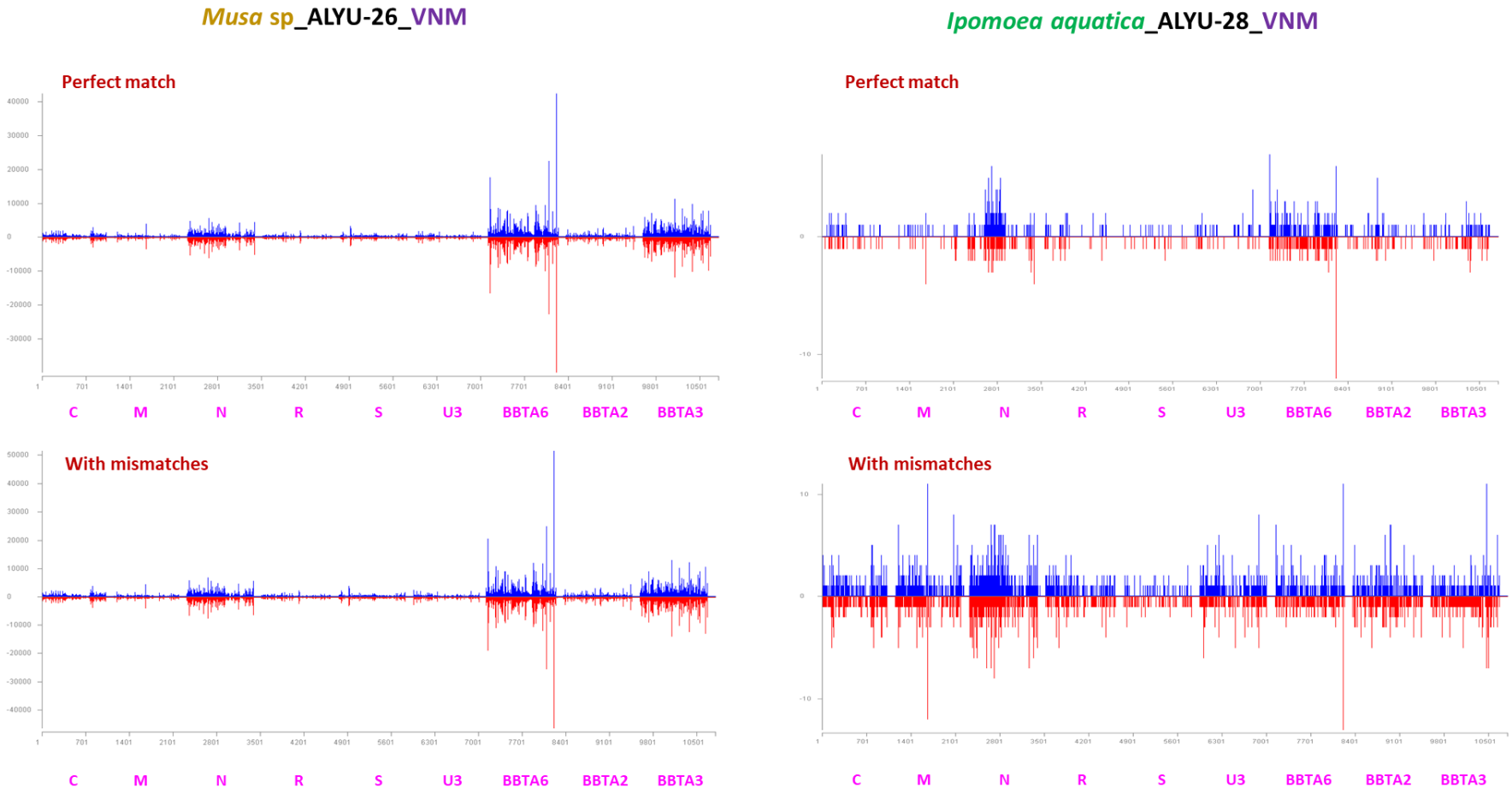
S2 Fig. *Bidens pilosa* and its neighbor banana plants

White arrows point at virus-like chlorotic symptoms *in B. pilosa* plants

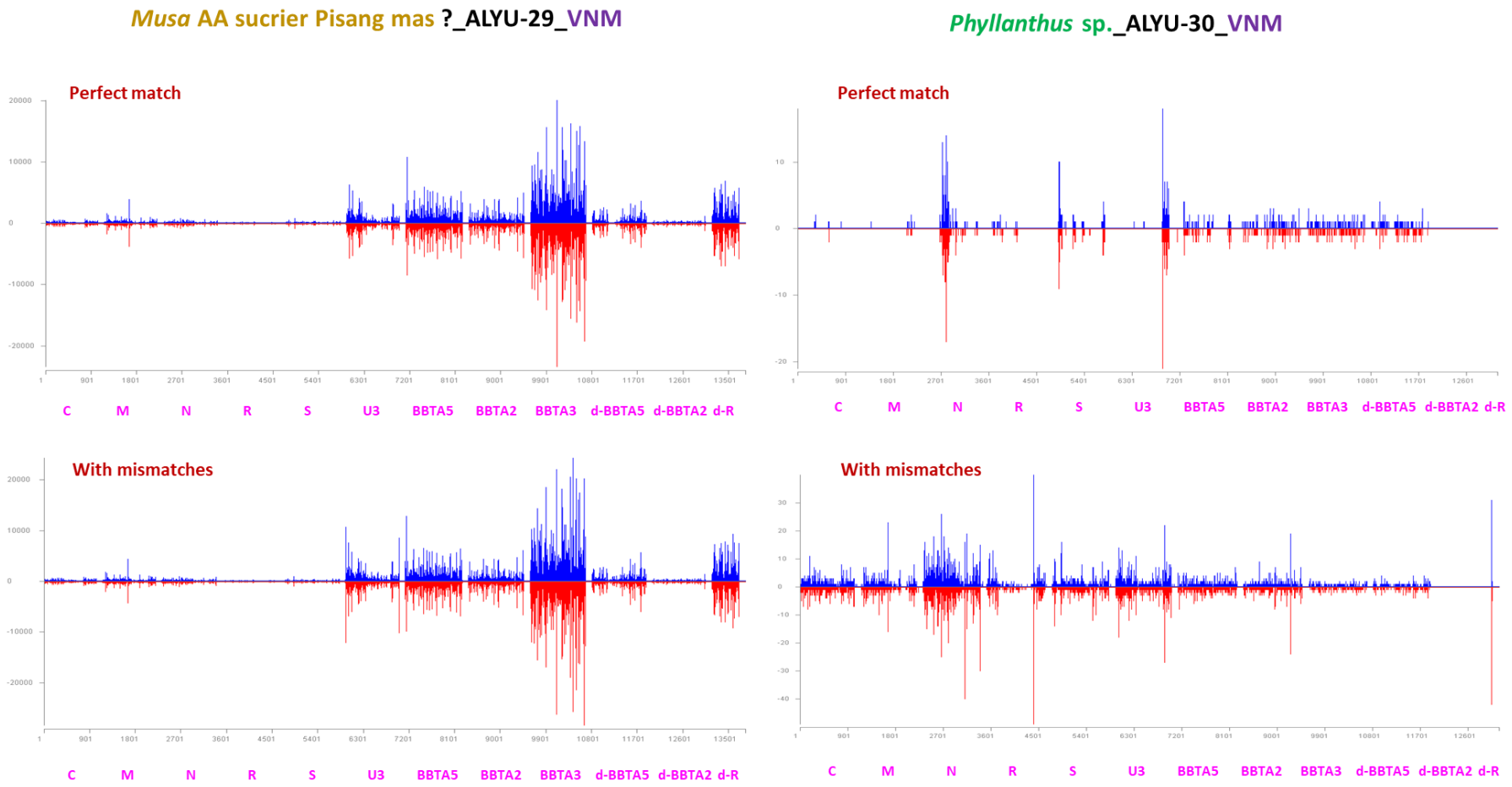
S3A Fig. Illumina read maps on concatenated reference sequences of the virome components



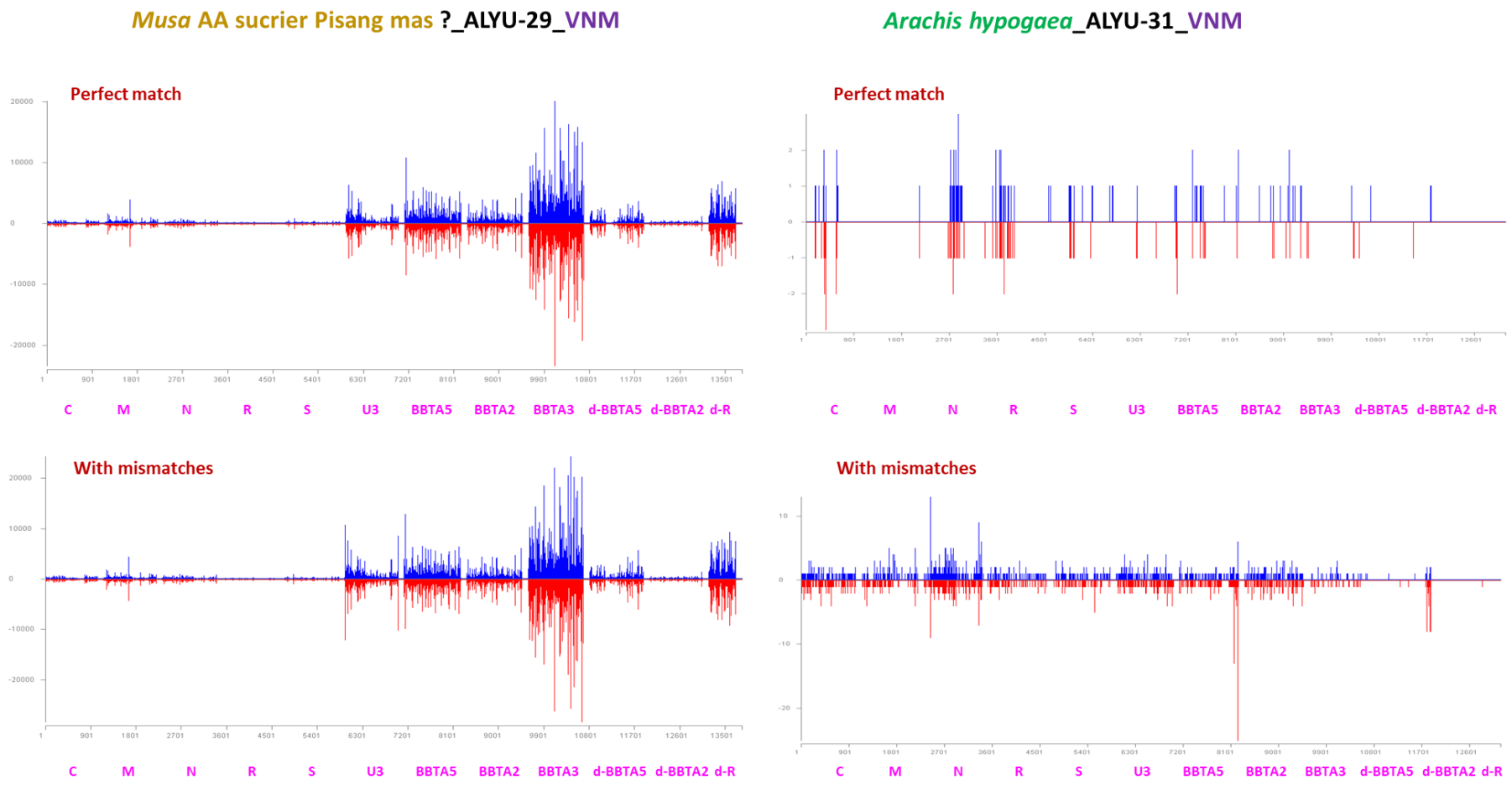
S3B Fig.


Musa sp_ALYU-26_VNM
Perfect match
With mismatches
C M N R S U3 BBTA6 BBTA2 BBTA3
*Ipomoea aquatica*_ALYU-28_VNM
Perfect match
With mismatches
C M N R S U3 BBTA6 BBTA2 BBTA3

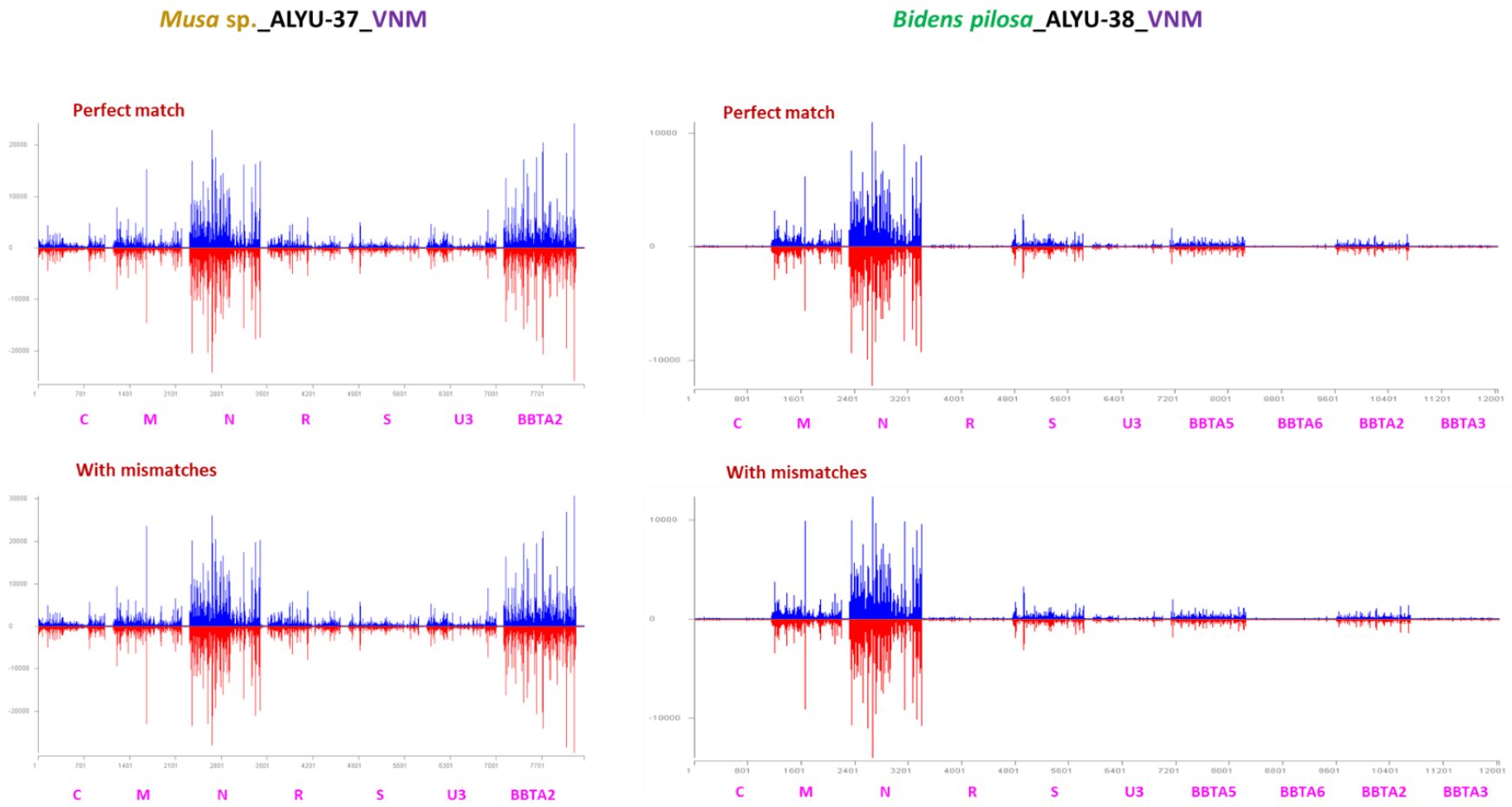
S3C Fig.


Musa AA sucrier Pisang mas ?_ALYU-29_VNM
Phyllanthus sp._ALYU-30_VNM
Perfect match
Perfect match
With mismatches
With mismatches
C M N R S U3 BBTA5 BBTA2 BBTA3 d-BBTA5 d-BBTA2 d-R
C M N R S U3 BBTA5 BBTA2 BBTA3 d-BBTA5 d-BBTA2 d-R
C M N R S U3 BBTA5 BBTA2 BBTA3 d-BBTA5 d-BBTA2 d-R
C M N R S U3 BBTA5 BBTA2 BBTA3 d-BBTA5 d-BBTA2 d-R

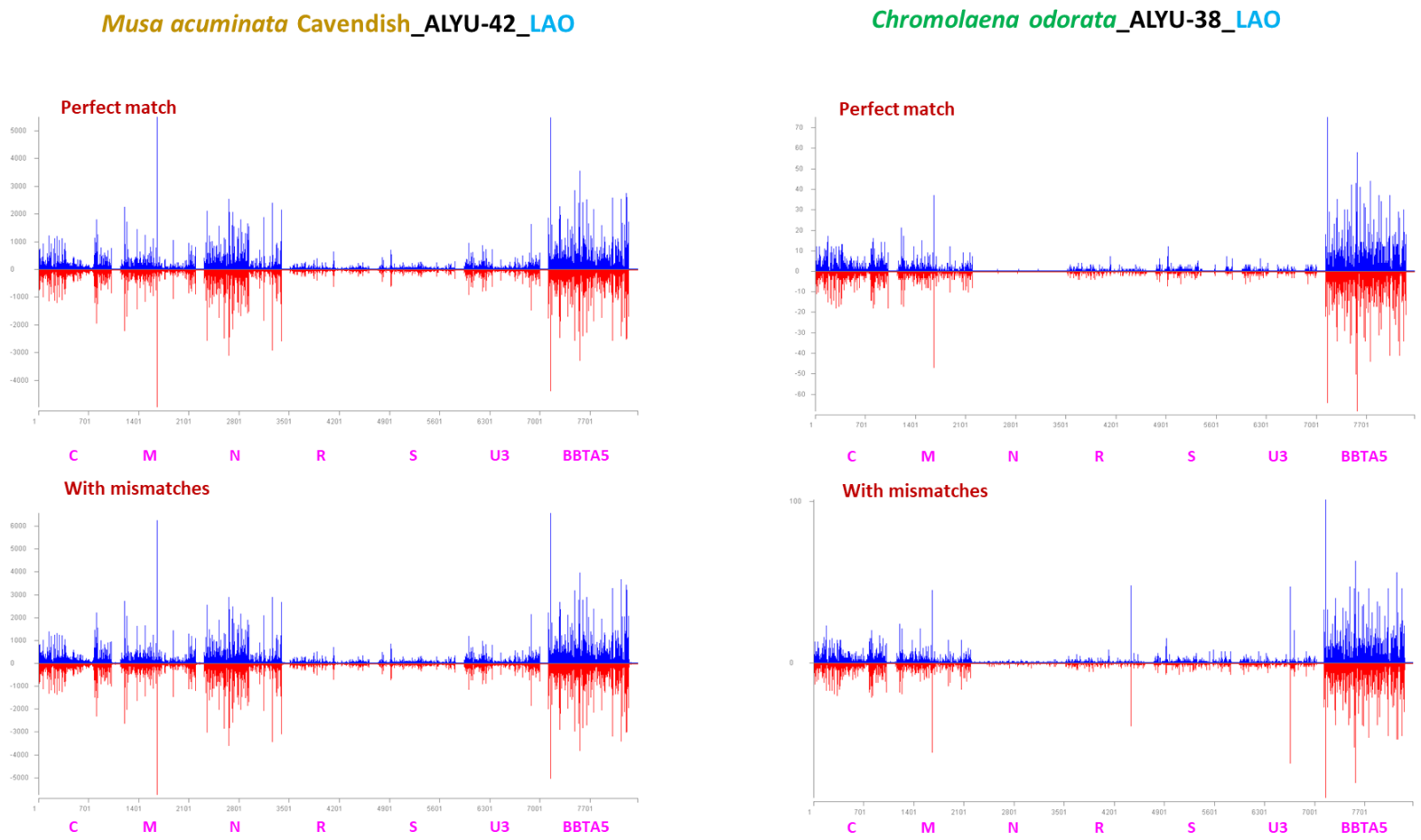
S3D Fig.


Musa AA sucrier Pisang mas ?_ALYU-29_VNM
*Arachis hypogaea*_ALYU-31_VNM
Perfect match
Perfect match
With mismatches
With mismatches
C M N R S U3 BBTA5 BBTA2 BBTA3 d-BBTA5 d-BBTA2 d-R
C M N R S U3 BBTA5 BBTA2 BBTA3 d-BBTA5 d-BBTA2 d-R
C M N R S U3 BBTA5 BBTA2 BBTA3 d-BBTA5 d-BBTA2 d-R
C M N R S U3 BBTA5 BBTA2 BBTA3 d-BBTA5 d-BBTA2 d-R

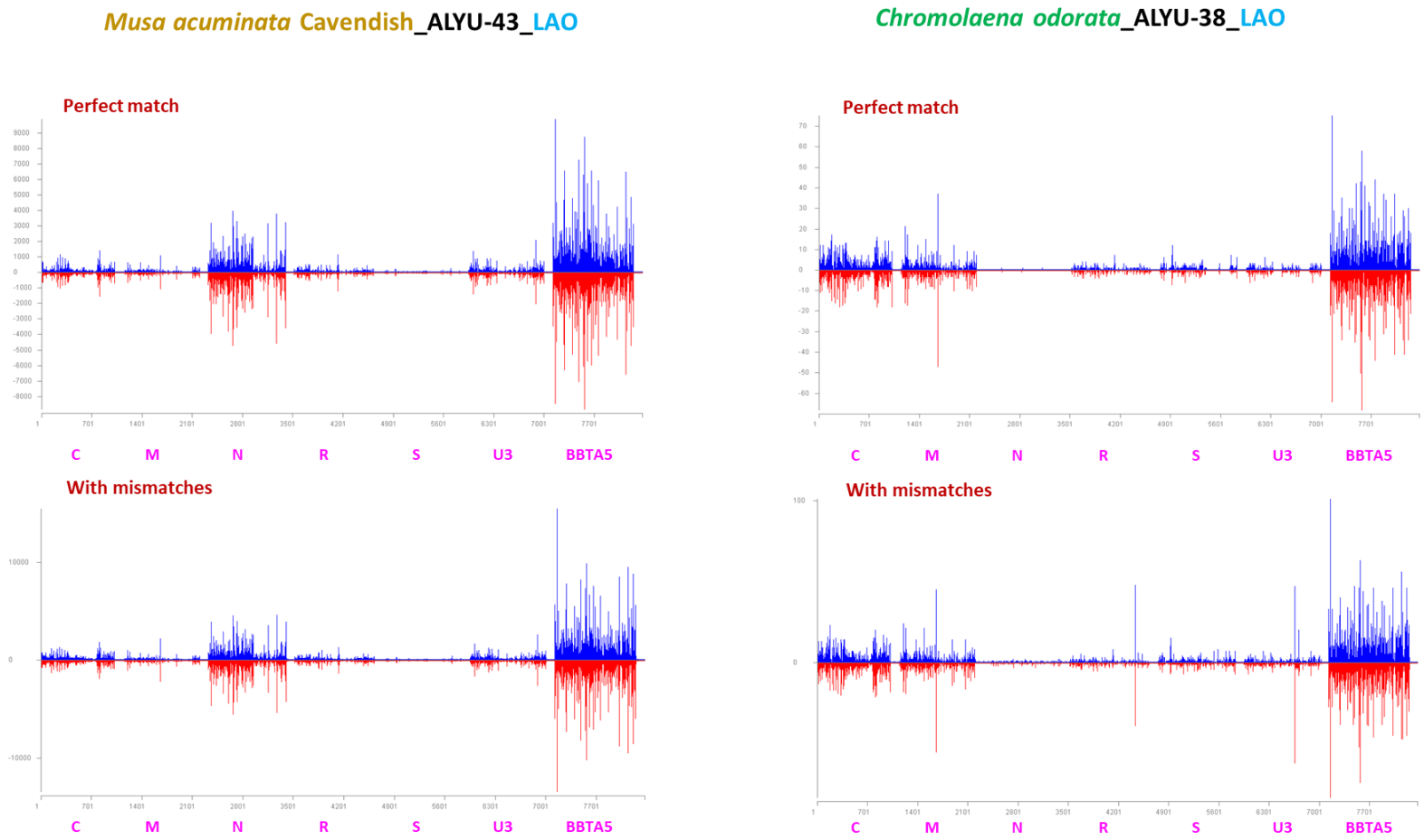
S3E Fig.


Musa sp._ALYU-37_VNM
*Bidens pilosa*_ALYU-38_VNM
Perfect match
With mismatches
C
M
N
R
S
U3
BBTA2
BBTA5
BBTA6
BBTA3

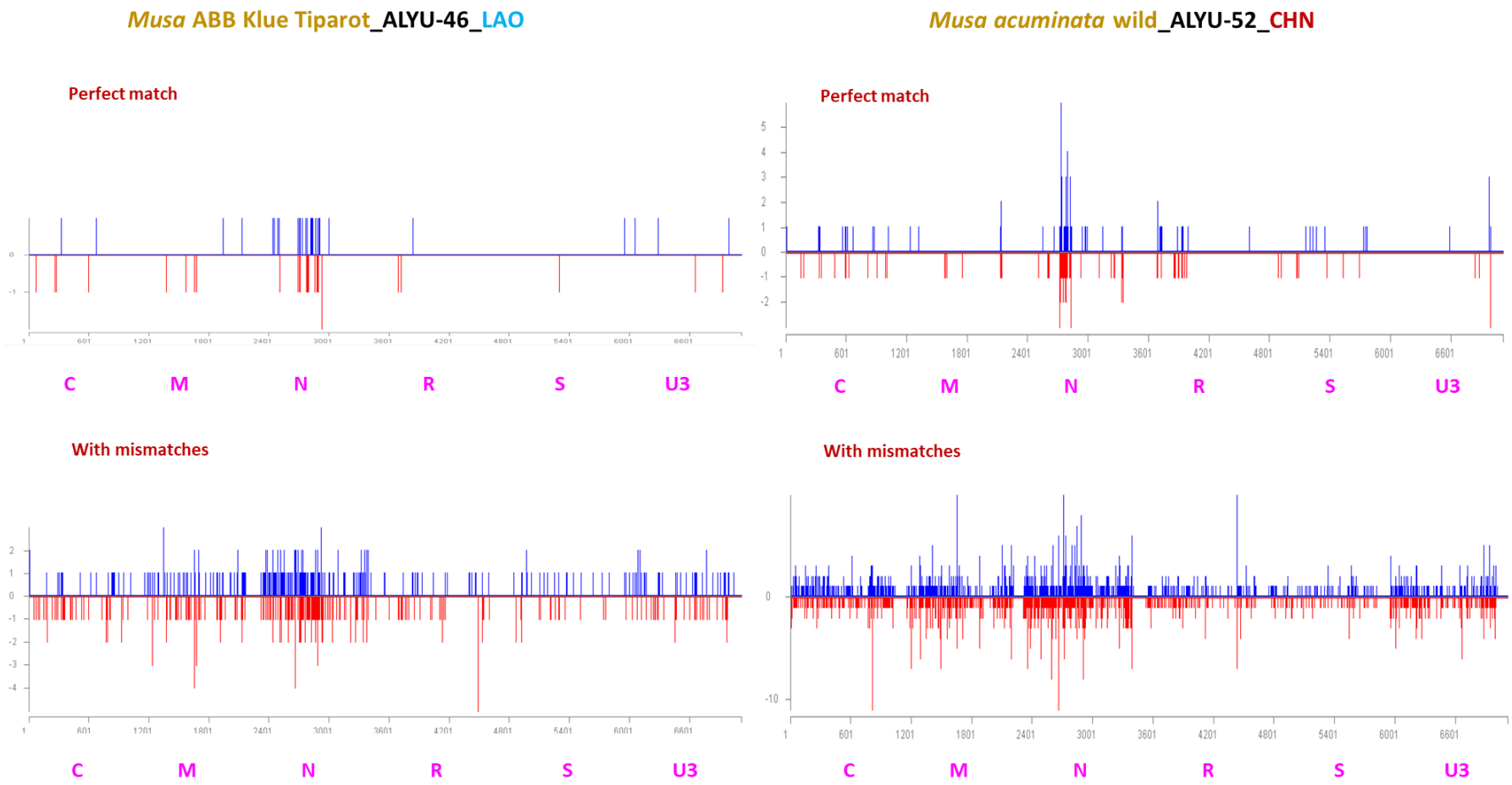
S3F Fig.


Musa acuminata Cavendish_ALYU-42_LAO
Perfect match
With mismatches
C M N R S U3 BBTA5
*Chromolaena odorata*_ALYU-38_LAO
Perfect match
With mismatches
C M N R S U3 BBTA5

S3G Fig.

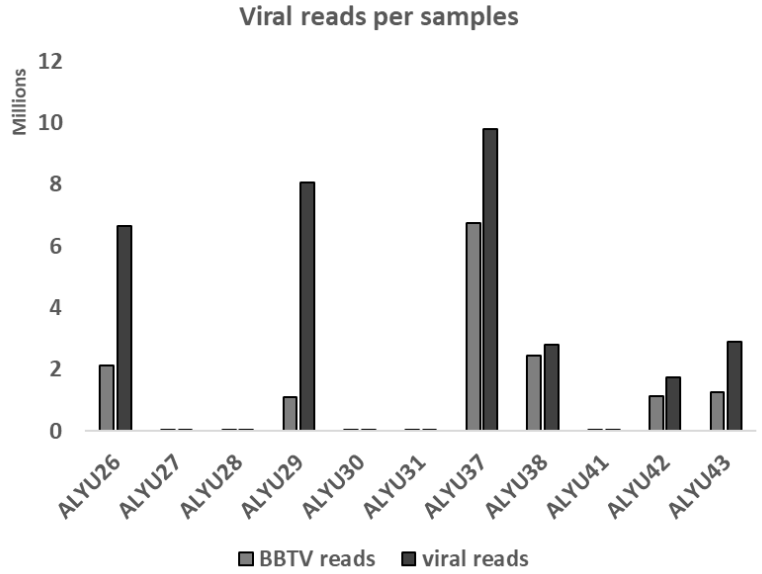

Musa acuminata Cavendish_ALYU-43_LAO
Perfect match
C M N R S U3 BBTA5
With mismatches
C M N R S U3 BBTA5
*Chromolaena odorata*_ALYU-38_LAO
Perfect match
C M N R S U3 BBTA5
With mismatches
C M N R S U3 BBTA5

S3H Fig.


Musa ABB Klue Tiparot_ALYU-46_LAO
Musa acuminata wild_ALYU-52_CHN
Perfect match
Perfect match
With mismatches
With mismatches
C M N R S U3
C M N R S U3
C M N R S U3
C M N R S U3

Plant	Samples	BBTV reads	All viral reads	Total number reads	% of viral reads
Musa	ALYU-26	2,343,872	7,190,098	10,455,772	0.6376
Ipomoea	ALYU-28	2,170	3,376	12,488,618	0.0003
Musa	ALYU-29	1,088,396	8,063,084	11,642,444	0.6926
Phyllanthus	ALYU-30	6,022	7,842	13,503,218	0.0006
Arachis	ALYU-31	1,042	1,452	10,347,832	0.0001
Musa	ALYU-46	259	259	18,154,862	0.0000
Musa	ALYU-52	1,685	1,685	10,763,616	0.0002

Plant	Samples	BBTV reads	All viral reads	Total number reads	% of viral reads
Musa	ALYU-26	2,343,872	7,190,098	10,455,772	0.6376
Commelina	ALYU-27	14,932	53,512	11,731,770	0.0043
Musa	ALYU-37	6,742,128	9,793,147	10,828,680	0.9044
Bidens	ALYU-38	2,435,471	2,808,035	11,763,264	0.2387
Chromolaena	ALYU-41	6,999	17,142	11,534,294	0.0015
Musa	ALYU-42	1,132,366	1,747,470	13,002,052	0.1344
Musa	ALYU-43	1,278,482	2,905,982	13,238,836	0.2195



S4 Fig. Counts and percentages of BBTV and BBTV+alphasatellite reads in total number of Illumina 125-nt reads

Résumé

Le banana bunchy top virus (BBTV) est le virus le plus dommageable de la culture de la banane, une monocotylédone considérée comme la 4ème plus grande culture mondiale. Appartenant à la famille des *Nanoviridae* et au genre *Babuvirus*, le BBTV est un virus à ADN circulaire simple brin (ssDNA) multipartite composé de 6 ADN viraux différents. Transmis de plante à plante par le puceron vecteur *Pentalonia nigronervosa*, ce virus peux également être retrouvé associé avec un alphasatellite. Un alphasatellite est une molécule d'ADN circulaire simple brin appartenant à la famille des *Alphasatellitidae*, qui code pour une protéine nommée Rep-like lui permettant de s'autorépliquer. Toutefois, il ne peut pas trans-répliquer son virus assistant et nécessite ce dernier afin de se déplacer dans la plante, être encapsider et être transmis de plante à plante. Très peu d'étude ont été menées sur les alphasatellites et leur rôle dans l'infection virale n'est pas défini. Dans ces travaux de thèse, nous montrons l'émergence d'un nouvel alphasatellite associé au BBTV provenant de République Démocratique du Congo. C'est le premier alphasatellite découvert associé au BBTV en Afrique. Sa caractérisation a montré qu'il impactait son virus assistant au travers de la transmission, de la réplication et la transcription virale, et de la production des petits ARNs viraux. De plus, cet alphasatellite est plus proche phylogénétiquement des alphasatellites découverts chez les dicotylédones que ceux découverts associés au BBTV et forme le genre des *Banaphisatellite* que nous décrivons. De plus, notre étude d'échantillons de BBTV provenant d'Asie du sud-est montre la forte prévalence d'alphasatellite dans cette région du monde et la découverte des alphasatellites formant deux autres espèces différentes du genre *Banaphisatellite*. Enfin, notre caractérisation des plantes dicotylédones montre qu'elles sont des hôtes alternatifs du BBTV et leur alphasatellites. Possédant également des alphasatellites du genre *Banaphisatellite*, elles appuient l'hypothèse d'un passage des dicotylédones à monocotylédone de ce genre d'alphasatellite.

Abstract

Banana bunchy top virus (BBTV) is the most damaging virus of monocot bananas and plantains, considered the 4th largest crop plants in the world. Belonging to the family *Nanoviridae* and the genus *Babuvirus*, BBTV is a multipartite circular single-stranded DNA (ssDNA) virus composed of six viral DNA components. Transmitted from plant to plant by the aphid vector *Pentalonia nigronervosa*, this virus can also be associated with one or more alphasatellites. Alphasatellites are circular ssDNA molecules belonging to the family *Alphasatellitidae* which encode a Rep-like protein that allows alphasatellites to self-replicate. However, alphasatellites cannot trans-replicate helper virus components and require the helper virus for encapsidation, movement and transmission by insect vectors. Very few studies have been conducted on alphasatellites, and their role in viral infection and disease transmission is not defined. In this thesis, we report the emergence of a new alphasatellite associated with BBTV in Democratic Republic of the Congo. This is the first BBTV alphasatellite discovered in Africa. We demonstrate that this alphasatellite affects its helper virus transmission, replication, gene expression and evasion of antiviral defences generating viral small interfering RNAs. Furthermore, this alphasatellite forms a new genus *Banaphisatellite* and is more phylogenetically related to alphasatellites infecting dicots than those previously found to be associated with monocot-infecting BBTV in South-East Asia. In addition, our surveys of BBTV-infected plants from South-East Asia showed the high prevalence of alphasatellites in this region of the world and among other alphasatellites we discovered two additional novel species representing the genus *Banaphisatellite*. Finally, we demonstrate that dicot plants can serve as alternative hosts of BBTV and its alphasatellites including the banaphisatellites, thus supporting the hypothesis of a dicot to monocot passage of this alphasatellite genus.